

治療効果に対する代替性の評価尺度

小林 史明

博士 (統計科学)

総合研究大学院大学

複合科学研究科

統計科学専攻

2014

概要

医薬品の臨床試験では、臨床エンドポイントの代用として、代替エンドポイントを用いた評価が行われることがある。代替エンドポイントを適切に選択することは、新薬の開発戦略の要であるが、既存の統計的な代替エンドポイントの評価方法はいくつかの問題点を抱えている。そこで本論文では、既存の代替性評価尺度の問題点を解決し、かつ臨床家でも理解しやすい性質を備えた、統計的関連性を用いた代替性の評価尺度 **proportion of the treatment effect captured by candidate surrogate endpoint (PCS)**、および推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法を提案する。また、統計的因果推論の考え方を用いた代替性評価尺度への改良も行う。さらに、数値実験と無作為化比較試験への適用事例をとおして、提案方法が代替エンドポイントの候補の代替性を適切に評価できることを示す。本論文の結果は、治療効果に対する代替性を信頼性を持って評価することに貢献する。

目次

第1章 導入	5
1.1 臨床試験におけるエンドポイント	5
1.2 代替性の評価尺度	6
1.2.1 既存の統計的な代替エンドポイントの評価方法	6
1.2.2 本論文で取り組む課題1	7
1.2.3 本論文で取り組む課題2	8
1.3 本論文の構成と取り扱うデータの概要	9
第2章 統計的因果推論に関する基礎的背景	12
2.1 はじめに	12
2.2 構造的因果モデル	15
2.2.1 データ生成過程と有向グラフ	15
2.2.2 自律性と因果効果	20
2.3 構造的因果モデルと潜在反応モデルの関係	25
2.3.1 Pearl 流の潜在反応モデル	25
2.3.2 因果効果	30
2.4 グラフィカルモデルに基づく統計的因果推論の有用性	33
2.4.1 グラフィカルモデルの問題点	33
2.4.2 共変量選択問題	36
2.4.3 グラフによる記述能力	40

2.5	まとめ	41
第3章	統計的関連性を用いた治療効果に対する代替性の評価尺度	43
3.1	はじめに	43
3.2	準備	44
3.3	提案手法	48
3.3.1	Proportion of the treatment effect captured by candidate surrogate end-point (PCS)	48
3.3.2	代替性の水準を判断するためのカットオフ値	50
3.3.3	符号付 PCS による代替エンドポイントの区別	52
3.3.4	推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法	54
3.4	数値実験	55
3.4.1	設定	55
3.4.2	推定	57
3.4.3	結果	57
3.5	適用例	63
3.5.1	MEGA Study	63
3.5.2	ARMD Study	66
3.6	まとめ	69
3.7	付録	70
3.7.1	治療効果が比によって定義される場合の CP と NCP の定式化	70
3.7.2	数値実験による平均二乗誤差の評価	72
3.7.3	ARMD study での PIG の値に関する考察	74

第4章	自然な因果効果を用いた治療効果に対する代替性の評価尺度	76
4.1	はじめに	76
4.2	準備	77
4.3	自然な因果効果を用いた代替性評価尺度：C-PCS	80
4.4	仮想的な数値例	81
4.5	適用例	83
4.5.1	データの概要と代替性評価の必要性	83
4.5.2	事前検討	84
4.5.3	解析結果1 (X, S, Y のみを用いた検討)	85
4.5.4	解析結果2 (X, S, Y と Z を用いた検討)	89
4.5.5	解析結果3 (治療期間平均を用いた検討)	94
4.6	まとめ	99
第5章	総括	100
5.1	本論文で取り組んだ課題と提案手法の特徴	100
5.2	本論文の貢献	101
5.2.1	臨床的観点からみた本論文の貢献	101
5.2.2	統計的観点からみた本論文の貢献	103
5.3	今後の展開および課題	104
5.4	まとめ	105

第1章 導入

1.1 臨床試験におけるエンドポイント

患者にとって価値のある医薬品を、1日でも早く提供するにはどうすれば良いか。科学的な観点からこの問いに答えるため、様々な医薬品の評価方法が議論されてきた。その中でも、無作為化比較試験において、治療効果をどのような測定値で評価するかは、最も重要なテーマの一つである。本論文ではこのような治療効果を判断するための評価項目のことをエンドポイントと呼ぶ。高橋 (1967, p.192-p.193) は、エンドポイントの設定にあたっては、“本質的には専門的な医学知識のほうから決められるべきものである。”さらには、“病気の治癒のしくみにとって的を射たものであるかどうかがよく吟味されていなければならない。”と述べている。また Fleming (1996) は、エンドポイントの選択における二つの重要な基準として、治療効果を判断するための感度があること、および臨床的な妥当性があることを挙げている。このような観点によって選ばれたエンドポイントのことを、本論文では臨床エンドポイントと呼ぶことにする。

冒頭の問いのうち、“患者にとって価値のある”については、上記の臨床エンドポイントを直接評価する試験において、薬効が認められた医薬品を提供することで応えることができる。しかし、“1日でも早く提供する”に応えるには、困難を伴うことがある。例えば、臨床エンドポイントが死亡であった場合に、その臨床エンドポイントを観測するには長い試験期間が必要となることがある。また、臨床エンドポイントが心筋梗塞や末期腎不全といった稀なイベントである場合、十分なイベント数を確保するために、試験の被験者数を増やす必要がある。こういった測定に時間がかかる、もしくは発現が稀な臨床エンドポイント

を用いた試験が常に必須であったとき、“患者に早期に医薬品を届ける”ことは困難となる。この問題に対処するため、臨床エンドポイントの代わりに評価項目として、代替エンドポイントと呼ばれる評価項目を利用する考えが生まれた。代替エンドポイントに求められる実用上の性質として、臨床エンドポイントより早期に、より簡便に測定できることが求められている (Molenberghs et al., 2005; Stevens et al., 2006)。しかし、代替エンドポイントの候補と考えられる測定値 (治療開始から臨床エンドポイント測定の間測定される値。以降、代替エンドポイントの候補と呼ぶ) が、どのような要件を満たせば代替エンドポイントとして妥当であるかは明確ではない。この点を統計的に検討する取り組みが過去 20 年以上に渡って進められている (Prentice, 1989; Weir and Walley, 2006)。

1.2 代替性の評価尺度

本節では、既存の代替エンドポイントの統計的な評価方法について概観し、本論文で取り組む課題を明確にする。

1.2.1 既存の統計的な代替エンドポイントの評価方法

Prentice (1989) は代替エンドポイントの統計的な定義を提案し、続いて実利用のための統計的な基準 (Prentice’s operational criteria, 以降、Prentice 基準と呼ぶ) を示した。その具体的な基準とは、代替エンドポイントの候補に対し治療効果が存在すること、臨床エンドポイントに対し治療効果が存在すること、代替エンドポイントの候補は臨床エンドポイントと関連があること、そして、代替エンドポイントの候補を与えたとき、治療は臨床エンドポイントと条件付き独立であること、の四つの基準をすべて満たすことである。Prentice 基準は、臨床的に受け入れられる概念 “capture any relationship between the treatment and the true response” を背景に導き出されたものである (Prentice, 1989)。しかしながら、Prentice 基準が意味するところは、代替エンドポイントの候補が治療効果を ‘完全に’ 捕捉 (capture) してい

ることである。しかし、これは現実的には非常に厳しい基準であり、この基準を完全に満たすことは困難である。

1.2.2 本論文で取り組む課題 1

前記のような Prentice 基準を満たすことの困難さを回避するため、Freedman et al. (1992) は、代替エンドポイントが臨床エンドポイントへの治療効果を‘完全に’ではなく、‘部分的に’捉えていると考え、その割合を評価することに着目した。すなわち、治療効果の全体に対し、どの程度の割合を代替エンドポイントの候補が説明しているのかを評価する取り組みが始まった。本論文では、このような割合を‘代替性の評価尺度’と呼ぶ。Freedman et al. (1992) は、proportion of the treatment effect explained (PTE) という代替性の評価尺度を提案した。PTE は一つの臨床試験のデータから比較的簡単な計算で求められるメリットがあり、近年になっても臨床家に広く利用されている (例えば、DePrimo et al. (2009); Boekholdt et al. (2012))。これに対し Wang and Taylor (2002) は、PTE のばらつきが大きいことの問題点を指摘し、これを解決する proportion explained (PE) という代替性の評価尺度を提案した。また、Qu and Case (2007) は、カルバック-ライブラー情報量の増加度合に基づいた、proportion of information gain (PIG) という代替性の評価尺度を提案した。しかし、これら既存の代替性の評価尺度について、生物統計分野の研究者や実践家により次の問題点が指摘されている。

- (i) 既存の代替性評価尺度の多くは、特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとることがある (Li et al., 2001; Wang and Taylor, 2002)
- (ii) 既存の代替性評価尺度は、代替性の水準を判断するための適当なカットオフ値を定義できない (Weir and Walley, 2006; Qu and Case, 2007)
- (iii) 既存の代替性評価尺度の多くは、そのばらつきが大きく、そのため信頼区間が許容できないほどに広いことがある (Lin et al., 1997; De Gruttola et al., 1997)

そこで、本論文では、既存の代替性評価尺度の問題点を解決し、かつ臨床家でも理解しやすい性質を備えた、統計的関連性を用いた代替性の評価尺度 **proportion of the treatment effect captured by candidate surrogate endpoint (PCS)** を提案し、併せて推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法も提案する。

1.2.3 本論文で取り組む課題 2

Prentice 基準や PTE などとは異なるアプローチとして、複数の臨床試験データを用いたメタアナリシスアプローチ (Buyse et al., 2000)、Rubin 流の因果推論に基づく主要層別を用いたアプローチ (Frangakis and Rubin, 2002)、そして Pearl 流の統計的因果推論の議論をふまえたアプローチが近年提案されている (Lauritzen, 2004; Chen et al., 2007; Ju and Geng, 2010; VanderWeele, 2013)。しかし、メタアナリシスアプローチは適切なデータ収集のための基盤が十分に整備されていないという問題点があり (Buyse et al., 2010; Sargent and Mandrekar, 2013)、また主要層別を用いたアプローチも識別条件が検証不能であるという問題点がある (田中他, 2010)。一方で、Pearl 流の統計的因果推論の議論をふまえたアプローチでは、治療、代替エンドポイントそして臨床エンドポイントを取り巻く因果的な構造を明らかにしながら代替性を評価することが可能であり注目を集めている (Lauritzen, 2004; Chen et al., 2007; Ju and Geng, 2010; VanderWeele, 2013)。このことは、先ほど述べた薬効を評価するエンドポイントとしての“病気の治癒のしくみにとって的を射たものであるかどうかがよく吟味されていなければならない” (高橋, 1967) に呼応する取り組みと考えられる。そこで、本論文では、統計的因果推論に関する基礎的な概念を確認したうえで、Pearl 流の統計的因果推論を用いた代替性の評価尺度を提案する。

1.3 本論文の構成と取り扱うデータの概要

第2章では、統計的因果推論を用いた代替性評価尺度を提案する準備として、統計的因果推論に関する基礎的な背景(黒木・小林, 2012)を述べる。第3章では、新たな統計的関連性を用いた代替性評価尺度 PCS を提案し、併せて推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法も提案する(Kobayashi and Kuroki, 2014a)。第4章では、第3章で提案した統計的関連性を用いた代替性評価尺度を、統計的因果推論の観点から再定義し、自然な因果効果を用いた代替性評価尺度 Causal PCS (C-PCS) を提案する(Kobayashi and Kuroki, 2014b)。最後に、第5章において本論文の総括を述べる。

ここで、本論文で取り扱う具体的なデータの概要を表 1.1 に示す。

一つ目の Management of Elevated Cholesterol in the Primary Prevention Group of Adult Japanese study (MEGA study) は、日本で初めて行われたプラバスタチンの冠動脈疾患 (CHD) の予防効果を評価した無作為化比較試験である(Nakamura et al., 2006)。対象となった8,214例が、対照群(食事制限のみ)もしくはプラバスタチン群(食事制限とプラバスタチン投与)に無作為に割りつけられ、試験開始時から終了時までの脂質値(低比重リポ蛋白コレステロール [LDL-C], 非高比重リポ蛋白コレステロール [non-HDL-C]) が測定された。LDL-C については、従来からこれを代替エンドポイントとした新薬の臨床開発が行われていたが、近年になりこの妥当性を問い直すべき結果が得られている(Barter et al., 2007; Psaty and Lumley, 2008)。他方で、non-HDL-C については、CHD イベントに対する高い予測能をもつことが最近の研究で示されている(Boekholdt et al., 2012; Ingelsson et al., 2007; Liu et al., 2006)。これらの状況をふまえ、本論文では、CHD イベント発症予防効果に対する脂質値(LDL-C, non-HDL-C)の代替性を検討する。

二つ目の Age Related Macular Degeneration study (ARMD study) は、プラセボとインターフェロン α を加齢黄斑変性患者に投与した効果を比較することを目的とした試験である(Pharmacological therapy for macular degeneration study, 1997)。臨床エンドポイントは、1年

経過時の視力検査であり、この検査結果を用いてインターフェロン α の臨床効果が判定されている。これに対して、より早期に治療効果を判定するために、6ヵ月経過時の視力検査を代替エンドポイントの候補とすることが検討されている (Buyse and Molenberghs, 1998). 本論文においても, Buyse and Molenberghs (1998) が代替性を評価した際のデータを用いて, 臨床エンドポイント (1年経過時の視力検査) に対する, 6ヵ月経過時の視力検査の代替性を検討する.

三つ目の Olmesartan Reducing Incidence of End stage Renal Disease in Diabetic Nephropathy Trial (ORIENT) は, 日本および香港の顕性腎症を伴う 2 型糖尿病患者 566 例を対象に, プラセボを対照としてオルメサルタン投与による 腎症進展抑制効果や心血管イベント抑制効果を比較した無作為化比較臨床試験である (Imai et al., 2011). ORIENT の臨床エンドポイントは, 血清クレアチニン値 (血清 Cr 値) の 2 倍化, 末期腎不全 (血清 Cr 値 5.0mg/dL 以上, 透析, 腎移植), 死亡から構成される腎複合イベントである. ORIENT では, 尿蛋白の変化率, 血圧, 腎機能の低下速度 (血清 Cr 値の逆数の推移による評価) が, 代替エンドポイントの候補と考えられているが, これらの候補に対する治療効果の発現時期が異なることや併用薬による治療の有無によって治療効果の大きさが異なることが確認されている (Imai et al., 2011, 2013a,b). そこで本論文では, 腎複合イベントに対する尿蛋白の変化率, 血圧, 腎機能の低下速度の代替性を検討する.

本論文では, 第 3 章において一つ目の MEGA study と二つ目の ARMD study のデータを検討し, 第 4 章において三つ目の ORIENT データを検討する.

表 1.1: 本論文において取り扱うデータの概要

試験名	治療	代替エンドポイント の候補	臨床エンドポイント
MEGA study ^{*1}	食事制限のみ, 食事制限と プラバスタチン	LDL-C ^{*2} , non-HDL-C ^{*3}	冠動脈疾患
ARMD study ^{*4}	プラセボ, インターフェロン α	6ヵ月経過時の 視力検査	1年経過時の 視力検査
ORIENT ^{*5}	プラセボ, オルメサルタン	血圧, 尿蛋白, 腎機能の低下速度	腎複合イベント

*1: Management of Elevated Cholesterol in the Primary Prevention Group of Adult Japanese study (Nakamura et al., 2006); *2: 低比重リポ蛋白コレステロール; *3: 非高比重リポ蛋白コレステロール; *4: Age Related Macular Degeneration study (Pharmacological therapy for macular degeneration study, 1997); *5: Olmesartan Reducing Incidence of End stage Renal Disease in Diabetic Nephropathy Trial (Imai et al., 2011)

第2章 統計的因果推論に関する基礎的

背景

本章では、統計的因果推論を用いた代替性評価尺度を提案する準備として、統計的因果推論に関する基礎的な背景を述べる¹。具体的には、Pearl 流の統計的因果推論 (Pearl, 2009a) の基本的な考え方を、Rubin 流の統計的因果推論 (Rubin, 1974, 1978, 2006) とは何が異なるのかを意識しつつ述べる。

2.1 はじめに

潜在反応アプローチ (Potential Response Approach, Potential Outcome Approach) は、対象者に対してある治療を行った際に現れる反応を、その対象者が持つ特徴と決定論的に結びつけたうえで、ランダムサンプリングなどの統計的要素を付加して因果効果 (Causal Effect) の定量的評価を試みる統計的アプローチの一つである (Pearl, 2009a; Rubin, 2006)。現在、潜在反応アプローチの“中核”をなす潜在反応モデル (Potential Response Model, Potential Outcome Model) は、統計的因果推論 (Statistical Causal Inference) に関するあらゆる問題を議論するのに必要不可欠な統計的因果モデル (Statistical Causal Model) として位置づけられつつあるだけでなく、医学・疫学分野においても暴露効果や治療効果を評価するのに重要な役割を果たしている。このことは、国際的な計量生物学会誌である *Biometrics* や *Statistics in Medicine* をはじめとして、ここ数年の間に刊行された多くの医学統計関連の学術誌を開けば、毎年のように“Counterfactual (反事実)”, “Potential Outcome (潜在結果)”

¹本章の内容は、(黒木・小林, 2012) をまとめたものである。

あるいは“Potential Response (潜在反応)”といった用語を目にすることから明らかであろう (たとえば, Albert and Nelson (2011); Cai et al. (2007, 2008); VanderWeele and Shpitser (2011); VanderWeele et al. (2012)). また, 黒木 (2009), 宮川 (2004), 佐藤・松山 (2002) による統計因果推論や因果ダイアグラム (Causal Graph, Causal Diagram) の解説に始まり, 統計関連学会連合大会で“計量生物学における統計的因果推論の役割”(2010年度)や“統計学初級中級講座「統計学的因果推論入門」”(2011年度)と題する企画セッションが行われていたことからわかるように, 日本国内においても, 統計科学研究者・計量生物学研究者の統計的因果推論や潜在反応モデルに対する関心が徐々に高まっていることを伺い知ることができる.

表 2.1 に示すように, 潜在反応モデルに立脚した統計的因果推論には, Rubin 流の統計的因果推論 (Rubin's Causal Models [RCMs] が使われる) (Rubin, 1974, 1978, 2006) と, Pearl 流の統計的因果推論 (構造的因果モデル [Structural Causal Models, SCMs] が使われる) (Pearl, 2009a) の 2 つの“流派”がある. Rubin 流の統計的因果推論は欠損値データの解析法 (Little and Rubin, 2002) を基調とするものであり, 国内外の統計科学の研究者に広く受け入れられている (Pearl, 2009a). 一方, Pearl 流の統計的因果推論は構造方程式モデル (Structural Equation Model; Bollen (1989); Wold (1954)) やベイジアンネットワーク (Bayesian Network; Pearl (1988)) を基調としたもので, Wright (1923, 1934) のパス解析のノンパラメトリックモデルへの拡張 (Pearl, 1995, 2009a) ともいえるものである. その中でも, グラフィカルモデルを利用した統計的因果推論 (たとえば, 外的操作の数学的表現である“set”あるいは“do”オペレーションを用いた因果効果の識別可能性問題や観測データを利用した因果構造発見問題) については情報科学研究者を中心に積極的に研究されてきた. また, この基礎理論をつくりあげた Judea Pearl 自身, ベイジアンネットワークの理論を体系的に整備した人工知能研究者であることもあって, ほとんどの結果がグラフ用語で記述されている (あるいは, Judea Pearl によって執筆された統計的因果推論に関するほとんどの論文で, グラフ用語で記述された結果が与えられている) ことは, ある意味で Pearl 流の統計的因果

表 2.1: 統計的因果推論の 2 つの“流派”で用いられる基本的な仮定の違い

Rubin 流の統計的因果推論: Rubin's Causal Models (RCMs) <ul style="list-style-type: none"> - 潜在反応変数の存在, 一致性 (Consistency) - SUTVA(No Interference between Subjects, No Multiple Versions of Treatment)
Pearl 流の統計的因果推論: Structural Causal Models (SCMs, 構造的因果モデル) <ul style="list-style-type: none"> - 自律的なデータ生成過程 (Autonomous Data Generating Process) - No Interference between Subjects

推論の特徴であるともいえる。しかし、このことが、**Pearl 流**の統計的因果推論に基づいて議論する際にはグラフィカルモデルを必ず用いなければならないもの、あるいは極端なケースではグラフィカルモデルと構造的因果モデルとが同値であるかのような誤解を与えているようにも見受けられる。

このような状況をふまえて、本章では、**Pearl 流**の統計的因果推論、すなわち、構造的因果モデルの基本的な考え方を解説する。特に、**Rubin 流**の統計的因果推論とは何が異なるのかを意識しつつ、(1) グラフィカルモデルに基づく統計的因果推論よりも潜在反応モデルに基づく統計的因果推論 (**RCM** と **SCM**) のほうが因果関係を詳細に表現できること (したがって、**Pearl 流**の統計的因果推論は、グラフィカルモデルと同値ではない) こと、そして、その一方で (2) 因果関係をグラフによって視覚化しないとミスリーディングな結果を導く可能性がある (それゆえ、**Pearl 流**の統計的因果推論では因果関係をグラフを用いて表現することの重要性が強調される) ことを示す。具体的には、(1) については、潜在反応モデルの特徴を利用した因果効果の評価方法である **Bounding Method** (Balke and Pearl, 1997; Cai et al., 2007, 2008; Kuroki and Cai, 2008, 2011; Kuroki et al., 2010; MacLehose et al., 2005; Tian and Pearl, 2000) を例として、基本的な対象者レベルの因果的仮定である“単調性” (**Monotonicity Assumption**: この仮定はしばしば“あまのじゃくな対象者はいない”ことにたとえられる (佐藤, 2006)) でさえ、グラフィカルモデルでは記述できないことを説明する。(2) については、**M-バイアス** (**M-bias**: Greenland (2003); Greenland et al. (1999)) や操作変数法 (**Instrumental Variable [IV] Method**: Bowden and Turkington (1984); Greenland

(2000); Pearl (2009a)) を例として、Rubin 流の統計的因果推論でしばしば述べられている共変量選択指針がバイアスを導く可能性があることを述べる。加えて、グラフィカルモデルを用いることで因果効果を推定するのに十分な共変量を適切に選択できることを示す。

2.2 構造的因果モデル

前節で述べたように、本章の目的は Pearl 流の統計的因果推論 (構造的因果モデル) の基本的な考え方を解説することである。その準備として、本節では、構造的因果モデルの基本的な概念 (データ生成過程 [Data Generating Process], 自律性 [Autonomy]) を紹介するとともに、因果効果の定式化を行う。なお、本章では、単に潜在反応モデルといった場合は Pearl 流と Rubin 流の両方の潜在反応モデルを示し、それぞれの流派を区別して扱う場合には“Pearl 流の”もしくは“Rubin 流の”と冠をつけて示す。

2.2.1 データ生成過程と有向グラフ

構造的因果モデルは、因果関係は何らかの関数関係をとおして決定論的に記述できるという考え方に基づいて構築された因果モデルである。興味ある確率変数の集合 $\mathbf{V} = \{V_1, \dots, V_p\}$ の要素間の関係が構造方程式モデル (Structural Equation Model)

$$V_j = g_j(\text{pa}(V_j), \epsilon_j), \quad j = 1, 2, \dots, p \quad (2.1)$$

によって規定されており、それぞれが自律的 (Autonomous) なデータ生成過程をなすとき、式(2.1)を構造的因果モデルという。ここに、 $\text{pa}(V_j)$ は興味ある因果的な現象において V_j の直接的原因 (Direct Cause) と解釈される変数集合 ($\text{pa}(V_j) \subset \mathbf{V}$) である。また、 ϵ_j は $\text{pa}(V_j)$ では表現されることはない変数の集合を要約した錯乱項 (Random Disturbance) である。錯

乱項は、興味ある因果的現象を取り巻く環境など示す要因のうち \mathbf{V} で直接的に表現することが難しい個体差や要因、たとえば、手術の際の病院の設備や医師の技量など、 V_j と $\text{pa}(V_j)$ の間の決定論的關係を錯乱させる要因を要約したものと解釈される。年齢や性別、生活習慣など、一般に因果効果を評価するのに重要と考えられる要因であっても、因果的仮説を記述する際に重要なものと認識されていなかったり、観測されなければ、錯乱項として要約されることもある。したがって、何を錯乱項に含めるかは、興味ある因果的現象の範囲をどこまで捉えるかに依存する。自律的 (Autonomous) であるとは、ある関数 $V_k = g_k(\text{pa}(V_k), \epsilon_k) (V_k \in \mathbf{V})$ の関数形が変化してもそれが他の関数形を変化させることはない、すなわち、個々の構造方程式が独立したデータ生成メカニズム²であることを意味する。

さて、式 (2.1) から、以下のことがわかる。第一に、式 (2.1) は、 \mathbf{V} の各要素は他の変数によって規定される変数であることを意味しているだけであって、観測可能であるかどうか、外的操作 (Intervention) が可能であるかどうかを問わない。第二に、式 (2.1) の逐次的な代入を繰り返すことにより、 \mathbf{V} の各要素は錯乱項 $\{\epsilon_1, \dots, \epsilon_p\}$ のみ、あるいは錯乱項と外生変数 (錯乱項のみによって規定される変数) を用いて表現することができる (Pearl, 2009a)。この考察によって、次節で紹介する潜在反応モデルとノンパラメトリックな構造方程式モデルが結びつけられる。第三に、式 (2.1) は、 \mathbf{V} の各要素 V_j はその直接的原因に対応する変数集合 $\text{pa}(V_j)$ とそれに付随する錯乱項 $\epsilon_j (j = 1, \dots, p)$ によって生成されるのであって、式 (2.1) には現れていない構造方程式によって生成されることはない。第四に、錯乱項 $\epsilon_1, \dots, \epsilon_p$ が独立であるという仮定の下で、式 (2.1) が与えられたとき、これに対応する \mathbf{V} の同時分布 $\text{pr}(v_1, \dots, v_p)$ の逐次的因数分解 (Recursive Factorization) として

$$\text{pr}(v_1, \dots, v_p) = \prod_{j=1}^p \text{pr}\{v_j | \text{pa}(v_j)\} \quad (2.2)$$

²この仮定も緩めることができる (Kuroki, 2007)。

表 2.2: CCP データ (Kuroki and Cai, 2008; MacLehose et al., 2005)

	黒人 ($Z = z_0$)		白人 ($Z = z_1$)	
	β 遮断薬の服用		β 遮断薬の服用	
	なし ($X = x_0$)	あり ($X = x_1$)	なし ($X = x_0$)	あり ($X = x_1$)
生存 ($Y = y_0$)	4224	2143	63449	34868
死亡 ($Y = y_1$)	1254	60	22191	1103
総数	5478	2203	85640	35971

を与えることができる。ここに、錯乱項どうしに関連が見られる場合にも形式的に同時分布の逐次的因数分解を与えることは可能であるが、その場合には条件付き分布が因果的な意味を持つとは限らないことに注意する。このような場合には、関連を持つ錯乱項どうしの関係を新たな (関連のない) 錯乱項と共通原因 (Common Cause) によって表現した上で、同時分布の逐次的因数分解が行われる (Pearl, 2009a; Spirtes et al., 2001)。したがって、式 (2.1) において錯乱項どうしに関連はないと仮定しても議論の本質に影響を与えることはないため、特に断らない限り、本章では錯乱項どうしは独立であると仮定して議論を進めることとする。

例として、Gan et al. (2000), Kuroki and Cai (2008), そして MacLehose et al. (2005) によって解析された、CCP (Cooperative Cardiovascular Project) データ (Ellerbeck et al., 1995) を考える。CCP は、アメリカの公的医療保険制度であるメディケアの受益者に対し、提供される医療サービスの質の評価を目的としたプロジェクトであり、1994 – 1995 年に急性心筋梗塞疾患で入院した 20 万人以上のメディケア受益者全例について、各医療施設のカルテなどから診断や治療および予後に関する情報が収集されている (Gan et al., 2000; 野口他, 2003)。この CCP データを用いた解析の目的の一つに、 β 遮断薬の服用が心筋梗塞による死亡率をどの程度低下させるかがあった。Kuroki and Cai (2008) および MacLehose et al. (2005) によって解析されたデータを表 2.2 に与える。

X を β 遮断薬の服用 ($X = x_0$: 服用なし, $X = x_1$: 服用あり), Y を 30 日以内の対象者の状態 ($Y = y_0$: 生存, $Y = y_1$: 死亡), Z を人種 ($Z = z_0$: 黒人, $Z = z_1$: 白人) とす

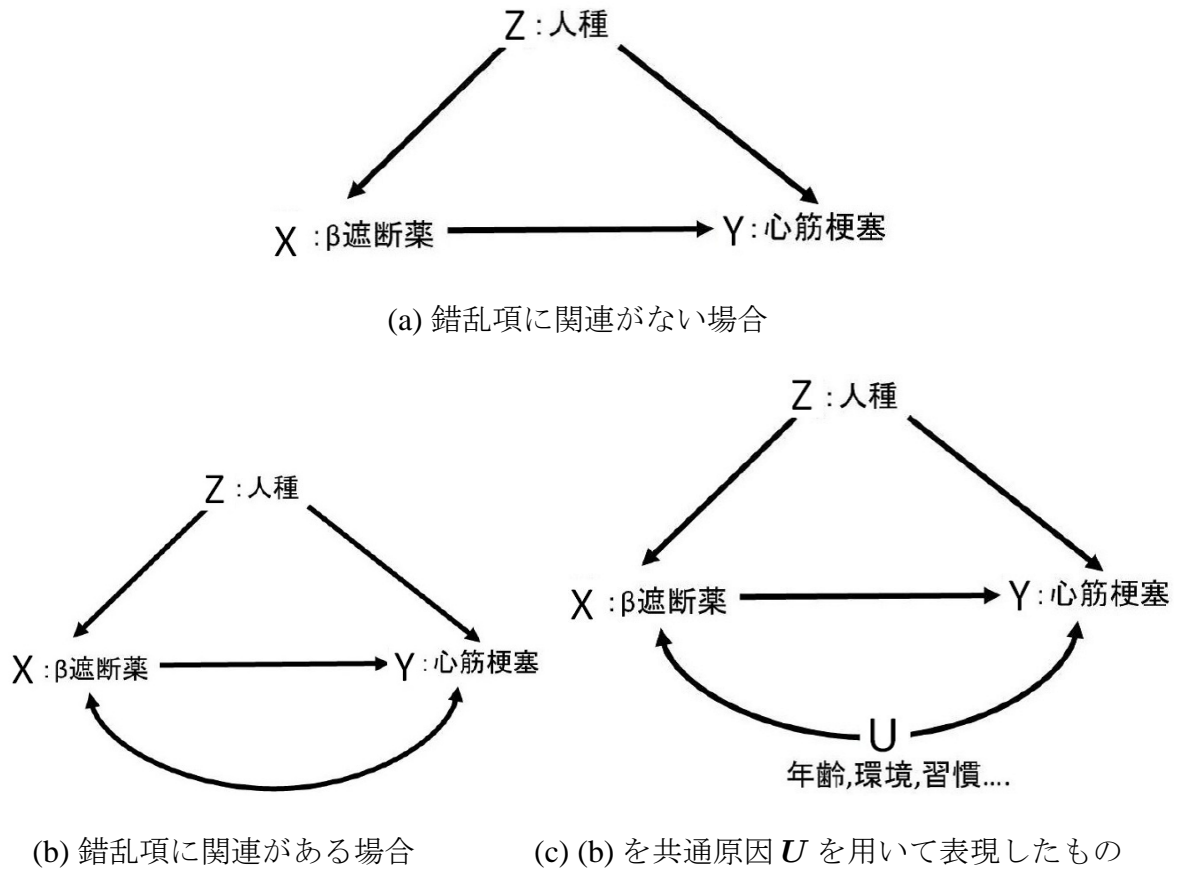


図 2.1: $\{\beta$ 遮断薬服用の有無, 心筋梗塞による生存・死亡, 人種 $\}$ の間の因果関係を表現した因果ダイアグラムの例. ただし, (a) では錯乱項どうしには関連がないことが仮定されており, (b) では錯乱項 ϵ_x と ϵ_y に関連があると仮定されている. (c) は錯乱項 ϵ_x と ϵ_y の間に共通原因を組み込むことにより, (b) を表現しなおしたものである.

る. このとき, 対象者の状態 (Y) は人種 (Z), β 遮断薬服用の有無 (X) およびこれらの項目では表現できない要因 (ϵ_y) によって決定され, β 遮断薬服用の有無 (X) は人種 (Z) および人種以外の要因 (ϵ_x) によって決定されると仮定する. ここに, $\epsilon_x, \epsilon_y, \epsilon_z$ は年齢や生活習慣などといった, X, Y, Z では表現されることのなかった要因の集合を要約したものと解釈することができる.

このときの構造的因果モデルは

$$X = g_x(Z, \epsilon_x), Y = g_y(Z, X, \epsilon_y), Z = g_z(\epsilon_z) \quad (2.3)$$

と記述されるが、この因果モデルを

$$X = g_x(Z, \epsilon_x), Y = g_y(Z, X, \epsilon_y) = g_y\{Z, g_x(Z, \epsilon_x), \epsilon_y\}, Z = g_z(\epsilon_z) \quad (2.4)$$

あるいは

$$\left. \begin{aligned} X &= g_x\{g_z(\epsilon_z), \epsilon_x\} (\triangleq h_x(\epsilon_x, \epsilon_z)), Z = g_z(\epsilon_z) (\triangleq h_z(\epsilon_z)) \\ Y &= g_y\{g_z(\epsilon_z), X, \epsilon_y\} = g_y[g_z(\epsilon_z), g_x\{g_z(\epsilon_z), \epsilon_x\}, \epsilon_y] (\triangleq h_y(\epsilon_z, \epsilon_x, \epsilon_y)) \end{aligned} \right\} \quad (2.5)$$

と書き換えることもできる。このことから、個々の対象者に関するデータは式 (2.4) の $\epsilon_x, \epsilon_y, Z$ 、あるいは式 (2.5) の $\epsilon_x, \epsilon_y, \epsilon_z$ に値が代入されることによって逐次的に生成されたものとみなされる。

Pearl 流の統計的因果推論では、構造的因果モデルが与えられたとき、これを因果ダイアグラムと呼ばれる有向グラフ (Directed Graph) を用いて表現する。因果ダイアグラムは、変数間に直接的な因果関係がみられる場合には原因に対応する変数からその結果に対応する変数へ矢線 (Arrow, Directed Edge; \rightarrow) を引き、錯乱項どうしに関連がみられる場合には、対応する変数どうしを双方向の矢線 (Bidirected Arrow, Bidirected Edge, Bow, Confounding Arc; \longleftrightarrow) で結ぶことによって構成される。したがって、構造的因果モデルでは変数の集合 $\text{pa}(V_j)$ が V_j の直接的原因と解釈されるので、対応する因果ダイアグラムでは $\text{pa}(V_j)$ の要素それぞれから V_j へ矢線が引かれることになる。なお、Pearl 流の統計的因果推論は共分散構造分析 (Bollen, 1989) とは異なり、因果ダイアグラム上には錯乱項を描かないのが慣習となっている。

CCP データの場合、錯乱項どうしに関連がないと仮定した場合の状況を有向グラフを用いて表すと図 2.1(a) のようになり、錯乱項 ϵ_x と ϵ_y との間に関連があると仮定した場合は図 2.1(b) のようになる。なお、Kuroki and Cai (2008) および MacLehose et al. (2005) では、表 2.2 のデータが図 2.1(b) の因果ダイアグラムに基づいて生成されていると仮定した

上で、因果効果の存在範囲を求めている。さて、錯乱項どうしに関連がない状況を示した因果ダイアグラム(図 2.1 (a)) が与えられたとき、対応する構造的因果モデル式 (2.1) における同時分布の逐次的因数分解(式 (2.2)) は、

$$\text{pr}(x, y, z) = \text{pr}(y|x, z)\text{pr}(x|z)\text{pr}(z) \quad (2.6)$$

と表現することができる。なお、錯乱項どうしに関連がある状況を示した因果ダイアグラム(図 2.1 (b)) が与えられた場合にも式 (2.1) のような構造的因果モデルを与えることはできる。しかし、これに対応する同時分布の逐次的因数分解として式 (2.6) を考え、これが図 2.1(b) の因果構造を適切に表現していると判断することは難しい(錯乱項 ϵ_x と ϵ_y の間に關連があるかどうかを判断できない)。このような場合、 X と Y の共通原因と解釈される変数の集合 \mathbf{U} を導入して

$$\text{pr}(x, y, z, \mathbf{u}) = \text{pr}(y|x, z, \mathbf{u})\text{pr}(x|z, \mathbf{u})\text{pr}(z)\text{pr}(\mathbf{u})$$

と表現することが多い。これに対応する構造的因果モデルは

$$X = g_x(Z, \mathbf{u}, \epsilon'_x), Y = g_y(Z, X, \mathbf{u}, \epsilon'_y), Z = g_z(\epsilon_z)$$

である。ここに、錯乱項 $\epsilon_z, \epsilon'_x, \epsilon'_y$ の間に關連はないと仮定される。この構造的因果モデルに対応する因果ダイアグラムは図 2.1(c) で与えられる。

2.2.2 自律性と因果効果

さて、Pearl 流の統計的因果推論では、自律性という仮定に基づいて、外的操作 (Intervention) という概念が数学的に定義される。すなわち、ある変数 V_k に対して外的操作を行って $V_k = v_k$ とするという行為は、数学的には、式 (2.1) において V_k に関する構造方程

式を定数関数 $V_k = v_k$ に置き換えることを意味している. この外的操作は原子的 (Atomic) であると呼ばれ (Pearl, 1995, 2009a), $\text{set}(V_k = v_k)$ や $\text{do}(V_k = v_k)$ と表現される. ここに, 本章では, 原子的な外的操作に限定して議論を行うが, 実際にはこれに限定されることはなく, 興味ある問題に応じて, 適用状況にあわせてさまざまな外的操作を考えることが可能であることに注意する (黒木, 2008; Kuroki, 2012; Kuroki and Miyakawa, 2003; Murphy, 2003; Pearl, 2009a).

さて, V_k に対する原子的な外的操作によって, 式 (2.1) は

$$\left. \begin{aligned} V_j &= g_j(\text{pa}(V_j), \epsilon_j), & j = 1, 2, \dots, p; j \neq k, \\ V_k &= v_k \end{aligned} \right\} \quad (2.7)$$

という新たな構造的因果モデルへ変化し, これに対応して式 (2.2) の同時分布も

$$\text{pr}\{v_1, \dots, v_p | \text{set}(V_k = v_k)\} = \prod_{j=1, j \neq k}^p \text{pr}\{v_j | \text{pa}(v_j)\} = \frac{\text{pr}(v_1, \dots, v_p)}{\text{pr}\{v_k | \text{pa}(v_k)\}} \quad (2.8)$$

と変わることになる. $\text{pr}\{v_1, \dots, v_p | \text{set}(V_k = v_k)\}$ は式 (2.1) において V_k に関する構造方程式を定数関数 $V_k = v_k$ に置き換えたときの V_1, \dots, V_p の同時分布を意味する (ただし, V_k は定数 v_k になっている). 式 (2.8) は式 (2.2) において $\text{pr}\{v_k | \text{pa}(v_k)\}$ を 1 と置き換えたものと同じであり, 外的操作が成功し $\text{pa}(v_k)$ や錯乱項 ϵ_k によらずに $V_k = v_k$ となっていることを意味する. つまり, 式 (2.8) は母集団に含まれる対象者全員に対して $V_k = v_k$ という外的操作を行ったときの同時分布とみなすことができる. ここに, 自律性があるがゆえに, 外的操作を行わない変数に対する条件付き確率については, 外的操作を行う前の条件付き確率がそのまま使われていることに注意が必要である. また, この外的操作によって, 因果ダイアグラムも V_k に入る矢線をすべて取り除いたものにも変わることにも注意されたい. Pearl 流の統計的因果推論の目的の一つは, (外的操作を行う前の) 式 (2.1) によって生成されたデータあるいは式 (2.2) の同時分布に基づいて, (外的操作を行った後の) 仮想的に定

式化された式 (2.7) の構造的因果モデル, 式 (2.8) の同時分布, あるいは式 (2.8) から導かれる周辺分布に関する性質を明らかにすることとってよい. もちろん, 因果ダイアグラム上の変数がすべて観測されていれば, 式 (2.1) によって生成されたデータから式 (2.8) の同時分布を推定できる.

例として, 図 2.1(a) の因果ダイアグラムにおいて, β 遮断薬服用の有無 (X) を操作して対象者全員に対して強制的に β 遮断薬を服用させる ($X = x_1$) とする. この外的操作によって, 構造的因果モデルは

$$X = x_1, Y = g_y(Z, x_1, \epsilon_y), Z = g_z(\epsilon_z)$$

あるいは

$$X = x_1, Y = g_y\{g_z(\epsilon_z), x_1, \epsilon_y\}, Z = g_z(\epsilon_z) \quad (2.9)$$

と書き換えられ, 同時分布も

$$\text{pr}\{y, z | \text{set}(X = x_1)\} = \text{pr}(y | x_1, z) \text{pr}(z) = \frac{\text{pr}(x_1, y, z)}{\text{pr}(x_1 | z)}$$

となる. ここに, 先に説明した自律性という仮定があるがゆえに, この外的操作によって X に関する構造方程式以外の方程式の形が変化することはないことに注意する. また, 式 (2.9) の Y は, X には外的操作前の構造方程式の代わりに値 $X = x_1$ を代入しかつ X 以外の変数 (Z) についてはそれに対応する構造方程式を代入したものとなっている, すなわち, 外的操作を行う前の X に関する構造方程式に関する情報は使われていないことを確認しておこう. 錯乱項による構造方程式の表現とあわせて, このことが次節において構造的因果モデルと潜在反応モデルを結びつける役割を果たす. なお, この外的操作に対応する因果ダイアグラムは図 2.2 で与えられる. ここに, β 遮断薬服用の有無 (X) に対して外的操作を行った後では, もはや X は人種 (Z) の関数ではないので, X に入る矢線はすべて取り

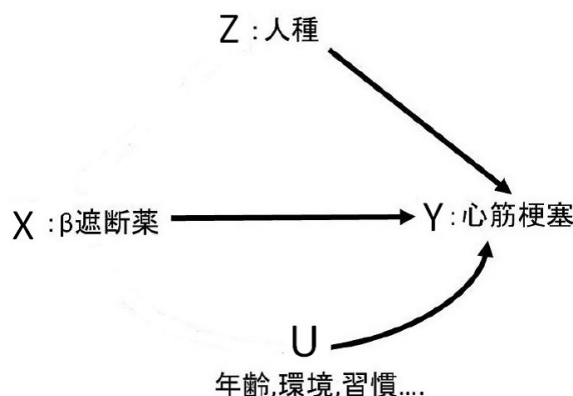


図 2.2: 因果ダイアグラム図 2.1(a)において, 外的操作により β 遮断薬 (X) を x_1 としたときの因果ダイアグラム. x_1 は定数なので, ダイアグラム上の X も x_1 とすべきかもしれないが, β 遮断薬を服用させない ($X = x_0$) とする外的操作を考えることも可能なので, あえて X のままにしてある.

除かれていることに注意が必要である.

以上の準備の下で, V_k に対する外的操作を行ったときの同時分布 (式 (2.8)) が与えられたとき, これから興味ある変数 V_l の周辺分布

$$\text{pr}\{v_l | \text{set}(V_k = v_k)\} = \sum_{v \setminus \{v_k, v_l\}} \frac{\text{pr}(v)}{\text{pr}\{v_k | \text{pa}(v_k)\}}$$

を考えることができ, この確率分布を V_k から V_l への因果効果と呼ぶ (Pearl, 1995, 2009a).

ここに, $\sum_{v \setminus \{v_k, v_l\}}$ は V_k と V_l を除く \mathbf{V} の要素について和をとることを意味する. この確率は, $\text{pr}(v)$ を “ $\text{pa}(v_k)$ を与えたときの V_k の条件付き確率 $\text{pr}\{v_k | \text{pa}(v_k)\}$ ” で割って計算したものであって, “ V_k の周辺分布 $\text{pr}(v_k)$ ” で割って計算した $V_k = v_k$ を与えたときの V_j の条件付き確率とは根本的に異なる.

一般に, 何らかの仮定をおかない限り, 上記の因果効果を推定することはできない. しかし, 幸いなことに, Pearl 流の統計的因果推論のフレームワークでは, 因果効果が識別可能であるための十分条件が数多く用意されている. その十分条件の一つを直観的にいうならば, 治療 (X) とその直接的な結果に対応する変数であり, かつ反応変数 (Y) に影響を

与える変数との間に交絡道(双方向矢線によって構成される道)がなければ因果効果は識別可能である(Tian and Pearl, 2002) というものである. 因果効果の識別可能性問題に関する詳細については黒木(2007a), Tian and Pearl(2002), Pearl(2009a)などを, 最近の発展についてはCai and Kuroki(2008), Huang and Valtorta(2006, 2008), Kuroki(2007), 黒木(2007b), 黒木・陳(2010), Kuroki and Miyakawa(1999), Kuroki and Pearl(2014), Shpitser and Pearl(2006a), Shpitser and Pearl(2006b)などを参照されたい.

図2.1(a)の例の場合, β 遮断薬を服用($X = x_1$)したときの対象者が死亡($Y = y_1$)する因果効果は

$$\text{pr}\{y_1|\text{set}(X = x_1)\} = \sum_z \text{pr}(y_1|x_1, z)\text{pr}(z) \quad (2.10)$$

と表現することができる. この結果を表2.2のCCPデータに適用することにより, X の各値に対する因果効果を求めることができ,

$$\text{pr}\{y_1|\text{set}(X = x_1)\} = 0.027236 \times 0.05941 + 0.030664 \times 0.94059 = 0.03046$$

$$\text{pr}\{y_1|\text{set}(X = x_0)\} = 0.22892 \times 0.05941 + 0.25912 \times 0.94059 = 0.25733$$

を得る. ちなみに, $X = x$ を与えたときの $Y = y_1$ の条件付き確率は, $\text{pr}(y_1|x_1) = 0.008995/0.29525 = 0.030466$, $\text{pr}(y_1|x_0) = 0.18133/0.70475 = 0.25730$ であることから, (この例ではわずかであるが)因果効果と条件付き確率が異なることが確認できる. また, 2.3.2節で定義する因果リスク差(Causal Risk Difference: x_1 と x_0 の因果効果の差)は, $\text{pr}\{y_1|\text{set}(X = x_1)\} - \text{pr}\{y_1|\text{set}(X = x_0)\} = 0.03046 - 0.25733 = -0.22687$ と求まり, β 遮断薬の服用により心筋梗塞後の死亡割合が減少したと解釈できる.

2.3 構造的因果モデルと潜在反応モデルの関係

本節では、Rubin 流の潜在反応モデルと Pearl 流の潜在反応モデルの区別を意識しながら、Pearl 流の潜在反応モデルを述べる。

2.3.1 Pearl 流の潜在反応モデル

簡単のために X を 2 種類の治療を表す変数とし、対象者 i が治療 $X = x_1$ を受けた場合に生じるであろう潜在的な反応を表す変数 (潜在反応変数 [Potential Response Variable, Potential Outcome Variable]) を $Y_{x_1}(i)$ 、治療 $X = x_0$ を受けた場合に生じるであろう潜在反応変数を $Y_{x_0}(i)$ と記す ($X, Y \in \mathbf{V}$)。また、本来は X も対象者 i に依存するので $X(i)$ と記すべきであるが、このときの i はしばしば省略される。このとき、 $Y_{x_1}(i) - Y_{x_0}(i)$ を対象者レベルの因果効果 (Unit-Level Causal Effect (Rubin, 2005)) と呼ぶ。

ここで、Pearl 流の潜在反応モデルと Rubin 流の潜在反応モデルでは、“対象者”の定義が異なることに注意しなければならない。Rubin 流の潜在反応モデルにおける対象者とは、個々の対象者を直接的に特徴づける属性からなるもの (たとえば、“黒木”本人や“小林”本人を完全に識別する情報すべて) であり、その対象者を取り巻く環境といった対象者を間接的に特徴づける情報は含まれない。これに対して、Pearl 流の潜在反応モデルにおける対象者には、個々の対象者を直接的に特徴づける属性だけでなく、対象者 i に対して治療を行ったときに生じる反応を決定論的に規定するあらゆるゆる要因が含まれている。このことは、Rubin 流の統計的因果推論の基本的フレームワークに基づいて因果効果の推測問題を扱う際には SUTVA (Stable Unit Treatment Value Assumption (Rubin, 1980)) と呼ばれる仮定を必要とするが、Pearl 流の統計的因果推論では必ずしもこの仮定を用いる必要がないという違いに結びつく。SUTVA は、No Interference between Subjects Assumption (Cox, 1958) と No Multiple Versions of Treatment Assumption (Neyman, 1935) からなる (Rubin, 1986)。前者は“対象者それぞれが治療を受けたときの潜在的な反応は、他の対象者が受ける治療

に依存しない”という仮定を，後者は“対象者が治療を受けたときの反応は，その治療をどのように受けたのかには依存しない”という仮定を意味する (VanderWeele and Hernán, 2011).

まず，**No Interference between Subjects Assumption** について，たとえば，対象者 i についてあるワクチンの効果を調べる場合，それを接種することによって対象者 i が感染症にかかりにくくなるかどうかは，対象者 i との接触が多い対象者 j がすでにワクチンを接種しているかどうかにも依存する．すなわち，対象者 j がワクチンを接種していた場合には対象者 j 自身が感染症にかかりにくくなるため，対象者 i が感染症にかかる可能性も小さくなるが，対象者 j がワクチンを接種していない場合には，対象者 i も対象者 j も感染症にかかりやすくなると考えられる．このことは，感染症への暴露状況が大きく異なると，一般にはワクチンの効果の大きさも異なるため，対象者 i に対するワクチンの効果は対象者 j がワクチンを接種したかどうかにも依存することを意味する．このような場合には，対象者 j に対してワクチン接種が行われた場合と行われていない場合のそれぞれについて，対象者 i に対するワクチン接種の反応を考えなくてはならない．**No Interference between Subjects Assumption** はこのような対象者どうしの依存関係が存在しないとする仮定である．

一方，**No Multiple Versions of Treatment Assumption** について，たとえば，対象者 i に対してある手術を行う場合，それが成功するかどうかは病院の設備や医師の経験や技術などの，その対象者を取り巻く環境にも依存する．すなわち，医師 B と比べて医師 A のほうがその手術に対する経験を豊富にもっていれば，(同じ環境であれば) 医師 B が対象者 i の手術を行うよりも医師 A が行うほうが成功する可能性が高くなると考えられる．しかし，医師 B の勤務する病院ではその手術を行うのに適切な設備がそろっているのに対して医師 A の勤務する病院の設備・環境が悪ければ，医師 B が手術を行えば成功するものの，医師 A が行った場合には，その経験・技術を十分に生かしきれずに失敗に終わることもありうるであろう．**No Multiple Versions of Treatment Assumption** とは，どの病院でどの医師がそ

の手術をしても対象者 i は同じ反応を示すことを意味している³。Rubin 流の潜在反応モデルで定義される対象者には、対象者 i を取り巻く環境、たとえば、その治療を誰が行ったのか、どういった条件の下でその治療が行われたのかなどといった情報は含まれていないため、SUTVA を仮定しない場合には、対象者 i が治療を受けたときの反応を決定論的に規定することは難しく、したがって、対象者レベルの因果効果も適切に定義することは難しい (Rubin 流の潜在反応モデルにおける対象者レベルの因果効果では、このような付帯状況が一定であることが暗黙に仮定されている)。これに対して、Pearl 流の潜在反応モデルで扱われる対象者 i には、治療 X が行われたときに対象者 i に対して生じる反応を決定づけるあらゆる要因が含まれているため、因果効果を評価するのに **No Interference between Subjects Assumption** は必要とするが、**No Multiple Versions of Treatment Assumption** は必要としない。また、Pearl 流の統計的因果推論では、潜在反応変数そのものは対象者 i の属性 (変数) と治療 X の値によって決定され、治療 X の値も対象者 i の属性によって決定される。

さて、上述の考え方を構造的因果モデルのフレームワークを用いて表現した場合、対象者 i の属性は変数集合 V の要素それぞれがとる値と錯乱項 $\epsilon_1, \dots, \epsilon_p$ のそれぞれがとる値の集まりによって規定されると考えてよい。したがって、2.2.1 節で述べたように、 V の要素それぞれは錯乱項の関数として表現できることから、対象者 i を規定する錯乱項がとる値を $\epsilon_1^i, \dots, \epsilon_p^i$ とおくと、結局のところ、潜在反応変数も錯乱項全体からなる関数の値、すなわち、 $Y_x(i) = Y(x, \epsilon_1^i, \dots, \epsilon_p^i)$ と記述できる。ここに、 X の値を定数 x としているので X に付随する錯乱項 ϵ_x や X に関する構造方程式を経由してのみ $Y_x(i)$ に現れるような錯乱項はこの関数には含まれていないことに注意する。これに対し、実際に観測される反応変数は $Y(i) = Y(\epsilon_x^i, \epsilon_1^i, \dots, \epsilon_p^i)$ と表現され、 X の部分にも構造方程式 $X = X(i)$ が代入されている。したがって、 $Y(i)$ は対象者 i について Y の直接的な原因や間接的な原因とみなされる変数に付随する錯乱項、すなわち、あるがままの対象者 i に関する情報がすべて含ま

³この仮定は、処置が複数選択ではなくユニークである、と説明されることもある (田中, 2014)。

れていることになる。この考え方は、後述する構造的因果モデルにおける一貫性 (Robins, 1986; Pearl, 2009a) の解釈に結びつく。

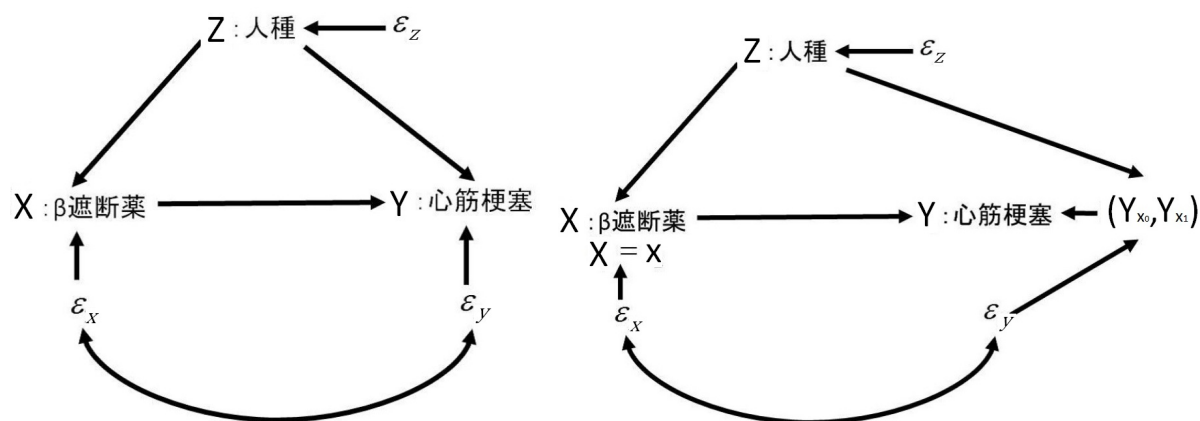
さて、対象者 i が $X = x_1$ を受けた場合には $Y_{x_1}(i)$ が観測されるがそのとき同時に $Y_{x_0}(i)$ が観測されることはなく、 $X = x_0$ を受けた場合には $Y_{x_0}(i)$ が観測されるがそのとき同時に $Y_{x_1}(i)$ が観測されることはない。なぜならば、実際に対象者 i が $X = x_1$ を受けたときには $x_1 = X(\epsilon_x^i, \epsilon_1^i, \dots, \epsilon_p^i)$ ⁴ ということになるわけであるが、これは対象者 i を規定した情報 $\epsilon_x^i, \epsilon_1^i, \dots, \epsilon_p^i$ から生成された X の値は x_1 であって x_0 ではないことを示しているからである。したがって、対象者 i について $X(i) = X(\epsilon_x^i, \epsilon_1^i, \dots, \epsilon_p^i) = x_1$ が実際に観測されたときには、 $Y_{x_1}(i)$ の値は $\epsilon_x^i, \epsilon_1^i, \dots, \epsilon_p^i$ とは矛盾することなく適切に定義され、 $Y_{x_1}(i)$ が観測されることになる。なお、ここで観測された値は、対象者 i についてあるがままに観測される反応 $Y(i)$ そのものである。一方、仮想的な状況である $Y_{x_0}(i) = Y(x_0, \epsilon_1^i, \dots, \epsilon_p^i)$ については、 $X(i) = x_1$ を代入することができないため (x_0 を対象者 i の属性によって表現できない)、 $Y_{x_0}(i)$ を対象者 i の属性によって適切に定式化することができない、すなわち、観測されないことになる。このように、Pearl 流の統計的因果推論のフレームワークを用いれば、潜在反応変数を $Y_x(i) = Y(x, \epsilon_1^i, \dots, \epsilon_p^i)$ と表現した上で、 $Y_x(i)$ の x に $X = x$ を代入することによって一貫性 (Consistency) “ $X = x \implies Y_x(i) = Y(i)$ ” を性質 (定理) として導くことができる (Galles and Pearl, 1997, 1998; Pearl, 2009a)。上述したように、 $Y(i)$ は対象者 i についてあるがままに観測される反応である。

一貫性は、欠測データ解析を基調とする Rubin 流の統計的因果推論では潜在反応変数と観測変数を結びつける仮定⁵として用いられるのに対して、ノンパラメトリック構造方程式モデルを基調とする Pearl 流の統計的因果推論ではデータ生成過程から導かれる性質 (定理) であることに注意する (Cole and Frangakis, 2009; VanderWeele, 2009a; Pearl, 2009a, 2010a)。すなわち、Rubin 流の統計的因果推論では、潜在反応変数は、一貫性を仮定した

⁴ X の結果系に付随する錯乱項は含まれないことに注意する。

⁵一貫性は No Multiple Versions of Treatment Assumption に含まれるが (VanderWeele and Hernán, 2011)、しばしばこれとは独立して仮定されることもあるため、本章でも後者の立場から議論する。

2.3. 構造的因果モデルと潜在反応モデルの関係



(a) 図 2.1(a) に錯乱項を加えたもの (b) 図 2.1(a) の潜在反応モデルのグラフ表現

図 2.3: {人種, β 遮断薬服用の有無, 急性心筋梗塞} の間の因果関係を表現した因果ダイアグラムに対する潜在反応モデル. (a) は図 2.1(a) の因果構造を錯乱項を加えて詳細に表現したものであり, (b) は潜在反応変数を加えて表現したものである.

うえて, 対象者 i が治療を受けたときに観測される反応として定義される (すなわち, 観測された反応変数 $Y(i)$ に基づいて潜在反応変数 $Y_x(i)$ を定義するために一致性が使われる) のに対して, Pearl 流の統計的因果推論では, 対象者 i が治療を受けたときに観測される反応は対象者 i の属性を規定する潜在的な反応 (錯乱項の関数) をとおして生成されたものと解釈される (Pearl, 2009a, 2010b). したがって, Rubin 流の統計的因果推論では一致性という仮定をどのように検証するのか, そして一致性が成り立たないときの潜在反応変数とはどのようなものかを明らかにすることが問題となるが, Pearl 流の統計的因果推論では一致性そのものを検証する必要はなく, むしろ式 (2.1) で表わされる “自律的なデータ生成過程” をどのように検証するかが問題となる. ただし, SUTVA や一致性を仮定するかどうかは統計的因果推論を議論する際の出発点の違いであって, これらに基づいて議論する限りにおいてはどちらのアプローチを用いても同様の結果が得られる (Pearl, 2009c).

ここで, 構造的因果モデルの観点から潜在反応変数のしくみをもう少し詳しく検討するために, 図 2.3 の因果ダイアグラムを考える. 図 2.3(a) は図 2.1(a) に錯乱項を加えたものであり, これに対応する構造的因果モデルは式 (2.5) あるいは式 (2.6) によって表現され

る。一方、 β 遮断薬を強制的に服用 ($X = x_1$) させたときの対象者の状態 (Y) に関する構造方程式は $Y = g_y\{g_z(\epsilon_z), x_1, \epsilon_y\}$ と表現することができる。そして、錯乱項に対象者 i が持つ値 $\epsilon_y^i, \epsilon_z^i$ を代入すれば対象者 i に β 遮断薬を服用 ($X = x_1$) させたときの対象者 i の状態 (Y) に関する潜在反応, すなわち, $Y = g_y\{g_z(\epsilon_z^i), x_1, \epsilon_y^i\}$ となる。このことは構造的因果モデルによって潜在反応モデルを特徴づけることができることを意味する。また, 対象者 i が β 遮断薬を服用 ($X = x_1$) していれば, $h_x(\epsilon_x^i, \epsilon_z^i) = x_1$ を対象者の状態 (Y) に関する構造方程式に代入することによって $Y = g_y\{g_z(\epsilon_z^i), h_x(\epsilon_x^i, \epsilon_z^i), \epsilon_y^i\} = h_y(\epsilon_x^i, \epsilon_y^i, \epsilon_z^i)$ となることから, 対象者 i が急性心筋梗塞によって死亡するかどうかに対象者 i を規定する錯乱項の値により一意に決定され, 観測されることとなる。これが一致性のメカニズムである。逆に, 実際とは異なり, 対象者 i が β 遮断薬を服用しなかった ($X = x_0$) 場合を考えると, $x_1 = h_x(\epsilon_x^i, \epsilon_z^i)$ を構造方程式 $Y = g_y\{g_z(\epsilon_z^i), x_0, \epsilon_y^i\}$ に代入することはできないため (x_0 を対象者 i の属性によって表現できない), 対象者 i が急性心筋梗塞によって死亡するかどうかは観測されないことになる。

この考察を因果ダイアグラム上に潜在反応変数を導入することによって表現したものを図 2.3(b) に与える。図 2.3(b) においては, 錯乱項 ϵ_y と人種 (Z) の両方から潜在反応変数 (Y_{x_0}, Y_{x_1}) へ矢線が引かれており, ϵ_x からは矢線が引かれていない。人種 (Z) と錯乱項 ϵ_y の両方から潜在反応変数 (Y_{x_0}, Y_{x_1}) へ矢線が引かれているのは潜在反応変数が錯乱項 ϵ_y と ϵ_z の関数でかつ人種 (Z) が ϵ_z に関する情報をすべて有していることが理由であり, ϵ_x からは矢線が引かれていないのは ϵ_x に関する情報をすべて有している X が (Y_{x_0}, Y_{x_1}) では定数として扱われているためである。

2.3.2 因果効果

ここで, 2.2.2 節で説明した因果効果について, Pearl 流の潜在反応モデルの観点からもう一度議論する。

前節で述べたように、一般に、対象者 i が $X = x_1$ あるいは $X = x_0$ のいずれか一方の治療を受けた時点で、その治療を受ける前の(まっさらな)対象者 i は存在しない。すなわち、特別な仮定をおかない限り対象者レベルの因果効果 $Y_{x_1}(i) - Y_{x_0}(i)$ を推定することはできない。しかし、たとえば、母集団レベルでの因果効果を評価対象として、対象者が無作為抽出されているような状況を考えた場合、 $Y_{x_1}(i)$ と $Y_{x_0}(i)$ はそれぞれ確率変数 Y_{x_1} と Y_{x_0} がとりうる値を表した確率変数とみなすことができる。そこで、本章では、 $Y_{x_k} = y_j$ である潜在反応変数の確率を $\text{pr}(Y_{x_k} = y_j) = \text{pr}(y_{j,x_k})$ とおく ($j, k = 0, 1$)。Lewis (1973) にならってこの確率が反事実確率と呼ばれることがあるが (Greenland et al., 1999; Dawid, 2000), $\text{pr}(y_{j,x_k})$ そのものは現実に観測された状態については関心がない、あるいは実際に治療を受けた場合にはこれらの変数の値が観測される(まったく観測されないものとの誤解を生む)といった理由から、少なくとも“反事実”という言葉は適切ではないという指摘がなされている (Rubin, 2005; Pearl, 2009a)。ここで、 $\text{pr}(y_{j,x_k})$ は、(i) X の値を観測しているわけではないため、直接的に推定することはできないこと、そして、(ii) 前節の議論からわかるように、 $\text{pr}(y_{j,x_k})$ は X の構造方程式に基づく情報を使うことなく $Y(x, \epsilon_1, \dots, \epsilon_p)$ における x に値を定数 x_k としたときの確率であるから、 $\text{pr}\{y_j | \text{set}(X_k = x_k)\}$ と同義であることに注意が必要である。本章では、二つの治療 ($X = x_0, x_1$) の因果効果を比較した $\text{pr}(y_{1,x_1}) - \text{pr}(y_{1,x_0})$ を因果リスク差と呼ぶことにする。この因果リスク差は変数 X の値を外的操作によって強制的に x_0 から x_1 に変化させたときの Y の平均的变化と解釈されることが多い。このような解釈が行われるのは、 $\text{pr}(y_{j,x_k})$ は $\text{pr}\{y_j | \text{set}(X_k = x_k)\}$ と同義であり、後者は“外的操作”という実質科学的に解釈可能でかつ物理的な意味を有していることが理由にあると推察される。なお、Pearl 流の統計的因果推論では、因果効果を考える際、変数 X に対して実際に外的操作が可能であるかどうかは問わないことに注意する (Pearl, 2009a, 2010b)。すなわち、Rubin 流の統計的因果推論では、“No Causation without Manipulation” (Holland, 1986) とあるように、操作可能性(実際に外的操作を行うことが可能であるかどうか)が変数間の因果関係を議論する意味があるかどうかを判断する重要な

要素となるが、Pearl 流の統計的因果推論では外的操作は因果関係を議論する科学的方法の一つであって、それによって因果推論に関する議論が制限されるものであってはならないという立場がとられている (Pearl, 2009a, 2010b). 実際、潜在反応変数の国語表現として用いられる (上述のとおり、潜在反応変数の国語表現としては厳密には適切な解釈とは言えないが) 反事実的表現“事象 A が起きれば事象 B は起きたであろう”は事実とは反対の仮定、想像あるいは実現不可能な願望を表現したものであることから、操作可能性を超えた意味を持っていることは明らかである。

ここで、 X に対する無作為化割りつけが適切に行われ、対象者がその割りつけにしたがう状況を考える。このとき、 X はそれ自身に直接付随する錯乱項 ϵ_x やそれに先立つ変数に依存せず、しかも、 (Y_{x_1}, Y_{x_0}) は ϵ_x 以外のすべての錯乱項の関数として表現できることから、 (Y_{x_1}, Y_{x_0}) と X は独立となり、さらに一致性より、因果リスク差は $\text{pr}(y_{1,x_1}|x_1) - \text{pr}(y_{1,x_0}|x_0) = \text{pr}(y_1|x_1) - \text{pr}(y_1|x_0)$ により推定できることがわかる。一方、観察研究においても、治療の割りつけが SITA 条件を満たす場合には因果効果は識別可能となる。すなわち、治療 X について、 X と (Y_{x_1}, Y_{x_0}) を条件付き独立にするような変数集合 \mathbf{T} が存在するとき、 \mathbf{T} を与えたときに治療の割りつけは強い意味で無視可能である (Rosenbaum and Rubin, 1983), あるいは \mathbf{T} は (X, Y) について SITA (Strongly Ignorable Treatment Assignment) 条件を満たすという (Pearl, 1994; Edwards, 2000). このとき、SITA 条件を満たす \mathbf{T} を観測できれば因果リスク差は識別可能であり

$$E_{\mathbf{t}}\{\text{pr}(y_1|x_1, \mathbf{t}) - \text{pr}(y_1|x_0, \mathbf{t})\} = \sum_{\mathbf{t}} \{\text{pr}(y_1|x_1, \mathbf{t}) - \text{pr}(y_1|x_0, \mathbf{t})\} \text{pr}(\mathbf{t})$$

で与えられる。なお、2.2.2 節ではグラフィカルモデルの観点から因果効果の識別可能条件を簡単に紹介したが、これらの識別可能条件は潜在反応モデルにおいても適用可能であることに注意する。逆に、潜在反応モデルに基づく因果効果の識別可能条件のうち、グラフィカルモデルでは表現できないものが存在する。これは、構造的因果モデルに理論基盤

をおく潜在反応アプローチでは構造方程式モデルによって規定される対象者レベルの因果的仮定や条件付き独立関係がそのまま保存されるのに対して、グラフィカルモデルはその条件付き独立関係の一部しか表現できないこと、そして、グラフィカルモデルでは対象者レベルの因果的仮定を表現できないことが理由にある (Pearl, 2009a). 次節では、後者について議論する.

2.4 グラフィカルモデルに基づく統計的因果推論の有用性

2.4.1 グラフィカルモデルの問題点

Pearl 流の統計的因果推論はしばしばグラフィカルモデルと同義とみなされ、解析を行うのに十分な因果構造を有向グラフによって記述することは難しいことが欠点として指摘されている. しかし、前節の議論より、有向グラフによる因果構造の記述は因果効果を推測するにあたって対象となる因果構造の理解を助けているだけ (“だけ”とはいっても、これが非常に重要である) であって、“Pearl 流の統計的因果推論 = グラフィカルモデル”ではないことがわかる. 一方、グラフィカルモデルを用いて因果効果を推測する際の問題点の一つとして、対象者レベルの因果的な仮定をグラフで表現することは困難であることがあげられる.

その例として、図 2.4 の因果ダイアグラムにおいて、 X と Y がともに 2 値の確率変数 ($x \in \{x_0, x_1\}, y \in \{y_0, y_1\}$) であり、共通原因 U を非観測変数とするとき、 X から Y への因果効果 $\text{pr}(y_{1,x_1})$ を評価することを考える.

この状況では、 U が非観測変数であるため $\text{pr}(y_{1,x_1})$ を推定することはできない. しかし、全確率の公式と一致性を使って、

$$\text{pr}(y_{1,x_1}) = \text{pr}(x_1, y_1) + \text{pr}(y_{1,x_1} | x_0, y_1) \text{pr}(x_0, y_1) + \text{pr}(y_{1,x_1} | x_0, y_0) \text{pr}(x_0, y_0)$$

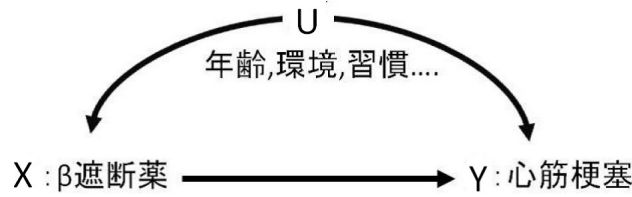


図 2.4: $\{\beta$ 遮断薬服用の有無, 急性心筋梗塞 $\}$ の間の因果関係を表現した因果ダイアグラム.

と変形できる. したがって, 推定不能な反事実確率 $\text{pr}(y_{1,x_1}|x_0, y_1)$ と $\text{pr}(y_{1,x_1}|x_0, y_0)$ について, $\text{pr}(y_{1,x_1}|x_0, y_1) = \text{pr}(y_{1,x_1}|x_0, y_0) = 1$ とおくことにより $\text{pr}(y_{1,x_1})$ の上限を, また $\text{pr}(y_{1,x_1}|x_0, y_1) = \text{pr}(y_{1,x_1}|x_0, y_0) = 0$ とおくことにより $\text{pr}(y_{1,x_1})$ の下限を求めることができ, これらより $\text{pr}(y_{1,x_1})$ の存在範囲として

$$\text{pr}(x_1, y_1) \leq \text{pr}(y_{1,x_1}) \leq 1 - \text{pr}(x_1, y_0) \quad (2.11)$$

を得ることができる (反事実確率 $\text{pr}(y_{1,x_1}|x_0, y_1)$ と $\text{pr}(y_{1,x_1}|x_0, y_0)$ を感度パラメータとみなすことにより感度分析を行うことも可能であるが, 本章ではこれについては扱わない). この存在範囲は, U が SITA 条件を満たしていることに注意すれば

$$\text{pr}(y_{1,x_1}) - \text{pr}(x_1, y_1) = \sum_u \text{pr}(y_1|x_1, u) \{1 - \text{pr}(x_1|u)\} \text{pr}(u) \geq 0$$

$$\begin{aligned} & 1 - \text{pr}(x_1, y_0) - \sum_u \text{pr}(y_1|x_1, u) \text{pr}(u) \\ &= 1 - \text{pr}(x_1) - \sum_u \text{pr}(y_1|x_1, u) \text{pr}(x_0|u) \text{pr}(u) \geq 0 \end{aligned}$$

となり, やはり式 (2.11) を得ることができる. さて, 因果効果の存在範囲を評価する際の大きな問題点の一つとして, (グラフから読み取れる限りの) 極小な因果仮定の下で求めた存在範囲は広くなりすぎることが指摘されている. この問題を解決する一つの方法

として、単調性、すなわち、 $\text{pr}(y_{1,x_1}, y_{0,x_0}) = 0$ という仮定が用いられる。この因果仮定は、 β 遮断薬を服用した対象者は急性心筋梗塞を引き起こして死亡し、 β 遮断薬を服用しない対象者が生存するという“あまのじゃく”な現象は起こらないことを意味している（もちろん、より狭い存在範囲を求めるのに必ずしもこの仮定を用いなければならないというわけではない (Pearl, 2009a)）。ここに、単調性は対象者レベルの仮定であり、条件付き独立関係を表現したグラフィカルモデルで記述することはできないことに注意する。単調性の仮定の下では $\text{pr}(y_{1,x_1}|x_0, y_0) = 0$ であるから、因果効果 $\text{pr}(y_{1,x_1})$ の存在範囲は

$$\text{pr}(x_1, y_1) \leq \text{pr}(y_{1,x_1}) \leq \text{pr}(y_1)$$

となり、式 (2.11) よりも狭い（厳密には、広くない）存在範囲を得ることができる。ここに、因果効果の存在範囲を定式化しておくことは、識別困難な状況において因果効果の定量的評価を行うのに有用であるだけでなく、因果的仮定を検証する上で大きな意味を持つ場合があることに注意が必要である。たとえば、上記の存在範囲では起こることはないが、より複雑な因果構造において導かれた存在範囲の公式にデータを代入すると、その上限の値と下限の値の大小関係が入れ替わることがある。このことは、採取されたデータと因果的仮説とが整合しないことを意味しており、この性質を用いることにより因果的仮説の検討を行うことが可能となる (Cai et al., 2008; Pearl, 2009a)。また、単調性に加えて $\text{pr}(y_{1,x_1}|x_0, y_1) = 0$ を仮定してよいのであれば、表 2.3 の観測変数の確率と潜在反応変数の確率との対応関係を参照することにより、因果効果 $\text{pr}(y_{1,x_1})$ は式 (2.11) の下限を用いて推定できることがわかる (Cai and Kuroki (2007) はノンコンプライアンスデータを用いた治療効果の評価問題のフレームワークにおいて、これと同様な議論を行っている)。このことから、条件付き独立関係の視覚的表現であるグラフィカルモデルに基づく統計的因果推論のフレームワークを用いた場合には、対象者レベルの因果仮定に基づいた結果を得ることが難しくなるケースがあることがわかる。

表 2.3: 観測確率と潜在反応変数の確率との対応関係

	x_0	x_1
y_0	$\text{pr}(x_0, y_0)$ $= \text{pr}(y_{1,x_1}, y_{0,x_0}, x_0)$ $+ \text{pr}(y_{0,x_1}, y_{0,x_0}, x_0)$	$\text{pr}(x_1, y_0)$ $= \text{pr}(y_{0,x_1}, y_{1,x_0}, x_1)$ $+ \text{pr}(y_{0,x_1}, y_{0,x_0}, x_1)$
y_1	$\text{pr}(x_0, y_1)$ $= \text{pr}(y_{1,x_1}, y_{1,x_0}, x_0)$ $+ \text{pr}(y_{0,x_1}, y_{1,x_0}, x_0)$	$\text{pr}(x_1, y_1)$ $= \text{pr}(y_{1,x_1}, y_{1,x_0}, x_1)$ $+ \text{pr}(y_{1,x_1}, y_{0,x_0}, x_1)$

2.4.2 共変量選択問題

前節では、グラフィカルモデルに基づく統計的因果推論の問題点を指摘した。しかし、グラフィカルモデルを用いることで、データを用いて因果関係を解明する際にどのような変数を観測すべきであるかを明らかにしたり、議論の対象となっている因果仮説のどこが正しく、どこが誤っているのか(あるいは因果仮説のどこの部分を詳細に検討すべきなのか)といった検討作業を容易にする特徴は非常に有用である。また、対象者レベルの因果仮定をどのように導入するかという指針を与えたり、対象者が治療を受けたのちどのような経過を得て反応変数に影響を与えているかといった因果メカニズムを視覚的に表現できることは、前節で述べた限界を補って余りあるほどの恩恵を我々に与える。しかも、グラフィカルモデルに基づく統計的因果推論で得られた結果はそのまま構造的因果モデルのフレームワークにおいても成り立つ (Pearl, 2009a) ことは、グラフィカルモデルにさらなる魅力を付加する。

グラフィカルモデルに基づく統計的因果推論の記述能力の限界とは異なる理由により、Pearl 流の統計的因果推論の利用に対して消極的な研究者や実務家も多い。その理由の多くは、実質科学においては変数間の因果関係を明確にすること、あるいは、有向グラフで記述することは困難であるというものである。因果効果を推定する際にこの立場をとる研究者や実務家がとっている変数選択の指針として、Paul Rosenbaum の著書 *Observational Study* にある “there is no reason to avoid adjustment for a variable describing subjects before

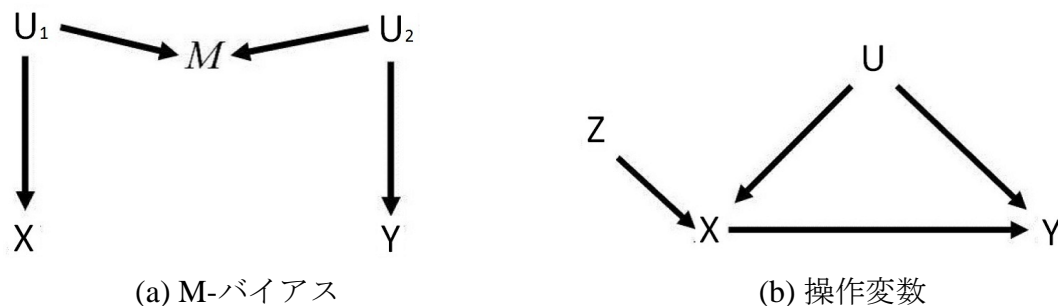


図 2.5: 因果構造を軽視することでバイアスを生じかねない状況

treatment”(Rosenbaum, 2002)が適用されていることが多い。これを文脈どおりに受け取るならば、治療に先立って観測される共変量で調整してはならないという理由はないということであるが、この指針を用いることで因果効果の推定に偏りが生じる場合があることが指摘されている (Pearl, 2009b,c; Shrier, 2008, 2009). その例として、図 2.5(a) の因果ダイアグラムを考える。

図 2.5(a) は、統計的因果推論の分野ではよく知られている M-バイアスを表現したものであり、共変量調整を行わなければ因果効果をバイアスなく推定できるのにもかかわらず、M を解析に加えることで因果効果の推定にバイアスが生じる状況である。すなわち、図 2.5(a) において、 U_1 と U_2 が観測されず、M が X に先だって観測されたとしても、因果リスク差は $\text{pr}(y|x_1) - \text{pr}(y|x_0)$ で与えられ、M を解析に加えてはならない。しかし、Rosenbaum の記述にしたがえば、状況によっては $\sum_m (\text{pr}(y|x_1, m) - \text{pr}(y|x_0, m)) \text{pr}(m)$ により評価しても構わないということになりかねず、その場合には因果効果の評価にバイアスが生じることになる (Pearl, 2009b,c; Rubin, 2008; Shrier, 2008, 2009). もちろん、X に先立って生成される変数(たとえば、 U_1 と M)を“すべて”観測した上でそれらを解析に取り入れるのであれば因果効果を偏りなく推定することができるが、すべての共変量を観測するのはほとんど不可能である場合もあり、因果効果を評価するのに十分な因果構造をグラフで記述するよりも困難な作業である場合もありうる。M-バイアス問題が与えるもう一つの教

訓として、実際のデータ解析において共変量選択を行う際にしばしば参考とされている指針 “strong ignorability requires that all variables affecting treatment assignment and response be measured”(Berk, 1987) に対する反例を与えていることにも注意が必要である。これは、SITA 条件を満たすためには治療と反応の両方に影響を与える共変量をすべて観測しなければならないことを意味している。しかし、たとえば、図 2.5(a) の M は X と Y の両方に影響を与えておらず、これを解析に加えることでバイアスを引き起こすが、 $\{U_1, M\}$, $\{U_2, M\}$, $\{U_1, U_2, M\}$ のいずれかの組み合わせが観測できれば因果効果はバイアスなく推定できる(もちろん、空集合(いずれの共変量も観測しない状況)でも因果効果をバイアスなく推定できる)。このとき、 U_1, U_2, M はいずれも高々 X または Y に対してだけ影響を与えているにすぎず、両方に影響を与えているわけではない。なお、図 2.5(a) は非常に単純な例であるが、M-バイアスに関する(仮想的ではあるが)より複雑なグラフィカル表現が Shrier and Platt (2008) によって検討されている。

一方、図 2.5(b) はデータ生成過程として線形構造方程式モデルを仮定したとき、 Z を操作変数 (Instrumental Variable) として用いれば因果効果をバイアスなく推定できるが、共変量とみなして解析に取り入れた場合には因果効果の推定にバイアスが生じる、しかも、 X から Y への単回帰係数で推定するよりも大きなバイアスが生じる状況である (Myers et al., 2011a,b; Pearl, 2010c, 2011)。ここに、 Z が (X, Y) について、(i) Z は治療 X と関連がある、(ii) Z は交絡因子と独立である、(iii) Z は X をとおして Y に影響を与える、をみたすとき、 Z を (X, Y) に関する操作変数という。この場合においても、Rosenbaum (2002) の記述にしたがえば、共変量として Z を解析に加えた重回帰分析を行っても構わないということになりかねない。しかし、因果ダイアグラムを描きさえすれば Z が操作変数であることが容易に確認することができるため、こういった問題は生じることはないであろう(この議論は、 Z が操作変数とは認識されなかったという意味で、因果構造があいまいであるゆえに生じた議論であるが、Pearl (2010c) はこの議論をとおして上述した Rosenbaum (2002) の記述の問題点を指摘している)。

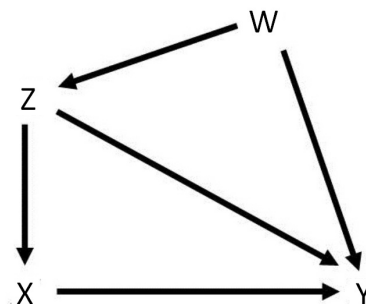


図 2.6: バックドア基準

最後に、SITA 条件に基づいて因果効果を推測する際の問題点について考えよう。SITA 条件に基づく方法論・応用研究では、“strong ignorability requires measurement of all covariates related to both treatment and outcome”という指針にしたがって、共変量がすべて観測されていると仮定することがある。しかし、SITA 条件は、その定義からわかるように、潜在反応変数と治療との(条件付き)独立性を要求しているだけであって、すべての共変量を観測しなければならないというわけではない(すべての共変量を観測することは困難であろう)。これに対して、グラフィカルモデルに基づいた統計的因果推論では、代表的な因果効果の識別可能条件としてバックドア基準 (Back Door Criterion; Pearl (1995, 2009a)) と呼ばれるものが用意されている。厳密なバックドア基準の定義は有向分離と呼ばれる概念を使うため、本章ではその詳細を解説することはしないが、直観的には、因果ダイアグラムより変数 X から出る矢線を取り除いたグラフにおいて X と Y の間の道のそれぞれに共変量集合 \mathbf{T} の要素がするとき、 (X, Y) について \mathbf{T} はバックドア基準を満たすといわれる。 \mathbf{T} が (X, Y) についてバックドア基準を満たすとき、 X から Y への因果効果は識別可能であり、

$$\text{pr}\{y|\text{set}(X = x)\} = \sum_t \text{pr}(y|x, t)\text{pr}(t)$$

で与えられる。この式は本質的に SITA 条件から導いたものと同じであるが、上述の経験的な指針とは異なり、すべての共変量を観測することを要求しているわけではない。たと

例えば、図 2.6 において、上述の指針に基づいて共変量を選択した場合には $\{W, Z\}$ となるが、バックドア基準に基づいて共変量を選択するならば Z だけを観測すれば十分であることがわかる。仮に、上述の指針に基づいて交絡因子を同定する場合、不必要に多くの共変量を観測する可能性があるため、因果効果の推測精度の低下や観測コストの増加といった問題を引き起こすこともありうる。

これらの例からわかるように、治療に先だって観測された共変量を安直に解析に取り入れることは、解析結果にバイアスを引き起こすことになりかねない。また、この問題は傾向スコア (Propensity Score) を使ったからと言って回避できるわけではない(しばしば、SITA 条件を検討することなしに傾向スコア (Rosenbaum and Rubin, 1983; Guo and Fraser, 2009) を使って因果効果の推定を行っているケースが見られるが、SITA 条件を満たしていない状況で傾向スコアを導入しても、(漸近的な意味で) 同じ大きさのバイアスを引き起こすだけである)。加えて、因果関係を有向グラフで記述しない場合には、因果効果を推定するのに十分な共変量を選択することが難しくなる可能性もある。これに対して、グラフィカルモデルに基づく統計的因果推論は視覚的にこのような誤った判断を避けることを可能にするだけでなく、交絡因子を適切に同定することができるという利点もある。加えて、交絡因子が同定できない場合においても、2.2.2 節で述べたように、グラフィカルモデルに基づく統計的因果推論にはさまざま因果効果の識別可能条件が用意されている。

2.4.3 グラフによる記述能力

前節では、グラフィカルモデルに基づく統計的因果推論を使うことに消極的な理由の一つとして、変数間の因果関係を明確にすることが困難であることを述べた。しかし、実際のところ、因果効果の識別問題において重要なことは、治療と反応変数の間の道がどのように有向分離されているかを考えるだけである。したがって、目的にもよるが、単に識別可能性だけを議論するのであれば不必要に複雑なグラフを用いる必要はないと考える。実

際、Geng and Li (2002) は治療の直接的原因が観測できれば、完全なグラフ表現がなくとも因果効果は識別可能であることを示している(もちろん、直接的原因をどのように同定するのかという問題は残っている)。また、Pearl (2009a) は、因果関係があることよりも因果関係がないことが本質であること、因果関係があるのにないと思って解析を行った場合には結果にバイアスを引き起こすことを指摘している。加えて、矢線の多くひかれたグラフにおいて因果効果が識別可能であれば、そこから興味の対象となる因果関係を保存する範囲でいくつかの矢線を取り除いたグラフにおいても因果効果は識別可能である(Pearl, 1995, 2009a)。したがって、因果効果をバイアスなく推定する際に因果関係の有無が明確でない場合には、因果関係が存在すると仮定して解析するのがベターであるということになるであろう。なお、因果効果を推定する上で、十分な変数集合を観測できない場合や因果構造があいまいな場合には、因果効果の存在範囲を評価するかあるいは感度分析を実施するという方策もありうる。

2.5 まとめ

本章では、統計的因果推論を用いた代替性評価尺度を提案する準備として、Pearl によって提案された統計的因果推論について、その基本的な考え方を述べた。その中で、背後にある仮定(哲学)は異なるものの、一貫性および SUTVA という 2つの仮定から出発する限りにおいては Pearl 流の統計的因果推論と Rubin 流の統計的因果推論のいずれを用いても得られる結果は同じであること、グラフィカルモデルに基づく方法よりも潜在反応モデルに基づく方法のほうが因果的仮説を詳細に表現できること、それでもなおグラフィカルモデルに基づく方法は有用であることを述べた。

Ronald Fisher はその著書 *Statistical Methods for Research Workers* (Fisher, 1925) の中で “The preliminary examination of most data is facilitated by the use of diagrams. Diagrams

prove nothing, but bring outstanding features readily to the eye; they are therefore no substitute for such critical tests as may be applied to the data, but are valuable in suggesting such tests, and in explaining the conclusions founded upon them”と述べている。この著書が刊行されたのはおおむね 100 年前であるが、この言葉はグラフィカル・モデルを用いた統計的因果推論にも当てはまるものと考えられる。本章の議論からわかるように、グラフィカルモデルを用いた統計的因果推論が、常に実質科学の要請に応えることができる統計解析法というわけではない。しかし、因果関係のグラフィカル表現は、潜在反応モデルだけでは把握することが困難な知見や因果仮説を視覚的に明らかにできるだけでなく、M-バイアスのような誤った共変量の導入や操作変数の誤用といった問題を回避でき、かつ因果効果の推定を行うための適切な共変量選択を実施できることから、計量生物学研究者・実務家に大きな恩恵をもたらすと考える。したがって、因果効果を評価する際に、因果関係が明確ではないとか複雑であるからといって最初から因果関係のグラフィカル表現を放棄することはせずに、まずは因果関係をグラフによって記述するよう努力すべきであり、その上で潜在反応モデルに基づいて統計的因果推論を議論すべきであると考えられる。

本章で述べた Pearl 流の統計的因果推論の議論をふまえたアプローチは、治療、代替エンドポイントそして臨床エンドポイントを取り巻く因果的な構造を考慮した代替性評価を可能とするものと考えられる。

第3章 統計的関連性を用いた治療効果に 対する代替性の評価尺度

本章では, 統計的関連性を用いた新たな代替性評価尺度として, proportion of the treatment effect captured by candidate surrogate endpoint (PCS) を提案し, 併せて推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法 (以降は'最頻値を用いた代替性の評価方法' と表記する) も提案する¹.

3.1 はじめに

代替エンドポイントの候補と考えられる測定値が, どのような要件を満たせば代替エンドポイントとして妥当であるかを統計的に検討する取り組みが, 過去 20 年にわたって進められている (Prentice, 1989; Weir and Walley, 2006). その中で代替性の評価尺度を用いたアプローチとして, たとえば Freedman et al. (1992) の proportion of the treatment effect explained (PTE), Wang and Taylor (2002) の proportion explained (PE), Qu and Case (2007) の proportion of information gain (PIG) が提案されている. しかしこれらの統計的関連性を用いた代替性評価尺度について, 生物統計分野の研究者や実践家により次の問題点が指摘されている. (i) 既存の代替性評価尺度の多くは, 特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとることがある (Li et al., 2001; Wang and Taylor, 2002), (ii) 既存の代替性評価尺度は, 代替性の水準を判断するための適切なカットオフ値を定義できない (Weir and Walley, 2006; Qu and Case, 2007), そして (iii) 既存の代替性評価尺度の多くは, そのばらつきが大

¹本章の内容は, Kobayashi and Kuroki (2014a) をまとめたものである.

大きく、そのため信頼区間が許容できないほどに広いことがある (Lin et al., 1997; De Gruttola et al., 1997). PTE と PE は上記三つの問題点をすべて抱えている. PIG は上記のうち問題 (i) と (iii) を解決しているが、問題 (ii) は解決できていない (Qu and Case, 2007).

そこで、これらの問題点を解決するために、統計的関連性を用いた新たな代替性評価尺度 PCS を提案する. 3.2 節では、準備として記号の表記、代替性の評価方法、および統計的関連性を用いた既存の代替性評価尺度の定式化やその特徴を示す. 3.3 節では、PCS が問題 (i) と (ii) を解決し、常に範囲 $[0, 1]$ 内の値をとること、代替性の水準を判断するための適当なカットオフ値を提示できることを示す. さらに、問題 (iii) を解決するため、推定された代替性評価尺度のブートストラップ分布 (Efron and Tibshirani, 1993) に対して Half-Range Mode (HRM) 法 (Bickel, 2002) を適用し代替性を評価する方法を提案する. 3.4 節では、数値実験をとおして、提案する尺度 PCS および最頻値を用いた代替性の評価方法が、既存の代替性評価尺度の問題点を解決するのに有効であることを示す. 3.5 節では、2つの無作為化比較試験 MEGA study (Nakamura et al., 2006) と ARMD study (Pharmacological therapy for macular degeneration study, 1997) への適用事例をとおして、提案する代替性評価尺度 PCS および最頻値を用いた代替性の評価方法によって代替エンドポイント候補の代替性が適切に評価されることを示す. 本章の結果を用いることで、代替エンドポイントの候補によってどの程度治療効果を捕捉できるかを信頼性を持って評価可能となることが期待される.

3.2 準備

本節では、記号の表記、統計的な代替性の評価方法、および統計的関連性を用いた既存の代替性評価尺度を紹介する.

X , S , および Y をそれぞれ、治療、代替エンドポイントの候補、および臨床エンドポイントを表す変数とする. D_X を X がとる値の領域とし、他の領域についても同様の表記を用いる. x , s , y を、それぞれ変数 X , S , Y の実現値とし、 $\text{pr}(X = x) = \text{pr}(x)$ と

$\text{pr}(S = s|X = x) = \text{pr}(s|x)$ をそれぞれ, $X = x$ の周辺確率 と $X = x$ を与えた下での $S = s$ の条件付き確率とする. さらに, $E(Y|x, s)$ を $(X, S) = (x, s)$ を与えた下での Y の条件付き期待値とし, 他の確率や期待値についても同様の表記を用いる.

本章では, 治療効果に対する代替エンドポイント候補 S の代替性について, 以下の統計的な評価を与える.

(a) 次の三つの条件を同時に満たすとき, S は完全な代替性を持つという

- ・ S を与えた下で, X と Y は条件付き独立である ($X \perp\!\!\!\perp Y|S$ と表記する)
- ・ X と S は独立ではない ($X \not\perp\!\!\!\perp S$ と表記する)
- ・ X を与えた下で, S と Y は条件付き独立ではない ($S \not\perp\!\!\!\perp Y|X$)

(b) 次の二つの条件のいずれかを満たすとき, S は代替性がないという

- ・ S と X は独立である ($X \perp\!\!\!\perp S$)
- ・ X を与えた下で, S と Y は条件付き独立である ($S \perp\!\!\!\perp Y|X$)

(c) 上記 (a), (b) 以外するとき, S は部分的な代替性を持つという

Wang and Taylor (2002) は, 代替性の評価尺度は, S が完全な代替性を持つときに 1 をとり, 代替性がないときに 0, そして, 部分的な代替性を持つときに 0 と 1 の間の値をとるべきであると述べている. 本章でもこの性質を満たす代替性の評価尺度を提案する. なお, 上記の代替性の評価に用いる統計的関連性は, 第 1 章で述べた Prentice 基準 (Prentice, 1989) が考慮している統計的関連性とわずかに異なる. 具体的には, 前者は X を与えた下での S と Y の条件付き独立ではないこと ($S \not\perp\!\!\!\perp Y|X$) を要求しているが, 後者の Prentice 基準では単に '代替エンドポイントの候補は臨床エンドポイントと関連がある' (すなわち, S と Y は条件付き独立ではない [$S \not\perp\!\!\!\perp Y$]) を求めているだけである. 本章では, Wang and Taylor (2002) の代替性評価にしたがい, X で条件付けした S と Y の独立関係を代替性評価に用いることにする.

以降, 統計的関連性を用いた代表的な代替性評価尺度を3つ紹介する. Freedman et al. (1992) は, 治療効果を代替エンドポイントの候補が説明する割合として PTE を提案した. PTE は, 次の一般化線形モデルにおける x の二つの回帰係数の比を用いて, $1 - \alpha_{yx \cdot s} / \alpha_{yx}$ と定義される.

$$g_1(E[Y|x]) = \alpha_{y \cdot x} + \alpha_{yx}x \quad (3.1)$$

$$g_2(E[Y|x, s]) = \alpha_{y \cdot xs} + \alpha_{yx \cdot s}x + \alpha_{ys \cdot x}s \quad (3.2)$$

ここで, $x \in D_X, s \in D_S$ とし, $g_1(\cdot)$ と $g_2(\cdot)$ は適当なリンク関数とする. また, $\alpha_{y \cdot x}$ と $\alpha_{y \cdot xs}$ は, それぞれ式 (3.1), 式 (3.2) における定数項とし, α_{yx} は, 式 (3.1) における x の回帰係数, $\alpha_{yx \cdot s}$ と $\alpha_{ys \cdot x}$ は, それぞれ式 (3.2) における x と s の回帰係数とする. 以降, 他のパラメータについても同様の表記を用いる. PTE は簡単に計算できるため, さまざまなモデルに適用可能である. 例えば, ロジスティックモデル (Freedman et al., 1992), Cox 比例ハザードモデル (Lin et al., 1997), そして線形構造方程式モデル (Qu and Case, 2006) などに適用されている. しかし, PTE は範囲 $[0, 1]$ 外の値をとることがあり (Li et al., 2001; Wang and Taylor, 2002), しばしば PTE の信頼区間は許容できないほどに広い (Lin et al., 1997; De Gruttola et al., 1997) という問題点が指摘されている.

Wang and Taylor (2002) は, Tsiatis et al. (1995) による以下の考察をふまえた代替性評価

尺度として PE を提案した。

“ if we use the definition of Prentice (1989), a good surrogate marker should have the following three properties: 1. The marker should be related to prognosis. 2. The distribution of values for the marker should be different for individuals receiving an effective treatment versus those receiving a placebo. 3. The beneficial effects of a good treatment should be mediated through its effect on the marker. That is, patients with the same value of a marker should have the same prognosis whether they are receiving a treatment or a placebo. In such a case the better prognosis associated with a good treatment could be explained by the change in the value of the marker for that treatment.”

Wang and Taylor (2002) の PE は, F と F' の 2 種類が定義されており, S が離散変数であるときは, 次のとおりに定められる。

$$F = \frac{h\left(\sum_{s \in D_S} g_{x_0}(s) \text{pr}(s|x_1)\right) - h\left(\sum_{s \in D_S} g_{x_0}(s) \text{pr}(s|x_0)\right)}{h\left(\sum_{s \in D_S} g_{x_1}(s) \text{pr}(s|x_1)\right) - h\left(\sum_{s \in D_S} g_{x_0}(s) \text{pr}(s|x_0)\right)} \quad (3.3)$$

$$F' = \frac{h\left(\sum_{s \in D_S} g_{x_1}(s) \text{pr}(s|x_1)\right) - h\left(\sum_{s \in D_S} g_{x_1}(s) \text{pr}(s|x_0)\right)}{h\left(\sum_{s \in D_S} g_{x_1}(s) \text{pr}(s|x_1)\right) - h\left(\sum_{s \in D_S} g_{x_0}(s) \text{pr}(s|x_0)\right)}$$

ここで $x_0, x_1 \in D_X, s \in D_S$ とし, g_{x_0} と g_{x_1} は, それぞれ $\text{pr}(y|x_0, s)$ と $\text{pr}(y|x_1, s)$ の関数, $h(\cdot)$ は単調な関数とする。なお, S が連続変数である場合は, \sum を積分記号に置き換えるものとする。本章では, F にのみ興味があるため, 以降はこれを PE と呼ぶ。PTE と同じように, PE もいくつかの条件を与えない限り, 範囲 $[0, 1]$ 外の値をとりうる (Wang and Taylor, 2002)。また, 式 (3.3) に $\text{pr}(s|x)$ が含まれていることから, X と S の関連の程度が, PE の値に影響を与えることがわかる。

Qu and Case (2007) は, カルバック・ライブラー情報量の増加度合に基づいた, PIG という代替性評価尺度を提案した。PIG は, S による情報量の増加と (X, S) による情報量の増

加の比を用いて $LRT(S : 1)/LRT(S, X : 1)$ と定義される. ここで, $LRT(S : 1)$ は, 次の二つの線形モデルを比較する尤度比検定統計量 (LRT) とする.

$$g_3(E[Y|s]) = \alpha_{y \cdot s} + \alpha_{ys}s \quad (3.4)$$

$$g_4(E[Y]) = \alpha_y \quad (3.5)$$

また, $LRT(S, X : 1)$ は, 式 (3.2) と式 (3.5) を比較する LRT とする. PIG は, 常に範囲 $[0, 1]$ 内の値をとるが, 代替性の水準を判断するための適切なカットオフ値は定義できない (Qu and Case, 2007). また, PE とは異なり, PIG は, その定義式から X と S の関連の程度をどの程度反映するかは明らかではない. 加えて, PTE と比較した際, PIG が考慮している条件付き従属関係 (すなわち, S を与えた下での X と Y の従属関係 (Qu and Case, 2007)) は, PTE が考慮している条件付き従属関係 (X を与えた下での S と Y の従属関係 (Freedman et al., 1992)) とは異なっている. これらの違いは 後述する数値実験や適用例において再度ふれることにする.

3.3 提案手法

3.3.1 Proportion of the treatment effect captured by candidate surrogate endpoint (PCS)

統計的関連性を用いた既存の代替性評価尺度の三つの問題点のうち, (i) 特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとることがある, および (ii) 代替性の水準を判断するための適切なカットオフ値を定義できない, という問題点を解決するため, 新たな代替性の評価尺度として, 以下の PCS を提案する.

$$PCS = \frac{CP^2}{CP^2 + NCP^2} \quad (3.6)$$

ここで、CP と NCP はそれぞれ、Captured Portion と Non Captured Portion の略であり、治療効果 (TE) のうち代替エンドポイントの候補 S によって捕捉される部分と、捕捉されない部分を表す。そして、CP と NCP は以下の条件を満たすものとする。

(a) CP と NCP の和は、TE と一致する

(b) CP と NCP は、TE と同じ型および単位をもつ値とする

(c) S が完全な代替性を持つとき $\text{NCP} = 0$ 、 S には代替性がないとき $\text{CP} = 0$ をとる
本章では、CP と NCP が共にゼロであるとき (すなわち、 $\text{TE} = 0$ のとき)、 $\text{PCS} = 0$ と定義し、 S には代替性がないと判断する。ここで、本章では代替エンドポイントの候補が一つである状況を考えているが、複数の代替エンドポイントの候補がある場合にも直接 適用可能である。

上記の条件 (a), (b), および (c) を満たすような、CP と NCP の定式化は複数考えることができるが、Tsiatis et al. (1995) の見解をふまえると、CP と NCP はそれぞれ、 $\sum_{D_s} \text{pr}(y|x_0, s)\text{pr}(s|x_1)$ と $\text{pr}(y|x_0)$ を比較する統計的な性質、および、 $\sum_{D_s} \text{pr}(y|x_0, s)\text{pr}(s|x_1)$ と $\text{pr}(y|x_1)$ を比較する統計的な性質を反映することが適当と考えられる。例えば、無作為化比較試験において、TE として、 $\text{TE} = E(Y|x_1) - E(Y|x_0)$ が評価される場合、CP と NCP は、それぞれ次のように定めることができる。

$$\text{CP} = \sum_{D_s} E(Y|x_0, s) \{ \text{pr}(s|x_1) - \text{pr}(s|x_0) \} \quad (3.7)$$

$$\text{NCP} = \sum_{D_s} \{ E(Y|x_1, s) - E(Y|x_0, s) \} \text{pr}(s|x_1) \quad (3.8)$$

ここで、PCS および CP, NCP は、この TE の設定に限定されるものではなく、TE がさまざまな比の型によって定義される場合にも適用できる。詳細は、付録 3.7 節を参照されたい。

式 (3.6) より、PCS は常に範囲 $[0, 1]$ 内の値をとることがわかる。加えて、式 (3.7) と (3.8)、もしくは付録 3.7 節に示すような CP と NCP の下では、 $X \perp\!\!\!\perp S$ もしくは $S \perp\!\!\!\perp Y|X$ が成立す

るとき $PCS=0$, そして $X \perp\!\!\!\perp Y|S$, $X \perp\!\!\!\perp S$ と $S \perp\!\!\!\perp Y|X$ が同時に成立するとき $PCS=1$, となることがわかる. ゆえに PCS は, 代替性の統計的な評価に求められる性質を満たしつつ, 既存の PTE や PE のような代替性の評価尺度の問題点 (範囲 $[0, 1]$ 外の値をとる) を解決していることがわかる.

CP と TE をそれぞれ, 式 (3.3) の PE の分子と分母によって定義したとき, PCS は PE の関数として記述することができる.

$$PCS = \frac{1}{1 + (1/PE - 1)^2} \quad (3.9)$$

これより, PCS は Wang and Taylor (2002) の PE を拡張した代替性評価尺度であると解釈することができる. そして, PE と同じように, PCS も $pr(s|x)$ に基づく定式化であるため, X と S の従属関係の程度が PCS の値にも反映される. そのため, S が完全に $X \perp\!\!\!\perp Y|S$ を満たす場合を除き, X と S の従属関係の程度が小さくなるのに応じて PCS は小さい値をとることになる.

3.3.2 代替性の水準を判断するためのカットオフ値

PTE , PE , や PIG のような既存の代替性評価尺度とは異なり, PCS はその微分を参考にすることで, 代替性の水準を判断するための客観的なカットオフ値の設定が可能である. これを示すため, TE の値を与えた下で, PCS を CP の関数 $PCS = CP^2 / \{CP^2 + (TE - CP)^2\}$ として表すことにする. このとき, CP について 2 階微分をとることにより, $CP = TE/2$, $TE \times (1 \pm \sqrt{3})/2$ が PCS の変曲点であることがわかる. さらに, CP について 3 階微分をとることにより, $CP = \pm TE/\sqrt{2}$, $TE \times (2 \pm \sqrt{2})/2$ が PCS の最小躍度の点であることがわかる.

例として, $TE = 1.000$ と固定したうえで CP のとりうる範囲が $(-1.0, 2.0)$ であるとした場合の PCS の概形を図 3.1 に与える. 図 3.1 より, CP が -0.366 , 0.5 , 1.366 であるとき, PCS の変曲点がそれぞれ 0.067 , 0.5 , 0.933 (○で示される点) であることがわかる. また, CP

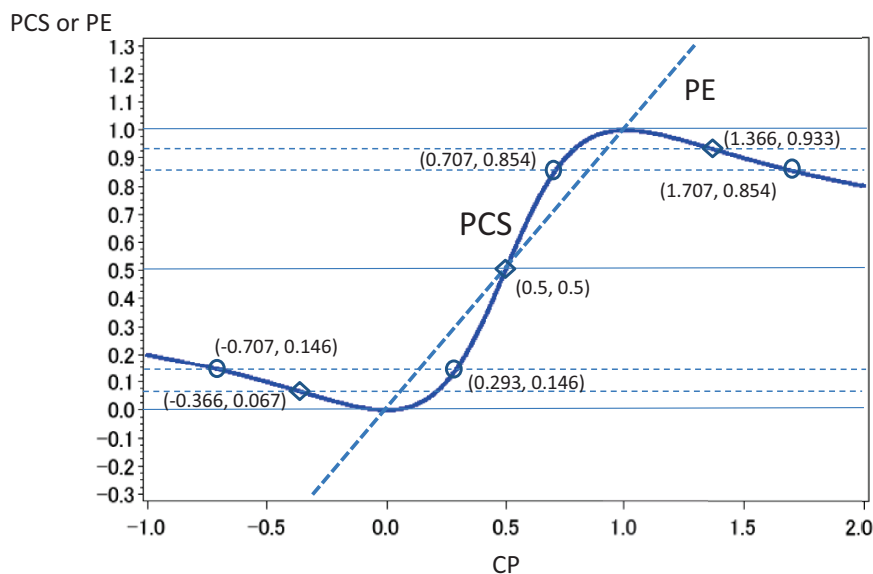


図 3.1: $TE = 1.000$ における PCS (実線) と PE (点線) ; 変曲点を○, 最小躍度の点を◇で示す

が $-0.707, 0.293, 0.707, 1.707$ のとき, PCS の最小躍度の点がそれぞれ $0.146, 0.146, 0.854, 0.854$ (◇で示される点) であることもわかる. なお, CP が ∞ もしくは $-\infty$ に近づくと, PCS は 0.5 に近づくことがわかる.

これら PCS の変曲点や最小躍度の点を参考にしながら, 代替性の水準を判断するための PCS のカットオフ値 (6 水準) を設定することができる (表 3.1 の 2 列目). なお, 水準の名称は, Cohen の κ 係数として知られる一貫性尺度に対して, Landis and Koch (1977) が利用した水準名を参考とした. ただし, “moderate” の範囲が実践的には広すぎると考えられる場合は, PCS のより高次の微分を参考にし, 7 水準 (表 3.1 の 3 列目) を設定することもできる. 本章では, 後者の 7 水準を代替性のベンチマークとして利用する. なお, 標本のばらつきを考慮すると, 表 3.1 に基づく代替性の評価は, PCS の点推定値より, その信頼区間の上限と下限を用いて行う方が適切であると考えられる.

最後に, 図 3.1 より, PCS と PE の違いが明確になる. すなわち, TE を与えた下で, PE は CP の線形な関数であるため, CP が 0 より小さいもしくは 1 より大きいときに, PE は範囲

表 3.1: PCS による代替性の水準

	代替性の水準 (6 水準)	代替性の水準 (7 水準)
PCS=1.000	perfect	perfect
$0.933 \leq \text{PCS} < 1.000$	almost perfect	almost perfect
$0.854 \leq \text{PCS} < 0.933$	substantial	substantial
$0.634 \leq \text{PCS} < 0.854$	moderate	moderate
$0.500 \leq \text{PCS} < 0.634$		good
$0.000 < \text{PCS} < 0.500$	poor	poor
PCS=0.000	useless	useless

[0, 1] 外の値をとる. さらに, PE は PCS のようなカットオフ値を設定することができないことがわかる.

3.3.3 符号付 PCS による代替エンドポイントの区別

Wang and Taylor (2002) は, PE が [0, 1] に収まるための十分条件を与え, 条件を満たさない場合は [0, 1] 外の値をとること, そしてその場合の PE に対して “most likely indicates the existence of unintended effects” という考察を加えている. 実践的には, このような意図しない効果 (unintended effects) が存在する状況も考えられるため, 本章の PCS の設定においても, これを評価する必要がある. 上記の状況を考察する前に, PCS の背景にある考え, すなわち, 図 3.2 において TE と CP をそれぞれベクトル (cp, ncp) と $(cp, 0.000)$ に対応させたときのクラスター分析でしばしば使われているコサイン類似度, を示すことが有用である. ここで, ベクトル (cp, ncp) と $(cp, 0.000)$ のコサイン類似度は, ベクトル $(cp, -ncp)$ と $(cp, 0.000)$ のそれとは区別ができないことがわかる. それは, 二つの角度 θ_1 と θ_2 が同じであることから明らかである. それゆえ, PCS は, TE と CP の類似度を反映した尺度であると考えられることができるが, 一方で CP と NCP が同じ符号となる場合の S と CP と NCP が互いに異なる符号となる場合の S を区別することはできない. 例えば, 図 3.1 の設定では, $0.0 < \text{CP} < 1.0$ の場合は, CP と NCP が同じ符号となる場合の S であると評価でき, 一方 $\text{CP} < 0.0$ もしくは $1.0 < \text{CP}$ の場合は, CP と NCP が互いに異なる符号となる場合の S と評価

されるがこれらを PCS は区別することができない. また, CP が 0.707 と 1.707 のとき, PCS は 0.854 をとり, 同じ代替性の水準であると判断されることになる.

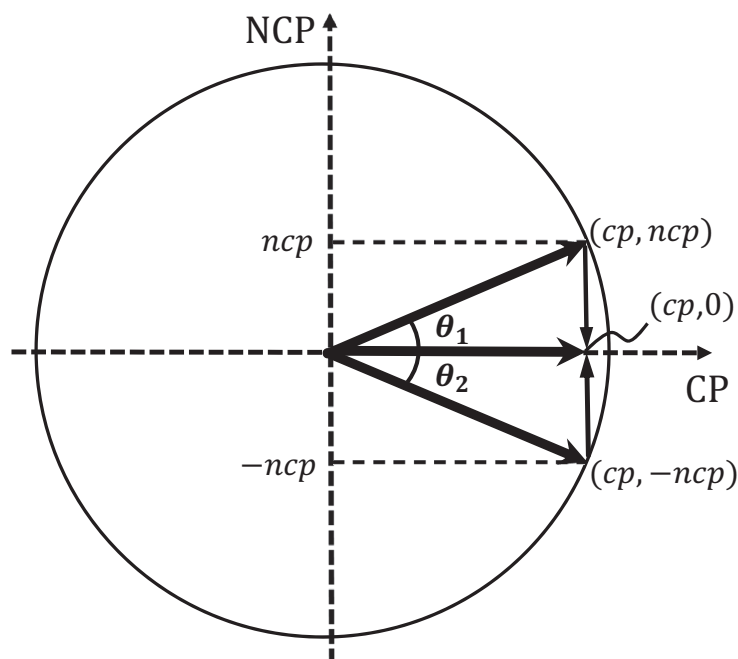


図 3.2: PCS の背景にある考え (コサイン類似度)

上記二つの状況を区別することを目的としたとき, PCS のフレームワークでは, 符号付 PCS として, $I(\text{CP}, \text{NCP}) \times \text{PCS}$ を提案することができる. ここで, $I(u, v)$ は, 実数 u と v を用いて, $u \times v$ が非負の値をとる時に 1 をとり, $u \times v$ が負の値をとるときに -1 をとる関数である. 符号付 PCS は, TE の値に応じて -1 と 1 の間の値をとる, $I(\text{CP}, \text{NCP}) = -1$ のときに ‘CP と NCP が互いに異なる符号となる場合の S' ’ と判断し, $I(\text{CP}, \text{NCP}) = 1$ のときに ‘CP と NCP が同じ符号となる場合の S' ’ と判断することが可能である. しかし, 本章では, 統計的な扱いやすさを考慮し, PCS のみを扱うことにする.

3.3.4 推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法

既存の代替性の評価尺度と同様な手続きにしたがって、適当な方法(本章では最尤法)により PCS の推定値を得ることができる。しかし、PTE をはじめとする多くの代替性の評価尺度と同様に、代替エンドポイントの候補と治療の間の相関が高い場合には、最尤法による推定精度は低下する。このとき、PCS も同様に推定精度が低下するため、信頼区間が広がるという問題を抱える。そこで、この問題を避けるための代替性の評価方法として、次の手順を提案する。

手順 (i). オリジナルデータからブートストラップ標本 (Efron and Tibshirani, 1993) を生成する

手順 (ii). 手順 (i) にて生成されたブートストラップ標本のそれぞれに対し、最尤法のような適当な方法により代替性の評価尺度の推定値を得る

手順 (iii). 手順 (ii) にて推定された代替性評価尺度のブートストラップ分布における最頻値を求める

最頻値は、一般に興味のある代替性の評価尺度の一致推定量とはならないが、推定された尺度の中で最も頻繁に出現するという意味で適当な値である。最頻値の推定において、通常のヒストグラム法 (Van Ryzin, 1973) を用いることも可能ではあるが、階級幅のとり方に依存して異なった推定値が得られるという問題が生じる。そこでこの恣意性を避けるために、HRM 法 (Bickel, 2002) を用いて、代替性の評価尺度のブートストラップ分布の最頻値を推定する (Efron and Tibshirani, 1993)。この手順は、多重共線性の問題を避けるための方法として、PCS に限らず、統計的関連性を用いた既存の代替性評価尺度にも適応可能である。

データが有限領域における単峰の連続分布から生成されたと仮定すると、HRM 法は与えられたデータの範囲 (range) を半分 (half) にしたときに、密度の高い範囲を繰り返し選ぶこと

で、最終的にそのデータの最頻値を含む範囲を選ぶことが可能となる (Bickel and Frühwirth, 2006). ここで, PCS を例にとって, HRM 法の直感的な解釈を示す. まず, B を反復回数とし, 得られた $\widehat{\text{PCS}}$ のブートストラップ分布を大きさの順に並び替え, これを $\widehat{\text{PCS}}_j, (j = 1 \dots B)$ とし, この分布の範囲を R_1 と表す. 次に, その範囲 R_1 を 2 等分した幅 ($R_1/2 = \widehat{\text{PCS}}_B - \widehat{\text{PCS}}_1$) を用いて, $j = 1 \dots B$ のそれぞれに対し, 範囲 $R_{1j} = [\widehat{\text{PCS}}_{1j}, \widehat{\text{PCS}}_{1j} + r_1/2]$ を設ける. そして, 範囲 R_{1j} に含まれる $\widehat{\text{PCS}}$ の個数が最も多い範囲を R_2 とし, この R_2 の内側に最頻値が含まれていると考える. 以降, 範囲 R_1 を R_2 に更新した上で, 上記と同様の手順を, 更新後の範囲に含まれる $\widehat{\text{PCS}}$ の個数が 2 になるまで繰り返し, 最後に残った 2 つの値の平均値を最頻値の推定値とする.

3.4 数値実験

3.4.1 設定

本節では, 数値実験を用いて, PCS と PTE, PE, そして PIG による代替性の評価性能を比較し, さらに最頻値を用いた代替性の評価方法と従来の最尤法による評価の比較も行う. 簡単のため, 得られるデータは以下の線形構造方程式モデルによって生成されているものとする.

$$S = \alpha_{sx}X + \varepsilon_{s,x} \quad (3.10)$$

$$Y = \alpha_{yx \cdot s}X + \alpha_{ys \cdot x}S + \varepsilon_{y,xs} \quad (3.11)$$

ここで $(\varepsilon_{s,x}, \varepsilon_{y,xs})$ は, 平均ベクトルを $(0.00, 0.00)$ とし, 分散共分散行列を $\begin{pmatrix} \sigma^2 & 0.00 \\ 0.00 & 1.00 \end{pmatrix}$ とする 2 変量正規分布にしたがう.

本節では, 式 (3.7) と (3.8) により, PCS を算出する. また, 式 (3.3) において, h, g_{x_0} , およ

び g_{x_1} をそれぞれ、恒等関数、 $E(Y|x_0, s)$ 、および $E(Y|x_1, s)$ としたとき、恒等リンク関数を用いた式 (3.1), (3.2), (3.4), (3.5) の回帰係数と尤度比検定統計量により、PTE と PIG を算出する。なお、この設定においては、PE は PTE と同じ定式化により求められる。ここで、 X は、確率 $1/2$ で 1 または 0 をとるベルヌイ分布にしたがうものとする。式 (3.10) と (3.11) の下で、大きさ 2000 の標本を生成する。

以降は、以下の 3 つの設定を考える。

設定 1: $\alpha_{sx} = \alpha_{ys\cdot x} = 1.00$, $\alpha_{yx\cdot s} = 0.00$ の下で、 σ^2 を $0.01, 0.10, 1.00, 2.00, 4.00$ とした時の、代替性の評価尺度の挙動を確認する

設定 2: $\alpha_{sx} = 1.00$, $\sigma^2 = 1.00$ の下で、 $(\alpha_{yx\cdot s}, \alpha_{ys\cdot x})$ を $(0.00, 1.00)$, $(0.10, 0.90)$, $(0.30, 0.70)$, $(0.50, 0.50)$, $(1.00, 0.00)$ とした時の、代替性の評価尺度の挙動を確認する

設定 3: $(\alpha_{yx\cdot s}, \alpha_{ys\cdot x}) = (0.50, 0.50)$, $\sigma^2 = 1.00$ の下で、 α_{sx} を $0.00, 0.20, 0.50, 1.00, 2.00$ とした時の、代替性の評価尺度の挙動を確認する

設定 1 の数値実験により、式 (3.11) における X と S の多重共線性が、代替性の評価尺度へ与える影響を評価することができる。このとき、 S を与えた下で X が Y と条件付き独立であって、それ以外の統計的な独立関係が成立しないとき、 σ^2 のそれぞれの値において、 S は完全な代替性を持つこと (すなわち、 $PCS=1$) が期待される。設定 2 の数値実験により、二つの条件付き独立 (S を与えた下での X と Y の条件付き独立、 X を与えた下での S と Y の条件付き独立) が、代替性評価尺度へ与える影響を評価することができる。ここで、 $(\alpha_{yx\cdot s}, \alpha_{ys\cdot x}) = (0.00, 1.00)$ のとき、 S は完全な代替性を持つこと ($PCS=1$) が期待され、一方で $(\alpha_{yx\cdot s}, \alpha_{ys\cdot x}) = (1.00, 0.00)$ のとき、 S は代替性がないこと ($PCS=0$) が期待される。設定 3 の数値実験により、 X と S の従属関係が、代替性評価尺度へ与える影響を評価することができる。ここで、 $\alpha_{sx} = 0.00$ のとき、 S は代替性がないこと ($PCS=0$) が期待される。

3.4.2 推定

$\alpha_{yx \cdot s}$ の最尤法による推定値を $\hat{\alpha}_{yx \cdot s}$ とし, 他の最尤法による推定値にも同様の表記を用いる. 回帰係数 α_{sx} , $\alpha_{ys \cdot x}$, $\alpha_{yx \cdot s}$ の最尤推定値を用いて, PCS は $\hat{\alpha}_{sx}^2 \hat{\alpha}_{ys \cdot x}^2 / (\hat{\alpha}_{sx}^2 \hat{\alpha}_{ys \cdot x}^2 + \hat{\alpha}_{yx \cdot s}^2)$ によって評価される. PTE も同様に $1 - \hat{\alpha}_{yx \cdot s} / \hat{\alpha}_{yx}$ によって評価される. ここで $\hat{\alpha}_{yx}$ は, 次の回帰モデルに基づいて推定される.

$$Y = \alpha_{yx}X + \varepsilon_{y \cdot x} \quad (3.12)$$

なお, 式 (3.10) と式 (3.11) を式 (3.12) と比較することによって, $\alpha_{yx} = \alpha_{yx \cdot s} + \alpha_{sx}\alpha_{ys \cdot x}$, $\varepsilon_{y \cdot x} = \alpha_{ys \cdot x}\varepsilon_{s \cdot x} + \varepsilon_{y \cdot xs}$ の関係が導かれる. PIG は, 式 (3.11) と $Y = \varepsilon_y$ を比較する尤度比検定統計量 $LRT(S, X : 1)$, および $Y = \alpha_{ys}S + \varepsilon_{y \cdot s}$ と $Y = \varepsilon_y$ を比較する尤度比検定統計量 $LRT(S : 1)$ を通じて評価される. 3.2.2 節で述べたように, 代替性の水準を判断するためのカットオフ値 (表 3.1) を用いる際は, PCS の点推定値ではなく, その信頼区間の上限と下限を用いて代替性を評価する. この時, 最尤法を用いて代替性の評価をする場合は, 代替性評価尺度の信頼区間として, 1000 個の最尤推定による推定値の 2.5% 点と 97.5 % 点を用いる. また, 最頻値を用いて代替性の評価をする場合は, 2 段階ブートストラップ法 (Davison and Hinkley, 1997) に基づき, 代替性評価尺度の 1000 個の最尤推定値からリサンプリングして得られる最頻値を 500 個用意し, その 2.5% 点, および 97.5% 点を用いる.

3.4.3 結果

表 3.2 に, 提案した尺度 PCS と既存の代替性評価尺度による代替性の評価の違いを比較した結果を示す. 表 3.2 より次の考察が得られる.

1. 表 3.2(a) より, 設定 1 に対して最尤法によって代替性を評価したとき, σ^2 の各値に対し, PTE の 97.5 % 点は 1.000 より大きい, PIG と PCS の 2.5% 点と 97.5 % 点は, 範

表 3.2: 数値実験結果

推定値 (2.5%点, 97.5%点)

(a) 設定 1: 最尤法による代替性評価			
σ^2	PTE	PIG	PCS
0.01	1.004 (0.393, 1.699)	0.996 (0.978, 1.000)	0.890 (0.295, 1.000)
0.10	1.013 (0.796, 1.252)	0.997 (0.984, 1.000)	0.988 (0.935, 1.000)
1.00	1.010 (0.881, 1.169)	0.999 (0.994, 1.000)	0.995 (0.976, 1.000)
2.00	1.006 (0.877, 1.147)	0.999 (0.996, 1.000)	0.995 (0.977, 1.000)
4.00	1.007 (0.890, 1.156)	0.999 (0.997, 1.000)	0.995 (0.977, 1.000)

(b) 設定 1: 最頻値を用いた代替性評価			
σ^2	PTE	PIG	PCS
0.01	0.992 (0.784, 1.119)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
0.10	1.039 (0.937, 1.051)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
1.00	0.980 (0.945, 1.034)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
2.00	1.001 (0.958, 1.033)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
4.00	0.988 (0.972, 1.016)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)

(c) 設定 2: 最尤法による代替性評価			
$(\alpha_{yx \cdot s}, \alpha_{ys \cdot x})$	PTE	PIG	PCS
(0.00,1.00)	1.004 (0.874, 1.164)	0.999 (0.994, 1.000)	0.995 (0.976, 1.000)
(0.10,0.90)	0.904 (0.789, 1.058)	0.996 (0.985, 1.000)	0.981 (0.933, 1.000)
(0.30,0.70)	0.704 (0.596, 0.819)	0.966 (0.931, 0.991)	0.840 (0.685, 0.953)
(0.50,0.50)	0.503 (0.418, 0.604)	0.879 (0.813, 0.937)	0.505 (0.339, 0.699)
(1.00,0.00)	0.001 (-0.065, 0.068)	0.187 (0.099, 0.302)	0.001 (0.000, 0.005)

(d) 設定 2: 最頻値を用いた代替性評価			
$(\alpha_{yx \cdot s}, \alpha_{ys \cdot x})$	PTE	PIG	PCS
(0.00,1.00)	0.953 (0.949, 1.025)	1.000 (1.000, 1.000)	1.000 (1.000, 1.000)
(0.10,0.90)	0.906 (0.860, 0.926)	1.000 (1.000, 1.000)	1.000 (0.999, 1.000)
(0.30,0.70)	0.695 (0.677, 0.702)	0.967 (0.961, 0.974)	0.838 (0.822, 0.885)
(0.50,0.50)	0.495 (0.479, 0.524)	0.886 (0.873, 0.896)	0.531 (0.441, 0.547)
(1.00,0.00)	-0.005 (-0.014, 0.008)	0.183 (0.153, 0.184)	0.000 (0.000, 0.000)

(e) 設定 3: 最尤法による代替性評価			
α_{sx}	PTE	PIG	PCS
0.00	-0.004 (-0.132, 0.117)	0.776 (0.683, 0.871)	0.004 (0.000, 0.019)
0.20	0.165 (0.065, 0.253)	0.793 (0.695, 0.877)	0.042 (0.005, 0.103)
0.50	0.336 (0.252, 0.422)	0.828 (0.742, 0.902)	0.209 (0.102, 0.347)
1.00	0.504 (0.415, 0.604)	0.880 (0.810, 0.938)	0.508 (0.335, 0.699)
2.00	0.665 (0.571, 0.762)	0.946 (0.907, 0.976)	0.791 (0.638, 0.911)

(f) 設定 3: 最頻値を用いた代替性評価			
α_{sx}	PTE	PIG	PCS
0.00	0.020 (-0.038, 0.030)	0.765 (0.763, 0.797)	0.000 (0.000, 0.000)
0.20	0.173 (0.149, 0.184)	0.807 (0.782, 0.817)	0.034 (0.021, 0.042)
0.50	0.325 (0.316, 0.355)	0.840 (0.810, 0.858)	0.188 (0.163, 0.229)
1.00	0.504 (0.486, 0.525)	0.881 (0.876, 0.897)	0.508 (0.466, 0.549)
2.00	0.651 (0.637, 0.690)	0.943 (0.943, 0.961)	0.829 (0.760, 0.853)

囲 (0.000, 1.000) に含まれることがわかる. 他方で, 代替性評価尺度のそれぞれに対し, $\sigma^2 = 0.01$ における推定精度は, $\sigma^2 \geq 0.10$ に比べて悪い. これは, 最尤法を用いた代替性評価尺度は, 多重共線性の推定精度悪化の問題の影響を受けてしまうことを意味している.

2. 表 3.2(b) より, 設定 1 に対して最頻値を用いて代替性を評価した場合は, σ^2 の各値に対し, 代替性評価尺度それぞれの 2.5% 点と 97.5 % 点に基づく範囲は, 最尤法によるそれらより狭いことがわかる. それでもなお, PTE の 97.5 % 点は 1.000 より大きい. これは, PTE の真値が 1.000 に近いときは, 何らかの仮定をおかない限り, PTE が 1.000 より大きい値をとるという状況は改善されにくいことを意味している. 一方で, PCS によれば, 表 3.1 の代替性の水準にしたがった S の代替性の水準は, “perfect” であると判断することができる.
3. 表 3.2(c) より, 設定 2 に対して最尤法によって代替性を評価したとき, $(\alpha_{yx \cdot s}, \alpha_{ys \cdot x}) = (0.00, 1.00)$ と $(0.10, 0.90)$ の場合に, PTE の 97.5 % 点は 1.000 より大きい, PIG と PCS の 2.5% 点と 97.5 % 点は, 範囲 (0.000, 1.000) に含まれることがわかる. 加えて, $\alpha_{ys \cdot x}$ が小さくなるにしたがって, PIG と PCS の 2.5% 点と 97.5 % 点は共に小さくなるが, 0.000 を下回ることはない. PTE の 2.5% 点と 97.5 % 点 も同様に小さくなるが, $(\alpha_{yx \cdot s}, \alpha_{ys \cdot x}) = (1.00, 0.00)$ の場合には, 2.5 % 点が 0.000 を下回ってしまう. 他方で, 同じ $(\alpha_{yx \cdot s}, \alpha_{ys \cdot x}) = (1.00, 0.00)$ の場合に, PCS の 2.5% 点と 97.5 % 点に基づく範囲は 0.000 を含むが, PIG のそれは 0.000 を含まない. すなわち, PCS によれば, 表 3.1 を用いた S の代替性の水準は “poor” もしくは “useless” であると判断することができる. 一方で, PIG によると “useless” であると判断することはできない. これらの結果から, PCS は, X と S の従属関係を PIG より正確に反映していることが確認できる.
4. 表 3.2(d) より, 設定 2 に対して最頻値を用いて代替性を評価した場合は, 代替性評価尺度それぞれの 2.5% 点と 97.5 % 点に基づく範囲は, 最尤法によるそれらより狭いこ

とがわかる. $(\alpha_{yx.s}, \alpha_{ys.x}) = (0.00, 1.00)$ の場合, **PIG** と **PCS** の 2.5%点と 97.5 %点 はともに 1.000 である. これは, S は完全な代替性をもつ評価項目であると判断可能であることを意味する. 加えて, $\alpha_{ys.x}$ が小さくなるにしたがって, 代替性評価尺度それぞれは小さい値をとる. さらに, $(\alpha_{yx.s}, \alpha_{ys.x}) = (1.00, 0.00)$ の場合は, **PIG** と **PCS** の 2.5 %点は 0.000 を下回っていないが, **PTE** のそれは 0.000 を下回っている. 同じ $(\alpha_{yx.s}, \alpha_{ys.x}) = (1.00, 0.00)$ の場合に, **PIG** と **PCS** を比べたとき, **PCS** の 2.5%点と 97.5 %点 はともに 0.000 であるが, **PIG** のそれらは 0.000 ではない. すなわち, **PCS** によれば, 表 3.1 を用いた S の代替性の水準は “useless” であると判断することができるが, **PIG** によると “useless” であるとは判断できない. これらの結果から, 他の代替性評価尺度に比べて, **PIG** は代替性を高く評価してしまうことが確認できる.

5. 表 3.2(e) より, 設定 3 に対して最尤法によって代替性を評価したとき, $\alpha_{sx} = 0.00$ の場合に, **PCS** の 2.5 %点は 0.000 であるが, **PTE** のそれは 0.000 を下回っていることがわかる. ここで, **PIG** は **PTE** と **PCS** より明らかに大きい値をとっている. 加えて, α_{sx} が大きくなるにしたがって, 代替性評価尺度のそれぞれは大きい値をとるが, $\alpha_{sx} \leq 1.00$ の場合は, **PIG** の 2.5 %点は, 他の代替性評価尺度の 97.5 %点より, 明らかに大きい値をとる. これらの結果から, **PIG** に比べ, **PCS** は X と S の従属関係の程度を適切に反映していることが確認できる.
6. 表 3.2(f) より, 設定 3 に対して最頻値を用いて代替性を評価した場合は, 代替性の評価尺度それぞれの 2.5%点と 97.5 %点に基づく範囲は, 最尤法によるそれらより狭いが, **PTE** の 2.5%点は, $\alpha_{sx} = 0.00$ の場合に負の値をとることがわかる. 他方で, **PCS** の 2.5%点と 97.5 %点 はともに 0.000 であるため, S の代替性の水準は “useless” であると判断することができる. 加えて, α_{sx} が大きくなるにしたがって, 代替性評価尺度のそれぞれは大きい値をとっているが, **PIG** の 2.5 %点は, 他の代替性評価尺度の 97.5 %点より, 明らかに大きい値をとる. それゆえ, **PIG** は, 他の代替性評価尺度に比

べて、**PIG** は代替性を高く評価してしまうことが確認できる。

以上までの考察は、提案した尺度 **PCS** と既存の代替性評価尺度による代替性の評価の違いに注目していた。そこでは、最尤法による評価と提案手法である **HRM** 法を用いた最頻値による評価の違いにも言及していたが、両者の推定方法に関する統計的な性質は明確ではなかった。そこで、最尤法による代替性評価と最頻値を用いた代替性評価のそれぞれの場合のバイアスと分散を表 3.3 に示す。表 3.3 より、**PTE** に関しては、設定 1 から設定 3 の多くで最頻値を用いた代替性評価のバイアスは最尤法のそれより大きい、分散についてはすべての設定で最頻値を用いた代替性評価は最尤法のそれより小さい。**PCS** に関しては、設定 1 のそれぞれの場合に対し、最頻値を用いた代替性評価のバイアスと分散は、最尤法のそれより小さい。設定 2 と設定 3 については、最頻値を用いた代替性評価は最尤法のそれより大きいバイアスをとることもあるが、分散についてはすべての場合に対し、最頻値を用いた代替性評価は最尤法のそれより小さいことがわかる。なお、付録の表 3.8 に、平均二乗誤差の平方根を用いて最尤法による代替性評価と最頻値を用いた代替性評価を比較した結果を示す。この結果から、すべての設定に対し最頻値を用いた代替性評価の平均二乗誤差の平方根は最尤法のそれより小さく、本章で提案した最頻値を用いた代替性の評価方法は、代替性評価尺度の 2.5%点と 97.5%点に基づく範囲を狭くしていることが確認できる。

表 3.3: 数値実験におけるバイアスと分散

設定 1						
	σ^2	真値	最尤法による代替性評価		最頻値を用いた代替性評価	
			バイアス	$\sqrt{\text{分散}}$	バイアス	$\sqrt{\text{分散}}$
PTE	0.01	1.000	0.004	0.328	-0.008	0.070
	0.10	1.000	0.013	0.115	0.039	0.033
	1.00	1.000	0.010	0.074	-0.020	0.017
	2.00	1.000	0.006	0.069	0.001	0.017
	4.00	1.000	0.007	0.071	-0.012	0.013
PCS	0.01	1.000	-0.110	0.175	0.000	0.000
	0.10	1.000	-0.012	0.018	0.000	0.000
	1.00	1.000	-0.005	0.007	0.000	0.000
	2.00	1.000	-0.005	0.006	0.000	0.000
	4.00	1.000	-0.005	0.007	0.000	0.000
設定 2						
	$(\alpha_{yx \cdot s}, \alpha_{ys \cdot x})$	真値	最尤法による代替性評価		最頻値を用いた代替性評価	
			バイアス	$\sqrt{\text{分散}}$	バイアス	$\sqrt{\text{分散}}$
PTE	(0.00,1.00)	1.000	0.004	0.071	-0.047	0.024
	(0.10,0.90)	0.900	0.004	0.069	0.006	0.019
	(0.30,0.70)	0.700	0.004	0.058	-0.005	0.007
	(0.50,0.50)	0.500	0.003	0.048	-0.005	0.012
	(1.00,0.00)	0.000	0.001	0.032	-0.005	0.005
PCS	(0.00,1.00)	1.000	-0.005	0.007	0.000	0.000
	(0.10,0.90)	0.988	-0.007	0.020	0.012	0.000
	(0.30,0.70)	0.845	-0.005	0.070	-0.007	0.018
	(0.50,0.50)	0.500	0.005	0.092	0.031	0.025
	(1.00,0.00)	0.000	0.001	0.002	0.000	0.000
設定 3						
	α_{sx}	真値	最尤法による代替性評価		最頻値を用いた代替性評価	
			バイアス	$\sqrt{\text{分散}}$	バイアス	$\sqrt{\text{分散}}$
PTE	0.00	0.000	-0.004	0.065	0.020	0.022
	0.20	0.167	-0.002	0.048	0.006	0.009
	0.50	0.333	0.002	0.045	-0.009	0.011
	1.00	0.500	0.004	0.049	0.004	0.007
	2.00	0.667	-0.001	0.050	-0.016	0.019
PCS	0.00	0.000	0.004	0.007	0.000	0.000
	0.20	0.038	0.004	0.026	-0.004	0.006
	0.50	0.200	0.009	0.067	-0.013	0.014
	1.00	0.500	0.008	0.094	0.008	0.014
	2.00	0.800	-0.009	0.072	0.029	0.020

3.5 適用例

3.5.1 MEGA Study

本節では、提案する尺度 PCS および最頻値を用いた代替性の評価方法を、日本で初めて行われたプラバスタチンの CHD 予防効果を評価した無作為化比較試験である MEGA study (Nakamura et al., 2006) に適用した結果を示す。8,214 例が、対照群 (食事制限のみ; $X = x_0$) もしくはプラバスタチン群 (食事制限とプラバスタチン投与; $X = x_1$) に無作為に割りつけられ、1, 3, 6ヵ月時とそれに続く 6ヵ月毎に、脂質値が検査された。代替エンドポイントの候補 S は脂質値の連続量とした。具体的には、LDL-C, non-HDL-C の連続量を S とした。臨床エンドポイント Y は、初発の CHD イベント発症までの時間とした。無作為化後の 3, 6, 12ヵ月時の脂質値の検査値がある 7,265 例を対象に解析した。

MEGA study では、脂質値は正規分布にしたがうこと (Teramoto et al., 2010)、また、CHD イベントは少ないこと (対照群 5.0/1000 人年; プラバスタチン群 3.3/1000 人年) (Nakamura et al., 2006) から、 S は $X = x (x \in \{x_1, x_0\})$ を与えた下で正規分布にしたがい、CHD イベントの発症は少ないと仮定した。これらの仮定のもと、以下に示す S の X への線形モデルと Y の Cox 比例ハザードモデルを用いた。

$$S = \alpha_{s \cdot x} + \alpha_{sx}x + \varepsilon_{s \cdot x} \quad (3.13)$$

$$\lambda(y|x, s) = \lambda_{xs}(y) \exp(\alpha_{yx \cdot s}x + \alpha_{ys \cdot x}s) \quad (3.14)$$

ここで、 $\varepsilon_{s \cdot x}$ は、平均 0、分散 $\sigma_{ss \cdot x}$ の正規分布にしたがう誤差項とし、 $\lambda_{xs}(y)$ は、式 (3.14) の Cox 比例ハザードモデルにおけるベースラインハザードとした。同様の表記を他のベースラインハザードにも用いる。このとき、式 (3.7) と式 (3.8) をそれぞれ CP と NCP の評価に利用したとき、付録 3.7 節に示すように、CP と NCP はそれぞれ、 $\alpha_{sx}\alpha_{ys \cdot x}$ および $\alpha_{yx \cdot s}$ と表すことができる。

表 3.4: MEGA study の要約統計量と回帰係数の推定値

		脂質値の推移		回帰係数の推定値	
		対照群	プラバスタチン群	$\alpha_{ys \cdot x}$	$\alpha_{yx \cdot s}$
		平均 (標準偏差)		推定値 (標準誤差)	
LDL-C (0.01mg/dL)	開始時	1.557 (0.184)	1.561 (0.180)		
	3ヵ月時	1.530 (0.237)	1.261 (0.223)*	0.091 (0.549)	-0.315 (0.260)
	6ヵ月時	1.538 (0.252)	1.273 (0.237)*	0.451 (0.563)	-0.219 (0.262)
	12ヵ月時	1.535 (0.242)	1.274 (0.234)*	0.700 (0.587)	-0.154 (0.265)
Non-HDL-C (0.01mg/dL)	開始時	1.850 (0.186)	1.850 (0.188)		
	3ヵ月時	1.816 (0.245)	1.525 (0.245)*	0.593 (0.581)	-0.163 (0.275)
	6ヵ月時	1.824 (0.264)	1.542 (0.260)*	1.001 (0.568)†	-0.045 (0.272)
	12ヵ月時	1.813 (0.251)	1.536 (0.259)*	1.077 (0.599)†	-0.026 (0.276)

†: p 値 < 0.10; *: p 値 < 0.05

回帰係数 α_{sx} , $\alpha_{ys \cdot x}$, $\alpha_{yx \cdot s}$ の推定に最尤法を用いることで, PCS と PTE はそれぞれ, $\hat{\alpha}_{sx}^2 \hat{\alpha}_{ys \cdot x}^2 / (\hat{\alpha}_{sx}^2 \hat{\alpha}_{ys \cdot x}^2 + \hat{\alpha}_{yx \cdot s}^2)$ および $1 - \hat{\alpha}_{yx \cdot s} / \hat{\alpha}_{yx}$ によって評価される. ここで, $\hat{\alpha}_{yx}$ は $\lambda(y|x) = \lambda_x(y) \exp(\alpha_{yx}x)$ に基づいて推定される. PIG は, 式 (3.14) と $\lambda(y) = \lambda_\phi(y)$ を比較する尤度比検定統計量 $LRT(S, X : 1)$, および $\lambda(y|s) = \lambda_s(y) \exp(\alpha_{ys}s)$ と $\lambda(y) = \lambda_\phi(y)$ を比較する尤度比検定統計量 $LRT(S : 1)$ を通じて評価される. そして, 3.2.2 節で述べたように, 代替性の水準を判断するためのカットオフ値 (表 3.1) を用いる際は, PCS の点推定値ではなく, その信頼区間の上限と下限を用いて代替性を評価する. この時, 最尤法を用いて代替性の評価をする場合は, 代替性評価尺度の信頼区間として, 生データに基づくブートストラップ法による 1000 個の最尤推定値の 2.5% 点と 97.5 % 点を用いる. また, 最頻値を用いて代替性を評価する場合は, 2 段階ブートストラップ法 (Davison and Hinkley, 1997) に基づき, 代替性評価尺度の 1000 個の最尤推定値からリサンプリングして得られる最頻値を 1000 個用意し, その 2.5% 点, および 97.5% 点を用いる.

表 3.4 の 3 列目と 4 列目に, LDL-C と non-HDL-C の各時点の要約統計量を群ごとに示す. 対照群とプラバスタチン群の間に有意な差が認められることから, 脂質値に対する治療効果があることがわかる (すなわち, $X \perp\!\!\!\perp S$). 加えて, 表 3.4 の 5 列目と 6 列目に, 最尤法による式 (3.13) と式 (3.14) の回帰係数 $\alpha_{ys \cdot x}$ と $\alpha_{yx \cdot s}$ の推定値を示す. いずれも, 有意水準

3.5. 適用例

表 3.5: MEGA study での代替性の評価尺度の推定値: S 連続量, Y CHD イベント発症までの時間

		推定値 (2.5%点, 97.5 %点)		
(a) 最尤法による代替性評価				
		PTE	PIG	PCS
LDL-C	3ヵ月時	0.381 (-1.684, 3.107)	0.804 (0.001, 0.996)	0.275 (0.000, 0.982)
	6ヵ月時	0.646 (-1.911, 4.929)	0.956 (0.015, 0.999)	0.768 (0.000, 0.995)
	12ヵ月時	0.832 (-2.285, 6.417)	0.992 (0.065, 0.999)	0.961 (0.001, 0.997)
Non-HDL-C	3ヵ月時	0.930 (-2.024, 5.453)	0.999 (0.030, 1.000)	0.994 (0.001, 0.998)
	6ヵ月時	1.169 (-2.930, 7.899)	0.995 (0.202, 1.000)	0.980 (0.010, 0.999)
	12ヵ月時	1.230 (-3.513, 8.131)	0.991 (0.146, 1.000)	0.966 (0.012, 1.000)

(b) 最頻値を用いた代替性評価				
		PTE	PIG	PCS
LDL-C	3ヵ月時	0.293 (0.039, 0.565)	0.551 (0.536, 0.566)	0.052 (0.001, 0.079)
	6ヵ月時	0.702 (0.410, 0.984)	0.698 (0.684, 0.710)	0.084 (0.039, 0.376)
	12ヵ月時	0.978 (0.569, 1.366)	0.777 (0.766, 0.788)	0.327 (0.120, 0.513)
Non-HDL-C	3ヵ月時	0.871 (0.519, 1.180)	0.797 (0.786, 0.807)	0.483 (0.277, 0.841)
	6ヵ月時	1.266 (0.735, 1.755)	0.866 (0.859, 0.874)	0.814 (0.569, 0.978)
	12ヵ月時	1.356 (0.814, 1.961)	0.868 (0.860, 0.876)	0.851 (0.725, 0.993)

5%の下では有意ではないが、有意水準 10%の下では、6, 12ヵ月時の non-HDL-C の $\alpha_{yS,x}$ が有意である。LDL-C については、代替エンドポイントとして疑わしいという結果が最近得られており (Barter et al., 2007; Psaty and Lumley, 2008), これらのことが裏づけられたと考えられる。他方で、non-HDL-C については、良い CHD イベントの予測能をもつことが最近の研究で示されており (Boekholdt et al., 2012; Liu et al., 2006; Ingelsson et al., 2007), こちらも上記の結果を示唆するものであった。これらの考察をふまえると、脂質値 S の代替性については再検討が必要であると考えられる。

表 3.5 (a) と (b) に、最尤法による代替性評価と最頻値を用いた代替性評価の結果を示す。表 3.5 (a) より、最尤法によって代替性を評価したとき、PTE の 2.5%点と 97.5 %点に基づく範囲は 範囲 $[0, 1]$ を含むことがわかる。他方で、PIG と PCS の 2.5%点と 97.5 %点は、 $[0, 1]$ に含まれている。ただし、2.5%点と 97.5 %点に基づく範囲が最も狭い場合であっても (0.202, 1.000) と治療効果をどの程度 S によって捕捉されているかを判断するにはあまりにも広い。

表 3.5 (b) より, 最頻値を用いて代替性を評価したとき, 代替性の評価尺度それぞれの 2.5%点と 97.5 %点に基づく範囲は, 最尤法に比べて狭くなっているが, 12ヵ月時の LDL とすべての時点の non-HDL-C について, PTE の 97.5 %点は 1 を超えていることがわかる. 他方で, 3, 6ヵ月時の LDL-C について, PCS の 97.5 %点は 0.500 を下回っていることから, 表 3.1 を参照すると, LDL-C の代替性の水準は “poor” であると判断することができる. 加えて, 12ヵ月時の non-HDL-C について, PCS の 2.5 %点は 0.634 を超えているが, 1 には届いていないことから, 12ヵ月時の non-HDL-C の代替性の水準は “moderate,” “substantial,” もしくは “almost perfect” であると判断することができる. さらに, すべての時点において, non-HDL-C の 2.5 %点は, LDL-C の 97.5 %点を超えていることから, non-HDL-C は LDL-C より優れた CHD イベントに対する代替エンドポイントであることがわかる. HRM 法を用いた最頻値による PCS の結果は, 既存の医学的な解釈である, LDL-C は代替エンドポイントとしては疑わしく (Barter et al., 2007; Psaty and Lumley, 2008), non-HDL-C が CHD イベントに対する良い代替エンドポイントになりうることを支持している.

3.5.2 ARMD Study

本章で提案する代替性の評価尺度 PCS を, ARMD study (Pharmacological therapy for macular degeneration study, 1997) に適用する. この試験の目的は, プラセボ ($X = x_0$) と インターフェロン α の高用量 ($X = x_1$) を加齢黄斑変性患者に投与した効果を比較することである. 加齢黄斑変性患者 481 例が, プラセボ, インターフェロン α -1.5 百万国際単位 (million international units : MIU), インターフェロン α -3 MIU, もしくは インターフェロン α -6 MIU に無作為に割りつけられた. 臨床エンドポイント Y は, 1 年経過時の視力検査 (治療開始時に比べ少なくとも 3 本の文字列が見えなくなること ($Y = y_1$); それ以外 ($Y = y_0$)) であり, 代替エンドポイントの候補 S は, 6ヵ月経過時の視力検査 (治療開始時に比べ少なくとも 2 本の文字列が見えなくなること ($S = s_1$); それ以外 ($S = s_0$)) である. 表

表 3.6: ARMD study データ (Buyse and Molenberghs, 1998)

X	S	Y	
		y_1	y_0
x_1	s_1	38	9
	s_0	9	31
x_0	s_1	30	8
	s_0	9	56

3.6 に, Buyse and Molenberghs (1998) が代替エンドポイントの候補 S の妥当性を評価した際のデータを示す.

本節では, PE と PCS に関して, 式 (3.3) の h を恒等関数とし, 式 (3.3) の g_{x_0} と式 (3.6) と (3.7) の $E(Y|x_0, s)$ を $\text{pr}(Y = 1|x_0, s)$ に置き換え, そして, 式 (3.3) の g_{x_1} と式 (3.7) の $E(Y|x_1, s)$ を共に $\text{pr}(Y = 1|x_1, s)$ に置き換える. 加えて, PTE と PIG に関して, 式 (3.1), (3.2), (3.4) と (3.5) においてロジットリンク関数を用いる. そして, 3.2.2 節で述べたように, 代替性の水準を判断するためのカットオフ値 (表 3.1) を用いる際は, PCS の点推定値ではなく, その信頼区間の上限と下限を用いて代替性を評価する. この時, 最尤法を用いて代替性の評価をする場合は, 代替性評価尺度の信頼区間として, 表 3.6 のデータに基づく多項分布を使ったパラメトリックブートストラップ法により, 1000 個の最尤推定値を求め, その 2.5%点と 97.5 %点を用いる. また, 最頻値を用いて代替性を評価する場合は, 代替性評価尺度の 1000 個の最尤推定値からリサンプリングして得られる最頻値を 1000 個用意し, その 2.5%点, および 97.5%点を用いる.

表 3.7 の 2 列目と 3 列目にそれぞれ, 最尤法による代替性評価と最頻値を用いた代替性評価による結果を示す. 2 列目の結果より, 最尤法によって代替性を評価したとき, PTE の 2.5%点と 97.5 %点に基づく範囲は範囲 $[0, 1]$ を含むことがわかる. また, PE と PCS の 2.5%点と 97.5 %点はそれぞれ, $(0.191, 1.797)$ と $(0.075, 0.999)$ であるが, 治療効果を S がどの程度捕捉しているかを判断するには広すぎる. 3 列目の結果より, 最頻値を用いて代替性を評価したとき, 代替性の評価尺度それぞれの 2.5%点と 97.5 %点は範囲 $[0, 1]$ に含まれ

表 3.7: ARMD Study での代替性の評価尺度の推定値

代替性の評価尺度	推定値 (2.5%点, 97.5%点)	
	最尤法による 代替性評価	最頻値を用いた 代替性評価
PTE	0.445 (-0.355, 2.160)	0.245 (0.203, 0.451)
PE	0.690 (0.191, 1.797)	0.485 (0.410, 0.666)
PIG	0.989 (0.898, 1.000)	1.000 (1.000, 1.000)
PCS	0.832 (0.075, 0.999)	0.471 (0.326, 0.798)

ており、またそれらは、最尤法の結果より狭いことがわかる。加えて、提案手法による PTE の 97.5% 点は 0.500 を下回っている。同様に、PE と PCS の 2.5% 点と 97.5% 点はそれぞれ、(0.410, 0.666) と (0.326, 0.798) であるため、*S* の代替性の水準は、“moderate” と “poor” の間であると判断できる。この結果は、Buyse and Molenberghs (1998) の見解 “the loss of two lines of vision at 6 months (*S*) is an incomplete surrogate for the loss of three lines of vision at 1 year (*Y*)” (ただし、(*S*) と (*Y*) は本章において追記した) を反映したものと考えられる。他方で、PIG の結果に着目すると、*S* の代替性は “perfect” と判断される可能性がある。これは、臨床的な観点 (Csaky et al., 2008; Verteporfin in photodynamic therapy study group, 2001) から考えると、PIG は代替性の水準を高め評価している可能性があることを意味している。他の尺度に比べ、PIG が代替性を過大に評価することの理由として、次の二つが考えられる。一つ目の理由として、PE および PCS を PIG と比較した時、3.2 節と 3.3 節に記載したように、PIG は *X* と *S* の従属関係の程度を適切に反映していない可能性があることが挙げられる。二つ目の理由として、PTE を PIG と比較した時、この 2 つの尺度が考慮している条件付き従属関係が異なることが挙げられる (Qu and Case, 2007; Freedman et al., 1992)。具体的には、PTE は、*X* を与えた下での *S* と *Y* の従属関係を考慮し、PIG は *S* を与えた下での *X* と *Y* の従属関係を考慮している。なお、上記の二つの理由についての詳細な考察は、付録 3.7 節を参照されたい。

3.6 まとめ

本章では、統計的関連性を用いた新たな代替性評価尺度 PCS, および最頻値を用いた代替性の評価方法を提案し、生物統計分野の研究者や実務家によって指摘されていた以下の三つの問題点を解決していることを確認した。

一つ目の問題点は、既存の代替性評価尺度の多くは特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとることである。この問題に対し、治療効果のうち代替エンドポイントの候補が捕捉する部分と、捕捉しない部分への分解, およびその二乗和を用いて PCS を定義することで解決した。二つ目の問題点は、既存の代替性評価尺度は、代替性の水準を判断するための適当なカットオフ値を定義できないことであった。これに対し、PCS はその微分を用いて、代替性の水準を判断するためのカットオフ値を与えられることを示した。ここで、最終的な代替性の判断は、このカットオフ値によるものだけではなく、臨床的かつ生物学的な観点に基づいてなされるべきであることを強調しておく。最後の三つ目の問題点は、代替性評価尺度の多くは、そのばらつきが大きく、そのため信頼区間が許容できないほど広く代替性の評価が困難であることであった。この問題を解決するため、推定された代替性評価尺度のブートストラップ分布 (Efron and Tibshirani, 1993) を用いた HRM 法 (Bickel, 2002) に基づく代替性の評価方法を提案した。なお、本論文で提案した最頻値を用いた代替性の評価方法は、尺度そのものの推定精度を向上させているわけではなく、多重共線性の問題の回避を狙って、推定された尺度の中で最も頻繁に出現するという意味で適切な値を推定値としていることを強調しておく。

本章では、数値実験をとおしてこれらの問題点を PCS と最頻値を用いた代替性の評価方法が解決していることを確認した。さらに二つの無作為化比較試験に適用することで、提案した代替性評価尺度 PCS および HRM 法に基づく代替性の評価方法が、臨床家でも理解しやすい性質を備えた方法であることを示した。

3.7 付録

3.7.1 治療効果が比によって定義される場合の CP と NCP の定式化

最初に, X の Y への治療効果を $TE = \log \{E(Y|x_1)/E(Y|x_0)\}$ と定義するとき, CP と NCP はそれぞれ, 次のように定めることができる.

$$CP = \log \left(\frac{\sum_{D_s} E(Y|x_0, s) \text{pr}(s|x_1)}{E(Y|x_0)} \right) \quad (3.15)$$

$$NCP = \log \left(\frac{E(Y|x_1)}{\sum_{D_s} E(Y|x_0, s) \text{pr}(s|x_1)} \right) \quad (3.16)$$

また, 治療効果として $TE = \log \{\text{pr}(Y > y|x_1)/\text{pr}(Y > y|x_0)\}$ に興味がある場合は, CP と NCP はそれぞれ, 式 (3.16) の $E(Y|x, s)$ と $E(Y|x)$ を $\text{pr}(Y > y|x, s)$ と $\text{pr}(Y > y|x)$ ($x \in D_X, s \in D_S$) に置き換えて定めることができる. 同様の手続きを, 分布関数によって定義される差 $\text{pr}(Y > y|x_1) - \text{pr}(Y > y|x_0)$ を治療効果と定義したときにも適用可能である.

続いて, X と Y が共に 2 値変数で, X の Y への治療効果を対数オッズ比

$$TE = \log \left\{ \frac{\text{pr}(y_1|x_1)\text{pr}(y_0|x_0)}{(\text{pr}(y_1|x_0)\text{pr}(y_0|x_1))} \right\}$$

によって定義するとき, CP と NCP はそれぞれ, 次のように定めることができる.

$$CP = \log \left(\frac{\text{pr}(y_1|x_1) \sum_{D_s} \text{pr}(y_0|x_1, s) \text{pr}(s|x_0)}{\text{pr}(y_0|x_1) \sum_{D_s} \text{pr}(y_1|x_1, s) \text{pr}(s|x_0)} \right) \quad (3.17)$$

$$NCP = \log \left(\frac{\text{pr}(y_0|x_0) \sum_{D_s} \text{pr}(y_1|x_1, s) \text{pr}(s|x_0)}{\text{pr}(y_1|x_0) \sum_{D_s} \text{pr}(y_0|x_1, s) \text{pr}(s|x_0)} \right) \quad (3.18)$$

また、治療効果として分布関数によって定義される治療効果

$$\text{TE} = \log \left\{ \frac{\text{pr}(Y > y|x_1)\text{pr}(Y < y|x_0)}{\text{pr}(Y > y|x_0)\text{pr}(Y < y|x_1)} \right\}$$

に興味がある場合は、CP と NCP はそれぞれ、式 (3.18) の $\text{pr}(y_1|x, s)$, $\text{pr}(y_0|x, s)$, $\text{pr}(y_1|x, s)$ および $\text{pr}(y_0|x, s)$ を、 $\text{pr}(Y > y|x)$, $\text{pr}(Y < y|x)$, $\text{pr}(Y > y|x, s)$ および $\text{pr}(Y < y|x, s)$ ($x \in D_X, s \in D_S$) に置き換えて定めることができる。

最後に、ハザードモデルにおいて CP と NCP を定式化するため、 $\lambda(y|x, s) = \lambda_{xs}(y) \exp\{k_1(x) + k_2(x)s\}$ (ただし $k_1(x)$ と $k_2(x)$ は X の関数) とし、さらに $X = x$ を与えた下での S の確率分布関数 $f(s|x)$ に基づく S のモーメント母関数を $m_s(\theta; x) = E(e^{\theta s}|x)$ とする。このとき、次の式を考える。

$$\begin{aligned} \lambda\{y|x, (s, x')\} &= \lambda_{xs}(y) \exp\{k_1(x)\} \\ &\times \frac{\int_{D_s} \exp\{k_2(x)s\} \exp[-\Lambda_{xs}(y) \exp\{k_1(x) + k_2(x)s\}] f(s|x') ds}{\int_{D_s} \exp[-\Lambda_{xs}(y) \exp\{k_1(x) + k_2(x)s\}] f(s|x') ds} \end{aligned}$$

なお、 $\Lambda_{xs}(y)$ は累積ハザード $\Lambda_{xs}(y) = \int_0^y \lambda_{xs}(t) dt$ とする。 Y のイベント発現が低いという仮定の下では、VanderWeele (2011) にしたがって $\Lambda_{xs}(y)$ は非常に小さな値であることから $\Lambda_{xs}(y) \simeq 0$ とみなせる。この時に、次の関係が導かれる。

$$\begin{aligned} \lambda\{y|x, (s, x')\} &\simeq \lambda_{xs}(y) \exp\{k_1(x)\} \frac{\int_{D_s} \exp\{k_2(x)s\} f(s|x') ds}{\int_{D_s} f(s|x') ds} \\ &= \lambda_{xs}(y) \exp\{k_1(x)\} m_s\{k_2(x); x'\} \end{aligned} \quad (3.19)$$

また、 $\text{CP} = \log [\lambda\{y|x_1, (s, x_1)\}] - \log [\lambda\{y|x_1, (s, x_0)\}]$ と $\text{NCP} = \log [\lambda\{y|x_1, (s, x_0)\}] - \log [\lambda\{y|x_0, (s, x_0)\}]$ と定式化すると、それぞれ $\text{CP} \simeq \log [m_s\{k_2(x_1); x_1\}] - \log [m_s\{k_2(x_1); x_0\}]$

と $\text{NCP} \simeq k_1(x_1) - k_1(x_0) + \log [m_s\{k_2(x_1); x_0\}] - \log [m_s\{k_2(x_0); x_0\}]$ と展開できる. ゆえに, $\text{TE} \simeq k_1(x_1) - k_1(x_0) + \log [m_s\{k_2(x_1); x_1\}] - \log [m_s\{k_2(x_0); x_0\}]$ となる. 3.5.1 節にて, $k_1(x) = \alpha_{yx.s}x$, $k_2(x) = \alpha_{ys.x}$ および $m_s(\theta; x) = \exp(\alpha_{sx}x\theta + \theta^2\sigma_{ss.x}/2)$ とすることで, $\text{CP} \simeq \alpha_{sx}\alpha_{ys.x}(x_1 - x_0)$, $\text{NCP} \simeq \alpha_{yx.s}(x_1 - x_0)$, そして $\text{TE} \simeq (\alpha_{yx.s} + \alpha_{ys.x}\alpha_{sx})(x_1 - x_0)$ を得る. x_1, x_0 をそれぞれ 1, 0 に置き換えることで, 3.5.1 節の CP と NCP が得られる.

3.7.2 数値実験による平均二乗誤差の評価

3.4 節の数値実験において, 平均二乗誤差の平方根を用いて最尤法と提案手法を比較した結果を表 3.8 に示す. 具体的な評価の記述は 3.4 節を参照されたい.

表 3.8: 数値実験における平均二乗誤差の平方根

設定 1				
	σ^2	真値	最尤法による 代替性評価	最頻値を用いた 代替性評価
PTE	0.01	1.000	0.328	0.070
	0.10	1.000	0.116	0.051
	1.00	1.000	0.075	0.027
	2.00	1.000	0.069	0.017
	4.00	1.000	0.071	0.018
PCS	0.01	1.000	0.207	0.000
	0.10	1.000	0.022	0.000
	1.00	1.000	0.009	0.000
	2.00	1.000	0.008	0.000
	4.00	1.000	0.008	0.000
設定 2				
	$(\alpha_{yx \cdot s}, \alpha_{ys \cdot x})$	真値	最尤法による 代替性評価	最頻値を用いた 代替性評価
PTE	(0.00,1.00)	1.000	0.071	0.053
	(0.10,0.90)	0.900	0.069	0.020
	(0.30,0.70)	0.700	0.058	0.009
	(0.50,0.50)	0.500	0.048	0.013
	(1.00,0.00)	0.000	0.032	0.007
PCS	(0.00,1.00)	1.000	0.009	0.000
	(0.10,0.90)	0.988	0.021	0.012
	(0.30,0.70)	0.845	0.070	0.019
	(0.50,0.50)	0.500	0.093	0.040
	(1.00,0.00)	0.000	0.002	0.000
設定 3				
	α_{sx}	真値	最尤法による 代替性評価	最頻値を用いた 代替性評価
PTE	0.00	0.000	0.065	0.029
	0.20	0.167	0.048	0.011
	0.50	0.333	0.045	0.014
	1.00	0.500	0.049	0.008
	2.00	0.667	0.050	0.025
PCS	0.00	0.000	0.008	0.000
	0.20	0.038	0.027	0.007
	0.50	0.200	0.068	0.018
	1.00	0.500	0.095	0.016
	2.00	0.800	0.072	0.035

3.7.3 ARMD study での PIG の値に関する考察

3.5.2 節の ARMD study での代替性評価において、PTE, PE, PCS に比べ、PIG は代替性を過大に評価している可能性があり、その理由として次の二点を指摘していた。一つ目は、PE および PCS とは異なり、 X と S の従属関係の程度が PIG には適切に反映されていない可能性があること、二つ目は、PTE と PIG を比較した時に、この二つの尺度が考慮している条件付き従属関係が異なることである。ここでは、これら二つの理由について、詳細な考察を加えることにする。

一つ目の PIG が代替性を過大に評価している可能性がある点については、すでに 3.2 節と 3.3 節、および 3.4 節で述べたように、PE と PCS は、 S が $X \perp\!\!\!\perp Y | S$ を完全に満たす場合を除いて、 X と S の従属関係の影響を受けるのに対し、PIG は、 X と S の従属関係の影響を十分には受けない可能性がある。今回の ARMD study においては、仮説 $X \perp\!\!\!\perp Y | S$ は棄却されず (逸脱度 = 1.324, 自由度 = 2, p 値 = 0.516), PE と PCS は X と S の従属関係の影響を受けているものと考えられる。これに対し、PIG は X と S の従属関係の影響を十分に受けず、そして $\text{pr}(y|x, s)$ と $\text{pr}(y|s)$ が比較的近い値をとっているために、表 3.7 のように PIG は高い値をとったと考えられる。なお、 X と S の従属関係の影響を排除した極端な状況 ($\text{pr}(s_1|x_1)$ と $\text{pr}(s_0|x_0)$ の両方を 1 に置き換えた状況) を考えると、PE と PCS はそれぞれ 0.972 と 0.999 となり、これらは表 3.7 の PIG の値に近い値となっていることがわかる。このことから、PIG は X と S の従属関係の影響を適切に反映せずに代替性を過大に評価していることが確認できる。

二つ目の理由として、3.5.2 節において、PTE と PIG が考慮している条件付き従属関係が異なることを指摘していた。このことを直接考察する前に、両者の尺度に用いられている統計量も異なっていることに注意が必要である。具体的には、PTE に用いられている統計量は回帰モデルの回帰係数であるが、PIG には尤度比検定統計量が用いられている。ここでは二つの尺度が考慮している条件付き従属関係の違いに着目するため、同じ統計量(こ

ここでは回帰係数)を用いて二つの尺度を比較することにする. ARMD study では, x の回帰係数 ($\alpha_{yx}, \alpha_{yx.s}$) と s の回帰係数 ($\alpha_{ys}, \alpha_{ys.x}$) は, それぞれ 0.657, 0.364 と 2.962, 2.919 である (これらは 3.5.2 節と同様にロジットリンク関数を用いたモデルの回帰係数の最尤推定値として得られる). このとき, PTE が考慮する条件付き従属関係 (X を与えた下での S と Y の従属関係) に対して, x の二つの回帰係数を用いた PTE は 0.445 と算出される. 一方で, PIG が考慮する条件付き従属関係 (S を与えた下での X と Y の従属関係) に対して, s の二つの回帰係数を用いた仮の尺度として $\alpha_{ys.x}/\alpha_{ys}$ を考えたとき, その推定値は 0.986 となる. この値は, 表 3.7 の PIG に近い値であることが確認できる. すなわち, 同じ統計量を用いた場合であっても, 考慮している条件付き従属関係の差異によって, 両者は異なる値をとることがわかる.

以上, 二つの理由から, PTE, PE, PCS に比べ, PIG は ARMD study における代替性を高めに評価していることが確認できる.

第4章 自然な因果効果を用いた治療効果 に対する代替性の評価尺度

本章では, 第3章で提案した代替性評価尺度を, 統計的因果推論の観点から再定義し, 自然な因果効果を用いた代替性評価尺度 Causal PCS(C-PCS) を提案する¹.

4.1 はじめに

第3章において検討した代替性の評価尺度は, 統計的関連性に基づいた尺度であった. ただ, そこでの検討では, 評価する代替エンドポイントの候補が“病気の治癒のしくみにとつて的を射たものであるかどうかがよく吟味されていなければならない”(高橋, 1967)という要件を満たしているかを確認することは困難であった. すなわち, 治療, 代替エンドポイントそして臨床エンドポイントを取り巻く因果的な構造を踏まえた代替性の評価は, 第3章での議論では明らかではなかった. これに対応するには, 第2章で述べた Pearl 流の統計的因果推論の議論をふまえたアプローチが有用であると考えられる. 実際に, このアプローチに沿った自然な因果効果を用いた代替性評価尺度がいくつか提案されている (Taylor et al., 2005; Albert, 2008). しかし, これら既存の尺度は特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとる問題点を抱えている (Taylor et al., 2005; Albert, 2008).

そこで本章では, この問題を解決するために, 自然な因果効果を用いた治療効果に対する代替性の評価尺度 C-PCS を提案する. 4.2 節では, 準備として記号の表記, 潜在反応アプローチと自然な因果効果を扱う際の仮定や条件, 代替性の評価方法, そして自然な因果効

¹本章の内容は, Kobayashi and Kuroki (2014b) をまとめたものである.

果を用いた既存の代替性評価尺度の定式化やその特徴を示す。4.3節では、第3章で提案した代替性評価尺度を、統計的因果推論の観点から再定義し、自然な因果効果を用いた代替性評価尺度 **C-PCS** を提案する。4.4節では、簡単な数値例をとおして、提案する **C-PCS** が既存の尺度の問題点を解決し、常に範囲 $[0, 1]$ 内の値をとることを示す。最後に、4.5節では、**C-PCS** を無作為化比較試験 **ORIENT** (Imai et al., 2011) に適用し、代替エンドポイント候補の代替性を適切に評価できることを示す。本章の結果を用いることで、治療、代替エンドポイント、そして臨床エンドポイントを取り巻く因果メカニズムを視覚的に表現した上で、代替エンドポイントの候補によってどの程度治療効果を捕捉できるかを評価可能となることが期待される。

4.2 準備

X, S , および Y をそれぞれ、治療、代替エンドポイントの候補、臨床エンドポイントを表す変数とする。 D_X を X がとる値の領域とし、他の領域についても同様の表記を用いる。また、 x, s , および y を、それぞれ変数 X, S, Y の実現値とする。加えて、 $\text{pr}(X = x) = \text{pr}(x)$ と $\text{pr}(S = s|X = x) = \text{pr}(s|x)$ をそれぞれ、 $X = x$ の周辺確率、 $X = x$ を与えた下での $S = s$ の条件付き確率とする。さらに、 $E(Y|x, s)$ を $(X, S) = (x, s)$ を与えた下での Y の条件付き期待値とし、他の確率や期待値についても同様の表記を用いる。

続いて、本章では、潜在反応アプローチ (Pearl, 2009a; Rubin, 2006) に基づいて新たな代替性評価尺度を提案する。対象者 i が治療 $X = x_1$ を受けた場合に生じるであろう潜在反応変数を $Y_{x_1}(i)$ 、治療 $X = x_0$ を受けた場合に生じるであろう潜在反応変数を $Y_{x_0}(i)$ と記す。 $Y_{x,s}(i)$ についても同様に定義する。ここで、第2章で述べた自律的なデータ生成過程が成り立つこと、および **No Interference between Subjects Assumption** (Cox, 1958) を仮定する。また、本章では、入れ子型の潜在反応変数 $Y_{x'S_x}(i)$ を利用する。ここで、 $Y_{x'S_x}(i)$ は“仮に $X = x$ を受けた場合に生じるであろう S の値 $S_x(i)$ の状態を保持したまま、“仮に $X = x'$

であったならば”生じるであろう $Y(i)$ の値を意味する. 一般的な潜在反応アプローチでは, 個人レベルでの潜在反応変数を定義したうえで, これを対象集団からの無作為抽出とみなすことで, 潜在反応変数を確率変数と考え統計的推測を行う. 本章でもこれにしたがうことにする.

これらの準備の下で, X の Y への治療効果 (TE) を, $TE = E(Y_x) - E(Y_{x'})$ とした場合に, Pearl (2009a) にしたがって自然な直接効果 (NDE) と間接効果 (NIE) を以下のように定義する.

$$NDE(x) = E(Y_{xS_x}) - E(Y_{x'S_x}), \quad NIE(x) = E(Y_{x'S_x}) - E(Y_{x'S_{x'}})$$

$$NDE(x') = E(Y_{x'S_{x'}}) - E(Y_{x'S_x}), \quad NIE(x') = E(Y_{xS_x}) - E(Y_{xS_{x'}})$$

($x \neq x'$). この時, TE の NDE と NIE への分解 ($TE = NIE(x) + NDE(x) = NIE(x') + NDE(x')$) が成立する (Pearl, 2009a).

一般に, NDE と NIE は実験研究だけで評価することはできず, S と Y の間の交絡因子の同定も必要となる. NDE や NIE の識別可能条件として以下が知られている (VanderWeele, 2009b).

$$Y_{xs}, S_x, Z \perp\!\!\!\perp X \quad (4.1)$$

$$Y_{xs} \perp\!\!\!\perp S | \{X, Z\} \quad (4.2)$$

($\forall x, s$). この識別可能条件が成り立っている時の変数間の関係をグラフで表現したものを図 4.1 に与える. ここに, Z は S と Y の間の同定された交絡因子の集合である.

Wang and Taylor (2002) は, 代替性評価尺度は, S が完全な代替性を持つときに 1 をとり, 代替性がないときに 0, そして, 部分的な代替性を持つときに 0 と 1 の間の値をとるべきであると述べている. 本章でも, これを踏襲し, NIE を用いた代替性の評価尺度が満たすべき要件として, $NDE=0$ かつ $NIE=TE \neq 0$ のとき 1, $NIE=0$ かつ $NDE=TE \neq 0$ のとき 0,

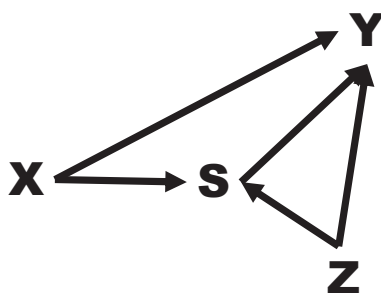


図 4.1: 問題設定

そしてこれ以外のとき範囲 $[0, 1]$ 内の値をとるという 3 つの条件を満たす代替性評価尺度を提案する.

上記の NIE は, 代替エンドポイント S を介する X の Y への効果と解釈できる. したがって, これを ‘治療効果を S がどの程度捕捉するか’ を評価する尺度に利用することは自然な考えである. 実際に, Joffe and Greene (2009) は, 代替エンドポイントの評価手法のレビューの中で, 間接効果を用いた代替性の評価方法を主要な研究の一つに据えている (彼らは, これを causal effect paradigm の一つに位置付けている).

Taylor et al. (2005) は, 単純に NIE を TE で除した, proportion indirect (PI) を代替性の評価尺度として提案している.

$$PI = \frac{NIE}{TE} \quad (4.3)$$

PI はその定式化が, 割合ではなく 2 つの効果の比であること, さらに PI の分母が TE であるため, PI も第 3 章で述べた PTE (Freedman et al., 1992) と同様の問題点を持つ. すなわち, 範囲 $[0, 1]$ 外の値をとる (Li et al., 2001; Wang and Taylor, 2002) という問題点が, PI にも生じる可能性がある. このことは, 4.4 節の仮想的な数値例や 4.5 節の適応例の中で確認していく. なお, 式 (4.3) の分子には, $NIE(x)$, $NIE(x')$ の二通りが考えられることに注意する.

なお、NIE の値そのものを代替性の尺度として利用することも考えられているが (Joffe and Greene, 2009), これは TE の型や単位に応じて、NIE の符号・大きさ自体が変わってしまうため、他のデータを用いた結果との比較は困難となり、一般に利用しやすい代替性の尺度とは考えにくい。

PI は 媒介分析 (mediation analysis) の文脈では、‘proportion mediated’ として知られる尺度と同一である (VanderWeele, 2013). そこで、媒介分析における尺度として提案されている R_G および R'_G (Albert, 2008) をここで述べる. 治療 X が 2 値 (x と x') の場合に、 R_G と R'_G は以下となる.

$$R_G = \frac{\left((h[E(Y_x)] - h[E(Y_{xS_{x'}})])^2 \right)^{1/2}}{\left((h[E(Y_x)] - h[E(Y_{x'})])^2 \right)^{1/2}} \quad (4.4)$$

$$R'_G = \frac{\left((h[E(Y_{x'})] - h[E(Y_{x'S_x})])^2 \right)^{1/2}}{\left((h[E(Y_{x'})] - h[E(Y_x)])^2 \right)^{1/2}} \quad (4.5)$$

ここで、 $h(\cdot)$ は単調な関数とする. なお、 $h(\cdot)$ を恒等関数とした場合は、 $R_G = |NIE(x')|/|TE|$, $R'_G = |NIE(x)|/|TE|$ となり、PI の分子と分母それぞれに対し絶対値をとった定式化であることが分かる. Albert (2008) は、上記の尺度を提案する際の議論の中で、PI が “not a true proportion in that it may take values less than 0 and greater than 1” と、その問題点を指摘しているが、 R_G および R'_G 自身は、範囲 $[0, 1]$ 外の値をとるという問題点は解決できていない. 実際に、式 (4.4) と (4.5) の分子と分母はそれぞれ正の値をとるため、 $R_G, R'_G \geq 0$ は満たすが、1 を超える可能性がある.

4.3 自然な因果効果を用いた代替性評価尺度：C-PCS

前節の Albert (2008) の指摘にもあったように、間接効果を用いた既存の代替性評価尺度は、割合としての意味を持たないものが多く、これらを利用して中間変数による代替性の程度を定量的に把握することは困難である. これに対して、第 2 章で提案した PCS は因果

的な尺度ではないために多くの現実問題に適用可能である。このことは、逆に言えば、PCS に因果的な解釈を与えるのは難しいことを意味する。そこで、本章では、処理変数が2値であるとき、統計的因果推論の観点から PCS を改良した代替性評価尺度として、Causal PCS (C-PCS) を提案する。

$$\text{C-PCS} = \alpha \left(\frac{\text{NIE}(x)^2}{\text{NDE}(x)^2 + \text{NIE}(x)^2} \right) + (1 - \alpha) \left(\frac{\text{NIE}(x')^2}{\text{NDE}(x')^2 + \text{NIE}(x')^2} \right) \quad (4.6)$$

ここに、 α は2つの項それぞれの重要性を表すパラメータである ($0 \leq \alpha \leq 1$)。また、それぞれの項は、 X から Y への TE が二通りの方法で直接効果と間接効果に分解できること、すなわち、 $\text{NIE}(x) + \text{NDE}(x)$ と $\text{NIE}(x') + \text{NDE}(x')$ と表現できることに着目したものである。

パラメータ α の値は、興味ある処理群の選択とその重要性に応じて選択されることになる。例えば、 $X = x'$ が比較の基準値であり、それに対する $X = x$ の間接効果を検討する場合には、 $\alpha = 0$ とすることになる。一方で、具体的な判断基準がない場合には、 $\alpha = 1/2$ としたり、 $\alpha = \text{pr}(x)$ とおくといったことも可能である。

4.4 仮想的な数値例

図 4.1 の問題設定の下で、簡単な仮想的な数値例により以下の性質を確認する。

- (1) Taylor et al. (2005) の PI は、範囲 $[0, 1]$ 外の値をとること
- (2) Albert (2008) の R_G は、1 より大きい値をとること
- (3) 提案手法である C-PCS は、範囲 $[0, 1]$ 内の値をとること

なお仮想的な数値例では、簡単のため $X \in (x_0, x_1)$, $S \in (s_0, s_1)$, $Y \in (y_0, y_1)$, かつ $Z = \emptyset$ とする。このとき、 X の Y への TE を $\text{pr}(y_1|x_1) - \text{pr}(y_1|x_0)$ とした場合に、NDE と NIE は次

のように定式化できる.

$$\text{NDE}(x_0) = \sum_{s \in \{s_0, s_1\}} \{\text{pr}(y_1|x_1, s) - \text{pr}(y_1|x_0, s)\} \text{pr}(s|x_0) \quad (4.7)$$

$$\text{NIE}(x_0) = \sum_{s \in \{s_0, s_1\}} \text{pr}(y_1|x_1, s) \{\text{pr}(s|x_1) - \text{pr}(s|x_0)\} \quad (4.8)$$

$$\text{NDE}(x_1) = \sum_{s \in \{s_0, s_1\}} \{\text{pr}(y_1|x_1, s) - \text{pr}(y_1|x_0, s)\} \text{pr}(s|x_1) \quad (4.9)$$

$$\text{NIE}(x_1) = \sum_{s \in \{s_0, s_1\}} \text{pr}(y_1|x_0, s) \{\text{pr}(s|x_1) - \text{pr}(s|x_0)\} \quad (4.10)$$

表 4.1 に, 三つの仮想的な数値例とそれぞれに対する代替性評価指標の算出結果を示す. なお, 表 4.1 には $\text{NDE}(x_1)$, $\text{NIE}(x_1)$ を用いた **C-PCS** および **PI** とこれに対応する R_G の結果を示した. $\text{NDE}(x_0)$, $\text{NIE}(x_0)$ を用いた場合も同様の検討が可能であるが, 冗長を避けるため割愛した. **C-PCS** は式 (4.6) において $\alpha = 1$ かつ $(x, x') = (x_1, x_0)$ とし, 同様に **PI** は式 (4.3) において分子を $\text{NIE}(x_1)$ とし, R_G は式 (4.4) において $(x, x') = (x_1, x_0)$ として算出した.

表 4.1 より次の考察が得られる.

1. 設定 1 では, すべての代替性評価指標が範囲 $[0, 1]$ 内の値をとるのに対し, 設定 2 では, **PI** は -0.12 と範囲 $[0, 1]$ 外の値をとる. 一方で, **C-PCS** と R_G は範囲 $[0, 1]$ 内の値をとる.
2. 設定 3 では, **PI** と R_G はともに 1.20 と範囲 $[0, 1]$ 外の値をとる. 一方で, **C-PCS** は 0.97 と範囲 $[0, 1]$ 内の値をとる.

以上の仮想的な数値例から, 既存の代替性評価尺度とは異なり, **C-PCS** は範囲 $[0, 1]$ 内の値をとるため, 代替性の評価尺度として適切であると解釈できる.

表 4.1: 仮想的な数値例

	$\text{pr}(y_1 x, s)$				$\text{pr}(s_1 x_1)$	$\text{pr}(s_1 x_0)$	C-PCS	PI	R_G
	x_1	x_1	x_0	x_0					
	s_1	s_0	s_1	s_0					
設定 1	0.7	0.4	0.3	0.1	0.7	0.4	0.07	0.21	0.21
設定 2	0.8	0.4	0.3	0.2	0.5	0.6	0.01	-0.12	0.12
設定 3	0.3	0.6	0.4	0.4	0.5	0.7	0.97	1.20	1.20

4.5 適用例

4.5.1 データの概要と代替性評価の必要性

ORIENT (Imai et al., 2011) は、日本および香港の顕性腎症を伴う 2 型糖尿病患者 566 例を対象に、プラセボを対照としてオルメサルタン投与による腎症進展抑制効果や心血管イベント抑制効果を比較した無作為化比較臨床試験である。

ORIENT では、血清 Cr 値の 2 倍化、末期腎不全、死亡から構成される腎複合イベント (主要評価項目) について、オルメサルタン群とプラセボ群で差がみられなかった (オルメサルタン群 116/282 [イベント発現例数/対象例数], 41.1%; プラセボ群 129/284, 45.4%; ハザード比 0.97 [95% 信頼区間 0.75-1.24, $p=0.791$])。一方で、副次評価項目である脳心血管複合イベントはハザード比 0.73 (95% 信頼区間 0.48-1.09, $p=0.126$; 開鍵後に確認された患者背景の不均衡を調整した探索的な検討の結果、ハザード比 0.64 [95% 信頼区間 0.43-0.98, $p=0.039$]) であり、さらに、尿蛋白の変化率、血圧、腎機能の低下速度はオルメサルタン投与群で有意に減少していた。

尿蛋白の変化率、血圧、腎機能低下速度は、腎複合イベントに対する代替エンドポイントの候補に挙げられているが、上記の ORIENT の結果は、それを支持しない (つまり代替性がない) 結果であった。この結果については十分な考察が与えられておらず、代替性の再検討が求められている (Imai et al., 2011, 2013a,b)。

4.5.2 事前検討

本節では、解析の実施に先立ち、以下の事前検討を行う。

1. 前節に記載したとおり、血圧、尿蛋白の変化率、血清 Cr 値の逆数はオルメサルタン投与により有意な減少が確認されている。ただし、血圧と血清 Cr 値の逆数に対する治療効果は、試験期間の早期ではわずかであり、試験期間の後半になって両群間に差があることが確認されている。一方で、尿蛋白の変化率に対する治療効果は、試験期間の早期から後半にわたり一貫して確認されている (Imai et al., 2013a)。代替エンドポイントに求められる実用上の性質として、'臨床エンドポイントより早期に測定できること'(Molenberghs et al., 2005; Stevens et al., 2006) が求められている。平均追跡期間が 3.2 年の ORIENT (Imai et al., 2011) の場合は、治療開始から 1 年以内に測定できる測定値を代替エンドポイントの候補として考えることが妥当と考えられる。この要求にしたがった場合に、血圧と血清 Cr 値の逆数の代替性評価 (すなわち、治療開始から 1 年以内の値を用いた代替性の評価) は、治療開始から 1 年以内の値への治療効果が小さいことをふまえると、代替性は低くなることが想定される。一方で、治療開始から 1 年以内の尿蛋白の変化率については、初期の治療効果が大きいため、適当な代替性があることが想定される。これらの考察をふまえ、次節の解析では、最初に 1 年以内の測定値として、治療開始から 12, 24, 48 週時点を評価対象とする。さらに、この時点を限定することの影響を確認するため、治療期間すべての値を用いた検討も実施する。
2. ORIENT は、アンジオテンシン変換酵素阻害薬 (ACE 阻害薬) による併用治療を条件付きで許容するデザインであった (Imai et al., 2011)。しかし、ORIENT の試験終了直前に報告された ONTARGET 試験 (The ONTARGET Investigators, 2008) によると、ACE 阻害薬の併用時のアンジオテンシン II 受容体拮抗薬 (ARB) の治療効果は小さいことが報告されている。実際に、ORIENT における ACE 阻害薬による併用治療有

集団での結果 (Imai et al., 2013a) は, 腎複合イベントへの治療効果は, ハザード比 1.02 (95% 信頼区間 0.75-1.39, $p=0.891$; オルメサルタン群 80/205 [イベント発現例数/対象例数], 39.0%; プラセボ群 87/209, 41.6%) であり, ONTARGET 試験 (The ONTARGET Investigators, 2008) と同様の結果であった. 一方で, ACE 阻害薬の非併用時のオルメサルタンの治療効果は, ハザード比 0.84 (95% 信頼区間 0.54-1.32, $p=0.450$; オルメサルタン群 35/75 [イベント発現例数/対象例数], 46.7%; プラセボ群 42/74, 56.8%) であった. この結果は ACE 阻害薬の非併用時のロサルタン (ARB の一種) の結果 (Brenner et al., 2001) と同様である. 併せて血圧, 尿蛋白の変化率, 血清 Cr 値の逆数に対するオルメサルタンの治療効果も, ACE 阻害薬の併用の有無による差異が確認されている (Imai et al., 2013a). これらの考察をふまえ, 次節の解析では, ACE 阻害薬の併用の有無を考慮した検討を実施する.

4.5.3 解析結果 1 (X, S, Y のみを用いた検討)

本節では, 図 4.1 において, 2 値変数 $X \in (x_0, x_1), S \in (s_0, s_1), Y \in (y_0, y_1)$, かつ $Z = \emptyset$ とした時の, S の代替性を評価する. なお, $Z = \emptyset$ であることから, NIE と NDE は式 (4.7), (4.8), (4.9), (4.10) によって算出できる. なお, このことは第 3 章で導入した PCS と, 本章の C-PCS が同値になることを意味する. なぜならば, 式 (3.3) の h を恒等関数とし, 式 (3.3) の g_{x_0} と式 (3.6) と式 (3.7) の $E(Y|x_0, s)$ を $\text{pr}(Y = 1|x_0, s)$ に置き換え, そして, 式 (3.3) の g_{x_1} と式 (3.7) の $E(Y|x_1, s)$ を共に $\text{pr}(Y = 1|x_1, s)$ に置き換えた PCS は, 式 (4.6) において $(x, x') = (x_1, x_0)$ かつ $\alpha = 1$ とした C-PCS と同じ定式化になるからである.

本節で検討する代替エンドポイントの候補は以下のとおりである.

- ・血圧: 治療開始後 12, 24, 48 週の 3 時点のそれぞれにおける収縮期血圧 (Systolic blood pressure [SBP]) / 拡張期血圧 (Diastolic blood pressure [DBP]) (mmHg) が, $<140/<90$ か否か, $<135/<85$ か否か, $<130/<80$ か否か, の 9 種類 (3 時点 \times 3 区分) を検討する

・尿蛋白変化率：治療開始後 12, 24, 48 週の 3 時点のそれぞれにおける治療前の値からの尿蛋白変化率 (%) が, <0 か否か, <20 か否か, <40 か否か, の 9 種類 (3 時点 × 3 区分) を検討する

・血清 Cr 値の逆数：治療開始後 12, 24, 48 週の 3 時点のそれぞれにおける血清 Cr 値の逆数 (dL/mg) が, プラセボ群の中央値より低いかな否か, および, 治療開始後 12 週と 48 週の両時点で中央値より低いかな否か, 治療期間全体の時点を用いた血清 Cr 値の逆数の推移 (症例ごとに線形回帰分析の傾き) のプラセボ群の中央値より低いかな否か, 治療開始後 48 週までの時点を用いた同様の定義, の 6 種類を検討する

代替性評価尺度の算出は, 式 (4.7), (4.8), (4.9), (4.10) を用い, ノンパラメトリックブートストラップ法により, 500 個の推定値を求め, その中央値, 2.5%点と 97.5 %点を出力する. また, C-PCS は式 (4.6) において $(x, x') = (x_1, x_0)$ とし, さらに $\alpha = 0, 1, \text{pr}(x_1)$ とした場合の三種類の結果を算出する. 比較のため, PI と R_G についても C-PCS と同様の観点から $\alpha = \text{pr}(x_1)$ を用いた以下の代替性評価尺度も出力する.

$$\begin{aligned} \text{PI} &= \alpha \frac{\text{NIE}(x_1)}{\text{TE}} + (1 - \alpha) \frac{\text{NIE}(x_0)}{\text{TE}} \\ R_G &= \alpha \frac{|\text{NIE}(x_1)|}{|\text{TE}|} + (1 - \alpha) \frac{|\text{NIE}(x_0)|}{|\text{TE}|} \end{aligned}$$

表 4.2 に, 血圧 (SBP/DBP), 尿蛋白変化率, そして血清 Cr 値の逆数の代替性を評価した結果を示す.

血圧 (SBP/DBP) について, $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると, すべての時点, かつすべての血圧区分において, その中央値は低い値であり (0.03 以下), 代替性はないと考えられる. また, $\alpha = 0$ もしくは 1 においても同様の結果であった. PI も同様に 0.09 以下と低い値であった. 一方で, R_G は他の代替性評価尺度に比べ高めの値 (0.2 以下) をとっていた. ただし, いずれの代替性評価尺度もその 2.5%点, 97.5 %点に基づく範囲は非常に広く, 信頼性を伴った判断をすることは困難である. それでも, C-PCS による評価であれば, 例えば 24

4.5. 適用例

週時点の $<135/<85$ を評価した場合に、 $\alpha = 1$ と $\text{pr}(x_1)$ の 97.5 % 点はともに 0.68 であることから、完全な代替性はないと判断することは可能である。

続いて、尿蛋白変化率について、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると、24 週時点の <0 を除いたすべての時点かつすべての変化率区分において、その中央値は 0.52~0.68 と中程度の値であった。また、 $\alpha = 0, 1$ とした場合では、 $\alpha = \text{pr}(x_1)$ を用いた場合に比べその範囲は広がった ($\alpha = 0$ のとき 0.37~0.70, $\alpha = 1$ のとき 0.38~0.69)。一方で、PI と R_G からは一貫した結果は得られていない (例えば、48 週時点の <40 についての $\alpha = 0$ における PI は 1.04 であったり、 $\alpha = \text{pr}(x_1)$ を用いた R_G に至っては 24 週時点の <0 を除いたすべての時点かつすべての変化率区分において、1 を超えていた)。血圧の検討時と同様に、いずれの代替性評価尺度もその 2.5% 点、97.5 % 点に基づく範囲は非常に広く、信頼性を伴った判断をすることは困難である。

最後に、血清 Cr 値の逆数について、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると、12, 24, 48 週の三時点それぞれ、および治療開始後 12 と 48 週時点の組み合わせはいずれも低い値であった (0.13 以下)。また、 $\alpha = 0, 1$ においても同様の結果であった ($\alpha = 0$ のとき 0.14 以下; $\alpha = 1$ のとき 0.12 以下)。一方で、血清 Cr 値の逆数の推移 (傾き [0-48 週時点]) の $\alpha = \text{pr}(x_1)$ を用いた C-PCS は 0.52 であり、傾き (全時点) の場合は 0.78 と、個別の 1 時点による結果より高い値であった。 $\alpha = 0, 1$ においても同様の結果である。しかし PI によると、12 週時点の中央値は -0.20 であり、C-PCS と同様の検討を PI から得ることは困難である。また、 R_G についてはいずれの α においても 1 を超えており、代替性の評価は困難であることが分かる。

表 4.2: ORIENT; 血圧 (SBP/DBP), 尿蛋白変化率, 血清 Cr 値の逆数の代替性評価

	NIE(x_0), NDE(x_0)による評価 ($\alpha = 0$)				NIE(x_1), NDE(x_1)による評価 ($\alpha = 1$)				中央値 (2.5%点, 97.5%点) による評価 ($\alpha = \text{pr}(x_1)$)			
	C-PCS	PI	RG	RG	C-PCS	PI	RG	RG	C-PCS	PI	RG	RG
SBP/DBP												
12 週時点, <140/<90	0.03 (0.00, 1.00)	0.07 (-1.87, 2.74)	0.16 (0.01, 4.68)	0.02 (0.00, 1.00)	0.06 (-1.23, 1.75)	0.12 (0.00, 2.85)	0.03 (0.00, 1.00)	0.08 (-1.35, 2.52)	0.15 (0.01, 3.43)			
12 週時点, <135/<85	0.05 (0.00, 1.00)	0.09 (-1.75, 2.75)	0.20 (0.01, 3.41)	0.01 (0.00, 1.00)	0.03 (-0.78, 1.17)	0.09 (0.00, 1.91)	0.03 (0.00, 1.00)	0.07 (-1.25, 1.95)	0.15 (0.01, 2.56)			
12 週時点, <130/<80	0.02 (0.00, 1.00)	0.05 (-1.66, 1.90)	0.14 (0.00, 3.08)	0.03 (0.00, 1.00)	0.07 (-1.80, 1.90)	0.15 (0.00, 3.39)	0.03 (0.00, 1.00)	0.07 (-1.65, 1.91)	0.15 (0.01, 3.00)			
24 週時点, <140/<90	0.02 (0.00, 1.00)	0.04 (-1.36, 1.23)	0.13 (0.00, 2.86)	0.01 (0.00, 1.00)	0.03 (-1.04, 1.16)	0.10 (0.00, 2.19)	0.02 (0.00, 1.00)	0.04 (-1.20, 1.19)	0.13 (0.00, 2.42)			
24 週時点, <135/<85	0.02 (0.00, 1.00)	0.02 (-1.17, 1.03)	0.13 (0.01, 1.78)	0.01 (0.00, 0.68)	0.01 (-0.90, 0.67)	0.08 (0.00, 1.40)	0.01 (0.00, 0.68)	0.03 (-1.01, 0.83)	0.11 (0.01, 1.70)			
24 週時点, <130/<80	0.02 (0.00, 1.00)	0.06 (-1.91, 1.10)	0.14 (0.00, 2.46)	0.02 (0.00, 1.00)	0.06 (-1.25, 1.59)	0.14 (0.00, 4.06)	0.03 (0.00, 1.00)	0.06 (-1.56, 1.24)	0.15 (0.01, 3.43)			
48 週時点, <140/<90	0.03 (0.00, 1.00)	0.08 (-2.00, 2.01)	0.16 (0.00, 3.92)	0.01 (0.00, 1.00)	0.02 (-0.69, 1.32)	0.08 (0.00, 2.60)	0.02 (0.00, 1.00)	0.05 (-0.97, 1.49)	0.13 (0.00, 3.11)			
48 週時点, <135/<85	0.02 (0.00, 1.00)	0.05 (-1.17, 1.78)	0.13 (0.00, 2.74)	0.00 (0.00, 0.58)	0.00 (-0.39, 0.71)	0.04 (0.00, 1.16)	0.01 (0.00, 1.00)	0.02 (-0.70, 1.22)	0.09 (0.00, 1.67)			
48 週時点, <130/<80	0.00 (0.00, 0.58)	0.01 (-0.44, 0.80)	0.05 (0.00, 0.92)	0.00 (0.00, 0.71)	0.00 (-0.37, 0.93)	0.04 (0.00, 1.22)	0.00 (0.00, 0.57)	0.00 (-0.26, 0.74)	0.05 (0.00, 1.00)			
尿蛋白変化率												
12 週時点, <0	0.37 (0.00, 1.00)	0.36 (-7.08, 7.06)	0.67 (0.03, 15.00)	0.69 (0.00, 1.00)	0.93 (-10.74, 16.40)	1.42 (0.37, 24.38)	0.52 (0.00, 1.00)	0.65 (-9.34, 10.06)	1.12 (0.27, 22.21)			
12 週時点, <20	0.69 (0.00, 1.00)	0.78 (-15.59, 13.13)	1.32 (0.27, 27.60)	0.48 (0.00, 1.00)	0.49 (-7.53, 9.74)	0.84 (0.06, 19.91)	0.56 (0.00, 1.00)	0.65 (-10.59, 11.64)	1.09 (0.28, 22.68)			
12 週時点, <40	0.65 (0.00, 1.00)	0.72 (-11.32, 12.34)	1.11 (0.30, 24.13)	0.59 (0.00, 1.00)	0.62 (-9.98, 7.29)	1.01 (0.09, 15.58)	0.59 (0.00, 1.00)	0.69 (-10.65, 11.31)	1.09 (0.30, 22.41)			
24 週時点, <0	0.10 (0.00, 1.00)	0.15 (-3.81, 3.29)	0.30 (0.01, 5.85)	0.19 (0.00, 1.00)	0.22 (-3.40, 5.98)	0.41 (0.04, 8.12)	0.19 (0.00, 1.00)	0.18 (-2.45, 4.04)	0.38 (0.06, 7.43)			
24 週時点, <20	0.66 (0.00, 1.00)	0.73 (-9.26, 11.03)	1.25 (0.32, 19.67)	0.46 (0.00, 1.00)	0.47 (-5.58, 9.90)	0.76 (0.08, 13.35)	0.54 (0.00, 1.00)	0.63 (-6.82, 9.61)	1.03 (0.29, 16.81)			
24 週時点, <40	0.65 (0.00, 1.00)	0.67 (-11.56, 12.91)	1.18 (0.31, 16.80)	0.57 (0.00, 1.00)	0.57 (-7.93, 10.85)	0.89 (0.12, 19.74)	0.60 (0.00, 1.00)	0.66 (-10.68, 11.51)	1.05 (0.33, 16.30)			
48 週時点, <0	0.67 (0.00, 1.00)	0.75 (-12.55, 12.34)	1.31 (0.32, 20.23)	0.47 (0.00, 1.00)	0.48 (-6.24, 6.69)	0.81 (0.19, 10.89)	0.61 (0.00, 1.00)	0.63 (-10.63, 8.99)	1.06 (0.31, 15.23)			
48 週時点, <20	0.70 (0.00, 1.00)	0.93 (-15.00, 14.10)	1.63 (0.47, 24.04)	0.61 (0.00, 1.00)	0.67 (-8.67, 9.40)	1.04 (0.27, 13.28)	0.68 (0.00, 1.00)	0.79 (-11.69, 11.71)	1.31 (0.41, 20.51)			
48 週時点, <40	0.70 (0.00, 1.00)	1.04 (-16.50, 16.62)	1.75 (0.52, 27.45)	0.38 (0.00, 1.00)	0.35 (-6.58, 5.51)	0.60 (0.05, 10.69)	0.54 (0.00, 1.00)	0.72 (-9.61, 10.22)	1.19 (0.33, 18.89)			
血清 Cr 値の逆数												
12 週時点	0.13 (0.00, 1.00)	-0.24 (-6.68, 4.84)	0.54 (0.04, 9.22)	0.08 (0.00, 1.00)	-0.16 (-4.91, 3.56)	0.35 (0.02, 6.65)	0.11 (0.00, 1.00)	-0.20 (-5.42, 4.23)	0.45 (0.04, 8.17)			
24 週時点	0.10 (0.00, 1.00)	0.03 (-3.80, 3.32)	0.36 (0.01, 10.32)	0.03 (0.00, 1.00)	0.02 (-2.17, 1.83)	0.19 (0.00, 3.98)	0.06 (0.00, 1.00)	0.02 (-2.79, 2.52)	0.27 (0.01, 7.25)			
48 週時点	0.12 (0.00, 1.00)	0.09 (-2.70, 4.90)	0.34 (0.01, 7.29)	0.10 (0.00, 1.00)	0.08 (-2.68, 4.25)	0.32 (0.01, 5.59)	0.12 (0.00, 1.00)	0.09 (-2.83, 4.53)	0.33 (0.01, 6.21)			
12*48 週時点	0.14 (0.00, 1.00)	0.14 (-3.32, 4.62)	0.37 (0.02, 7.38)	0.12 (0.00, 1.00)	0.12 (-3.14, 3.92)	0.34 (0.02, 6.33)	0.13 (0.00, 1.00)	0.13 (-3.09, 4.38)	0.36 (0.02, 6.76)			
傾き (全時点)	0.77 (0.00, 1.00)	0.97 (-8.84, 7.51)	1.23 (0.10, 14.01)	0.80 (0.00, 1.00)	0.94 (-7.57, 7.33)	1.15 (0.09, 11.40)	0.78 (0.00, 1.00)	0.96 (-8.31, 7.77)	1.18 (0.11, 12.65)			
傾き (0-48 週時点)	0.58 (0.00, 1.00)	0.61 (-8.82, 8.87)	0.91 (0.09, 18.65)	0.44 (0.00, 1.00)	0.44 (-6.80, 7.08)	0.64 (0.06, 15.09)	0.52 (0.00, 1.00)	0.54 (-7.98, 7.51)	0.78 (0.07, 16.82)			

4.5.4 解析結果 2 (X, S, Y と Z を用いた検討)

本節では, S と Y の間の同定された交絡因子 Z として ACE 阻害薬の併用 ($Z = z_1$)・非併用 ($Z = z_0$) を考慮した時の, S の代替性を評価する. このとき条件 (4.1), (4.2) の下で, NDE と NIE は次のように定式化できる.

$$\text{NDE}(x_0) = \sum_{s \in \{s_0, s_1\}, z \in \{s_0, s_1\}} \{\text{pr}(y_1|x_1, s, z) - \text{pr}(y_1|x_0, s, z)\} \text{pr}(s|x_0, z) \text{pr}(z) \quad (4.11)$$

$$\text{NIE}(x_0) = \sum_{s \in \{s_0, s_1\}, z \in \{s_0, s_1\}} \text{pr}(y_1|x_1, s, z) \{\text{pr}(s|x_1, z) - \text{pr}(s|x_0, z)\} \text{pr}(z) \quad (4.12)$$

$$\text{NDE}(x_1) = \sum_{s \in \{s_0, s_1\}, z \in \{s_0, s_1\}} \{\text{pr}(y_1|x_1, s, z) - \text{pr}(y_1|x_0, s, z)\} \text{pr}(s|x_1, z) \text{pr}(z) \quad (4.13)$$

$$\text{NIE}(x_1) = \sum_{s \in \{s_0, s_1\}, z \in \{s_0, s_1\}} \text{pr}(y_1|x_0, s, z) \{\text{pr}(s|x_1, z) - \text{pr}(s|x_0, z)\} \text{pr}(z) \quad (4.14)$$

なお, 上記の NDE と NIE を使った全例での検討に加え, ACE 阻害薬の併用・非併用のサブグループでの検討も行う. その際に利用する NDE と NIE は前節と同様に, 式 (4.7), (4.8), (4.9), (4.10) を用いる.

血圧の代替性評価の結果を表 4.3 に, 尿蛋白変化率の結果を表 4.4 に, そして血清 Cr 値の逆数の結果を表 4.5 に示す.

表 4.3 の血圧の代替性評価の結果から, $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると, 全例および ACE 阻害薬の併用・非併用のサブグループのいずれを対象としても, すべての時点, かつすべての血圧区分において, その中央値は低い値であり (0.06 以下), 代替性はないと考えられる. また, $\alpha = 0, 1$ においても同様の結果であった. また, このことは PI も同様に 0.13 以下であることから確認できる. ただし, 12 週時点の $<135/<85$ を評価した場合に, PI の中央値がマイナス ($\alpha = 0$ のとき -0.02; $\alpha = \text{pr}(x_1)$ のとき -0.01) であるため, 代替性の評価は困難である. 一方で, R_G は他の代替性評価尺度に比べ高めの値 (0.04~0.23) をとっていた. いずれの代替性評価尺度もその 2.5, 97.5 % 点は非常に広く, 信頼性を伴った判断をすることは困難である. それでも, C-PCS による評価であれば, 例えば 12 週時点の $<135/<85$

を評価した場合に、 $\alpha = 1$ と $\alpha = \text{pr}(x_1)$ の 97.5 % 点はそれぞれ、0.75 と 0.68 であることから、完全な代替性はないと判断することが可能である。上記の考察は、前節の解析結果 1 (表 4.2) と同様であり、血圧の代替性評価に対する ACE 阻害薬の併用・非併用による影響は小さいものと考えられる。

続いて、表 4.4 の尿蛋白変化率の結果から、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると、全例を対象としたとき、24 週時点の <0 を除いたすべての時点かつすべての変化率区分において、その中央値は 0.54~0.71 と中程度の値であった。これらの値は、前節の解析結果 1 (表 4.2) より若干ではあるが高い値である。ACE 阻害薬の非併用時の結果からは、さらに高い値が得られている (同じく 24 週時点の <0 を除く中央値の範囲として、0.51~0.79)。一方で、ACE 阻害薬の併用時の C-PCS は低い値であった (同じく、0.12~0.32)。また、 $\alpha = 0, 1$ においても、上記と同様の結果が得られている。しかし前節と同様に、PI と R_G は一貫した結果は得られていないため、これらの代替性評価尺度を用いた評価は限定的である。なお、前節と同様に、いずれの代替性評価尺度もその 2.5% 点、97.5 % 点に基づく範囲は非常に広く、信頼性を伴った判断をすることは困難である。

最後に、表 4.5 の血清 Cr 値の逆数の結果から、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると、全例の 12, 24, 48 週の 3 時点、および治療開始後 12 と 48 週時点の組み合わせはいずれも低い値であった (0.14 以下)。一方で、血清 Cr 値の逆数の推移 (傾き [0-48 週時点]) の $\alpha = \text{pr}(x_1)$ を用いた C-PCS は 0.52 であり、さらに傾き (全時点) の場合は 0.80 であった。これらの値は、個別の時点による結果より高い値であった。この傾向は、 $\alpha = 0, 1$ においても同様であった。さらに、ACE 阻害薬非併用時の場合も同様に、傾き (0-48 週時点) の C-PCS は、 $\alpha = 0, 1, \text{pr}(x_1)$ の順に、0.35, 0.07, 0.23 である一方で、治療開始後 12 と 48 週時点の組み合わせの C-PCS は 0.54, 0.19, 0.40 (同順) と高かった。これらの結果より、血清 Cr 値の逆数の推移を症例ごとの線形回帰分析の傾き (複数の時点を利用) によって評価するより、特定の 2 時点を利用する方が高い代替性を得られる可能性が示唆された。

表 4.3: ORIENT; 血圧 (SBP/DBP) の代替性評価

SBP/DBP	NIE(x_0), NDE(x_0) による評価 ($\alpha = 0$)			NIE(x_1), NDE(x_1) による評価 ($\alpha = 1$)			NIE(x), NDE(x) ($x \in (x_0, x_1)$) による評価 ($\alpha = \text{pr}(x_1)$)			
	C-PCS	PI	R_G	C-PCS	PI	R_G	C-PCS	PI	R_G	
ACE 阻害薬非併用	12 週時点, <140/<90	0.03 (0.00, 1.00)	0.09 (-1.89, 1.79)	0.17 (0.01, 3.15)	0.00 (0.00, 0.72)	-0.01 (-0.87, 1.47)	0.05 (0.00, 1.92)	0.02 (0.00, 1.00)	0.03 (-0.62, 1.01)	0.12 (0.00, 2.72)
	12 週時点, <135/<85	0.02 (0.00, 1.00)	-0.02 (-2.13, 1.44)	0.15 (0.01, 3.99)	0.00 (0.00, 0.75)	0.00 (-0.52, 0.72)	0.04 (0.00, 1.47)	0.01 (0.00, 0.68)	-0.01 (-0.80, 0.92)	0.10 (0.00, 2.58)
	12 週時点, <130/<80	0.02 (0.00, 1.00)	0.07 (-1.12, 1.56)	0.12 (0.00, 2.72)	0.00 (0.00, 0.84)	0.01 (-0.93, 0.84)	0.06 (0.00, 1.58)	0.02 (0.00, 1.00)	0.05 (-0.85, 1.14)	0.10 (0.00, 2.27)
	24 週時点, <140/<90	0.01 (0.00, 1.00)	0.01 (-0.99, 1.37)	0.10 (0.00, 1.83)	0.01 (0.00, 0.89)	0.01 (-1.03, 0.91)	0.08 (0.00, 2.05)	0.01 (0.00, 1.00)	0.01 (-0.93, 1.02)	0.10 (0.00, 1.87)
	24 週時点, <135/<85	0.05 (0.00, 1.00)	0.14 (-1.24, 2.69)	0.19 (0.01, 4.29)	0.01 (0.00, 0.92)	0.01 (-0.83, 1.30)	0.08 (0.00, 2.32)	0.03 (0.00, 1.00)	0.07 (-0.71, 1.54)	0.15 (0.01, 3.39)
	24 週時点, <130/<80	0.04 (0.00, 1.00)	0.12 (-1.07, 2.39)	0.18 (0.00, 4.19)	0.01 (0.00, 0.91)	0.01 (-0.62, 1.77)	0.09 (0.00, 2.70)	0.04 (0.00, 1.00)	0.07 (-0.56, 1.93)	0.14 (0.01, 3.22)
	48 週時点, <140/<90	0.07 (0.00, 1.00)	0.19 (-2.24, 2.41)	0.23 (0.01, 3.95)	0.01 (0.00, 0.94)	0.06 (-1.41, 2.04)	0.10 (0.00, 3.19)	0.06 (0.00, 1.00)	0.13 (-1.54, 1.75)	0.18 (0.01, 3.60)
	48 週時点, <135/<85	0.04 (0.00, 1.00)	0.10 (-2.18, 1.95)	0.17 (0.01, 5.04)	0.00 (0.00, 0.72)	0.00 (-0.77, 1.30)	0.05 (0.00, 1.90)	0.02 (0.00, 0.73)	0.04 (-1.11, 1.28)	0.12 (0.01, 3.53)
	48 週時点, <130/<80	0.05 (0.00, 1.00)	0.15 (-1.95, 2.50)	0.19 (0.01, 3.26)	0.00 (0.00, 0.94)	0.01 (-1.11, 1.66)	0.07 (0.00, 2.19)	0.03 (0.00, 1.00)	0.07 (-1.17, 2.24)	0.13 (0.00, 2.66)
ACE 阻害薬併用	12 週時点, <140/<90	0.01 (0.00, 1.00)	0.01 (-0.92, 1.16)	0.09 (0.00, 2.08)	0.03 (0.00, 1.00)	0.03 (-1.30, 2.02)	0.18 (0.01, 4.15)	0.03 (0.00, 1.00)	0.03 (-1.00, 1.23)	0.14 (0.01, 3.29)
	12 週時点, <135/<85	0.04 (0.00, 1.00)	0.03 (-2.88, 2.12)	0.19 (0.01, 4.56)	0.03 (0.00, 1.00)	0.03 (-1.51, 2.11)	0.17 (0.00, 4.29)	0.04 (0.00, 1.00)	0.04 (-2.35, 1.87)	0.19 (0.01, 5.08)
	12 週時点, <130/<80	0.01 (0.00, 0.85)	0.01 (-1.55, 0.94)	0.09 (0.00, 2.57)	0.04 (0.00, 1.00)	0.03 (-2.97, 2.06)	0.20 (0.00, 4.66)	0.03 (0.00, 1.00)	0.03 (-2.33, 1.40)	0.15 (0.00, 3.62)
	24 週時点, <140/<90	0.02 (0.00, 1.00)	0.02 (-1.42, 1.19)	0.15 (0.01, 2.64)	0.02 (0.00, 1.00)	0.02 (-1.08, 1.17)	0.13 (0.00, 2.52)	0.03 (0.00, 1.00)	0.03 (-1.31, 1.13)	0.15 (0.01, 2.00)
	24 週時点, <135/<85	0.02 (0.00, 1.00)	0.02 (-1.68, 1.53)	0.15 (0.00, 2.81)	0.02 (0.00, 1.00)	0.02 (-1.77, 1.09)	0.14 (0.00, 2.43)	0.02 (0.00, 1.00)	0.02 (-1.97, 1.24)	0.15 (0.01, 2.54)
	24 週時点, <130/<80	0.01 (0.00, 0.64)	0.00 (-0.90, 0.82)	0.08 (0.00, 2.71)	0.02 (0.00, 0.90)	0.00 (-1.02, 0.99)	0.14 (0.00, 3.26)	0.02 (0.00, 0.64)	0.00 (-1.04, 0.68)	0.12 (0.00, 2.45)
	48 週時点, <140/<90	0.01 (0.00, 0.83)	0.00 (-0.93, 0.81)	0.08 (0.00, 1.42)	0.00 (0.00, 1.00)	0.00 (-0.99, 1.13)	0.07 (0.00, 1.80)	0.01 (0.00, 0.52)	0.00 (-0.63, 0.70)	0.09 (0.00, 1.40)
	48 週時点, <135/<85	0.01 (0.00, 1.00)	0.00 (-0.82, 1.03)	0.08 (0.00, 1.67)	0.00 (0.00, 0.58)	0.00 (-0.64, 0.54)	0.05 (0.00, 1.27)	0.01 (0.00, 0.54)	0.00 (-0.63, 0.64)	0.07 (0.00, 1.34)
	48 週時点, <130/<80	0.00 (0.00, 0.71)	0.00 (-0.66, 0.64)	0.04 (0.00, 1.13)	0.00 (0.00, 0.51)	0.00 (-0.61, 0.51)	0.04 (0.00, 1.40)	0.00 (0.00, 0.49)	0.00 (-0.40, 0.39)	0.05 (0.00, 1.09)
全例	12 週時点, <140/<90	0.04 (0.00, 1.00)	0.08 (-1.84, 2.15)	0.18 (0.01, 4.54)	0.02 (0.00, 1.00)	0.05 (-1.58, 1.35)	0.14 (0.01, 2.95)	0.05 (0.00, 1.00)	0.07 (-1.34, 1.70)	0.18 (0.02, 3.28)
	12 週時点, <135/<85	0.04 (0.00, 1.00)	0.08 (-1.60, 2.94)	0.19 (0.01, 4.06)	0.02 (0.00, 1.00)	0.05 (-1.41, 1.60)	0.13 (0.00, 2.99)	0.04 (0.00, 1.00)	0.06 (-1.11, 2.09)	0.17 (0.02, 3.15)
	12 週時点, <130/<80	0.03 (0.00, 1.00)	0.08 (-1.61, 1.99)	0.15 (0.01, 3.52)	0.03 (0.00, 1.00)	0.07 (-1.61, 2.55)	0.16 (0.01, 3.13)	0.04 (0.00, 1.00)	0.08 (-1.36, 2.04)	0.16 (0.01, 3.07)
	24 週時点, <140/<90	0.02 (0.00, 1.00)	0.04 (-1.29, 1.27)	0.13 (0.00, 2.42)	0.01 (0.00, 1.00)	0.03 (-1.35, 1.49)	0.11 (0.00, 2.50)	0.02 (0.00, 1.00)	0.04 (-1.08, 1.30)	0.13 (0.01, 2.25)
	24 週時点, <135/<85	0.03 (0.00, 1.00)	0.05 (-1.46, 1.10)	0.16 (0.01, 2.45)	0.01 (0.00, 1.00)	-0.02 (-1.37, 1.31)	0.13 (0.01, 2.66)	0.03 (0.00, 0.67)	0.02 (-1.08, 0.86)	0.16 (0.02, 2.90)
	24 週時点, <130/<80	0.03 (0.00, 1.00)	0.08 (-2.73, 1.65)	0.17 (0.01, 3.54)	0.02 (0.00, 1.00)	0.04 (-1.94, 1.29)	0.13 (0.01, 2.54)	0.04 (0.00, 1.00)	0.07 (-2.08, 1.23)	0.17 (0.02, 3.44)
	48 週時点, <140/<90	0.06 (0.00, 1.00)	0.13 (-2.25, 2.24)	0.23 (0.01, 4.00)	0.01 (0.00, 1.00)	0.04 (-1.07, 1.67)	0.11 (0.00, 3.04)	0.05 (0.00, 1.00)	0.09 (-1.59, 1.58)	0.19 (0.02, 3.56)
	48 週時点, <135/<85	0.03 (0.00, 1.00)	0.07 (-1.52, 1.88)	0.17 (0.01, 3.16)	0.00 (0.00, 0.73)	0.00 (-0.78, 0.94)	0.07 (0.00, 1.73)	0.02 (0.00, 0.71)	0.03 (-0.85, 0.90)	0.13 (0.01, 2.44)
	48 週時点, <130/<80	0.03 (0.00, 1.00)	0.07 (-1.30, 1.99)	0.15 (0.01, 3.01)	0.00 (0.00, 1.00)	0.01 (-0.89, 1.10)	0.07 (0.00, 1.81)	0.02 (0.00, 0.67)	0.04 (-0.86, 0.93)	0.13 (0.01, 2.71)

ACE 阻害薬非併用集団での TE=0.12 (-0.05, 0.28); ACE 阻害薬併用集団での TE=0 (-0.1, 0.11); 全例での TE=0.04 (-0.06, 0.12)

表 4.4: ORIENT; 尿蛋白変化率の代替性評価

尿蛋白変化率	NIE(x_{t_0}), NDE(x_{t_0}) による評価 ($\alpha = 0$)			NIE(x_{t_1}), NDE(x_{t_1}) による評価 ($\alpha = 1$)			中央値 (2.5%点, 97.5 %点) による評価 ($\alpha = \text{pr}(x_{t_1})$)		
	C-PCS	PI	R_G	C-PCS	PI	R_G	C-PCS	PI	R_G
ACE 阻害薬非併用									
12 週時点, <0	0.50 (0.00, 1.00)	0.51 (-4.13, 6.58)	0.67 (0.04, 13.42)	0.79 (0.15, 1.00)	1.08 (-7.35, 15.68)	1.25 (0.30, 21.78)	0.59 (0.00, 1.00)	0.83 (-7.58, 10.33)	0.97 (0.28, 18.89)
12 週時点, <20	0.74 (0.00, 1.00)	0.87 (-9.50, 11.49)	1.00 (0.19, 20.36)	0.67 (0.00, 1.00)	0.73 (-6.42, 11.10)	0.87 (0.06, 15.56)	0.65 (0.00, 1.00)	0.81 (-7.34, 10.69)	0.94 (0.22, 19.29)
12 週時点, <40	0.72 (0.00, 1.00)	0.73 (-6.63, 10.84)	0.87 (0.27, 18.78)	0.88 (0.28, 1.00)	1.00 (-6.51, 11.83)	1.12 (0.49, 20.40)	0.78 (0.00, 1.00)	0.86 (-6.63, 11.03)	1.00 (0.41, 20.31)
24 週時点, <0	0.31 (0.00, 1.00)	0.34 (-3.90, 4.38)	0.48 (0.04, 7.99)	0.39 (0.00, 1.00)	0.42 (-3.26, 6.31)	0.49 (0.03, 10.50)	0.43 (0.00, 1.00)	0.40 (-4.65, 6.57)	0.52 (0.08, 9.73)
24 週時点, <20	0.52 (0.00, 1.00)	0.52 (-6.38, 6.27)	0.63 (0.06, 13.71)	0.80 (0.06, 1.00)	1.05 (-8.43, 13.14)	1.19 (0.20, 20.01)	0.63 (0.00, 1.00)	0.81 (-7.13, 10.77)	0.94 (0.23, 17.30)
24 週時点, <40	0.49 (0.00, 1.00)	0.49 (-5.35, 6.94)	0.57 (0.07, 11.00)	0.53 (0.00, 1.00)	0.55 (-6.35, 8.94)	0.72 (0.04, 12.25)	0.51 (0.00, 1.00)	0.52 (-5.90, 6.73)	0.65 (0.14, 12.91)
48 週時点, <0	0.81 (0.00, 1.00)	1.62 (-10.10, 22.72)	1.87 (0.76, 34.26)	0.81 (0.09, 1.00)	0.91 (-6.05, 11.07)	1.07 (0.23, 19.58)	0.79 (0.00, 1.00)	1.25 (-8.89, 17.74)	1.43 (0.60, 30.72)
48 週時点, <20	0.79 (0.00, 1.00)	1.64 (-13.27, 26.65)	1.93 (0.70, 39.20)	0.72 (0.02, 1.00)	0.91 (-7.68, 13.98)	1.08 (0.12, 18.49)	0.74 (0.00, 1.00)	1.28 (-8.88, 20.74)	1.48 (0.57, 28.39)
48 週時点, <40	0.80 (0.00, 1.00)	1.63 (-12.66, 26.74)	1.89 (0.68, 47.16)	0.52 (0.00, 1.00)	0.53 (-7.94, 10.46)	1.02 (0.05, 23.15)	0.59 (0.00, 1.00)	1.10 (-9.09, 16.42)	1.46 (0.49, 26.80)
ACE 阻害薬併用									
12 週時点, <0	0.17 (0.00, 1.00)	-0.02 (-5.31, 4.78)	0.52 (0.04, 10.43)	0.31 (0.00, 1.00)	0.09 (-7.11, 8.87)	0.81 (0.08, 15.63)	0.30 (0.00, 1.00)	-0.01 (-5.14, 6.21)	0.73 (0.12, 12.61)
12 週時点, <20	0.32 (0.00, 1.00)	0.03 (-6.82, 8.74)	0.87 (0.06, 12.30)	0.16 (0.00, 1.00)	0.02 (-5.70, 4.70)	0.52 (0.02, 10.27)	0.32 (0.00, 1.00)	0.02 (-5.15, 6.18)	0.71 (0.13, 12.04)
12 週時点, <40	0.27 (0.00, 1.00)	0.08 (-6.86, 8.20)	0.71 (0.08, 11.41)	0.24 (0.00, 1.00)	0.07 (-4.79, 10.78)	0.71 (0.04, 14.52)	0.29 (0.00, 1.00)	0.10 (-4.92, 8.65)	0.70 (0.16, 14.92)
24 週時点, <0	0.04 (0.00, 1.00)	0.01 (-1.96, 2.12)	0.20 (0.01, 3.92)	0.04 (0.00, 1.00)	-0.03 (-2.40, 2.24)	0.20 (0.01, 3.80)	0.05 (0.00, 1.00)	-0.01 (-1.33, 1.30)	0.22 (0.02, 3.95)
24 週時点, <20	0.36 (0.00, 1.00)	0.21 (-6.02, 9.65)	0.98 (0.20, 20.08)	0.08 (0.00, 1.00)	-0.01 (-2.61, 2.71)	0.30 (0.01, 6.78)	0.26 (0.00, 1.00)	0.10 (-3.29, 7.73)	0.68 (0.15, 13.07)
24 週時点, <40	0.35 (0.00, 1.00)	0.20 (-6.57, 7.58)	0.97 (0.20, 20.71)	0.24 (0.00, 1.00)	0.05 (-3.40, 5.98)	0.62 (0.04, 11.86)	0.32 (0.00, 1.00)	0.17 (-3.93, 6.90)	0.84 (0.17, 15.95)
48 週時点, <0	0.14 (0.00, 1.00)	0.07 (-3.32, 5.18)	0.46 (0.04, 9.36)	0.07 (0.00, 1.00)	0.04 (-2.37, 3.22)	0.28 (0.01, 6.22)	0.12 (0.00, 1.00)	0.06 (-2.56, 4.12)	0.36 (0.04, 7.03)
48 週時点, <20	0.25 (0.00, 1.00)	0.14 (-4.45, 7.99)	0.69 (0.07, 14.24)	0.15 (0.00, 1.00)	0.06 (-3.26, 6.07)	0.44 (0.04, 7.96)	0.20 (0.00, 1.00)	0.16 (-3.96, 6.41)	0.59 (0.07, 12.06)
48 週時点, <40	0.30 (0.00, 1.00)	0.16 (-6.04, 9.42)	0.80 (0.12, 14.88)	0.06 (0.00, 1.00)	0.02 (-1.91, 2.84)	0.25 (0.01, 5.26)	0.21 (0.00, 1.00)	0.12 (-3.66, 5.80)	0.54 (0.08, 10.36)
全例									
12 週時点, <0	0.46 (0.00, 1.00)	0.44 (-6.69, 7.43)	0.81 (0.04, 17.76)	0.70 (0.00, 1.00)	1.01 (-15.94, 11.25)	1.53 (0.41, 25.12)	0.56 (0.00, 1.00)	0.74 (-12.03, 8.70)	1.20 (0.33, 21.36)
12 週時点, <20	0.70 (0.00, 1.00)	0.86 (-16.51, 14.62)	1.35 (0.32, 23.43)	0.56 (0.00, 1.00)	0.57 (-9.54, 8.54)	0.93 (0.06, 17.34)	0.62 (0.00, 1.00)	0.72 (-12.77, 11.76)	1.21 (0.30, 20.70)
12 週時点, <40	0.69 (0.00, 1.00)	0.76 (-12.84, 10.72)	1.16 (0.36, 23.91)	0.73 (0.00, 1.00)	0.87 (-14.48, 11.44)	1.36 (0.39, 20.07)	0.71 (0.00, 1.00)	0.83 (-14.00, 10.65)	1.26 (0.41, 21.35)
24 週時点, <0	0.21 (0.00, 1.00)	0.22 (-4.47, 3.94)	0.42 (0.02, 6.41)	0.24 (0.00, 1.00)	0.27 (-5.26, 4.36)	0.46 (0.02, 8.28)	0.28 (0.00, 1.00)	0.26 (-4.78, 4.29)	0.47 (0.09, 7.25)
24 週時点, <20	0.68 (0.00, 1.00)	0.77 (-17.27, 8.12)	1.22 (0.30, 21.51)	0.65 (0.00, 1.00)	0.71 (-11.54, 8.30)	1.07 (0.18, 16.35)	0.65 (0.00, 1.00)	0.76 (-14.82, 7.62)	1.17 (0.33, 17.51)
24 週時点, <40	0.66 (0.00, 1.00)	0.70 (-13.31, 9.08)	1.14 (0.31, 20.99)	0.59 (0.00, 1.00)	0.59 (-10.12, 8.16)	0.98 (0.12, 14.17)	0.61 (0.00, 1.00)	0.65 (-11.03, 9.02)	1.06 (0.33, 17.52)
48 週時点, <0	0.72 (0.00, 1.00)	1.14 (-24.83, 12.85)	1.80 (0.59, 32.16)	0.64 (0.00, 1.00)	0.68 (-10.82, 7.45)	1.02 (0.25, 14.88)	0.70 (0.00, 1.00)	0.90 (-17.76, 10.00)	1.44 (0.46, 23.98)
48 週時点, <20	0.71 (0.00, 1.00)	1.24 (-26.24, 14.49)	2.09 (0.64, 34.33)	0.64 (0.00, 1.00)	0.77 (-18.49, 8.54)	1.18 (0.18, 22.76)	0.69 (0.00, 1.00)	1.03 (-20.79, 10.28)	1.59 (0.50, 26.70)
48 週時点, <40	0.70 (0.00, 1.00)	1.34 (-25.44, 15.33)	2.10 (0.62, 36.64)	0.41 (0.00, 1.00)	0.31 (-13.16, 6.41)	0.86 (0.07, 19.69)	0.54 (0.00, 1.00)	0.88 (-17.42, 9.94)	1.47 (0.41, 26.61)
ACE 阻害薬非併用集団での TE=0.12 (-0.05, 0.28); ACE 阻害薬併用集団での TE=0 (-0.1, 0.11); 全例での TE=0.04 (-0.06, 0.12)									

表 4.5: ORIENT; 血清 Cr 値の逆数の代替性評価

血清 Cr 値の逆数	NIE(x_0), NDE(x_0) による評価 ($\alpha = 0$)			NIE(x_1), NDE(x_1) による評価 ($\alpha = 1$)			中央値 (2.5%点, 97.5%点) による評価 ($\alpha = \text{pr}(x_1)$)		
	C-PCS	PI	R _G	C-PCS	PI	R _G	C-PCS	PI	R _G
ACE 阻害薬非併用									
12 週時点	0.03 (0.00, 1.00)	-0.14 (-4.18, 1.55)	0.19 (0.00, 5.81)	0.03 (0.00, 0.82)	-0.16 (-3.48, 1.89)	0.21 (0.00, 5.22)	0.03 (0.00, 1.00)	-0.16 (-2.98, 1.82)	0.20 (0.01, 5.39)
24 週時点	0.10 (0.00, 1.00)	0.12 (-2.31, 2.60)	0.31 (0.02, 7.88)	0.05 (0.00, 0.98)	0.09 (-1.71, 2.25)	0.22 (0.01, 4.57)	0.08 (0.00, 1.00)	0.11 (-1.89, 2.38)	0.26 (0.01, 6.19)
48 週時点	0.37 (0.00, 1.00)	0.41 (-2.63, 3.85)	0.48 (0.03, 7.45)	0.12 (0.00, 1.00)	0.21 (-2.16, 2.39)	0.27 (0.01, 5.54)	0.26 (0.00, 1.00)	0.32 (-2.67, 3.34)	0.38 (0.02, 6.57)
12 週時点*48	0.54 (0.00, 1.00)	0.54 (-3.80, 4.80)	0.59 (0.05, 6.86)	0.19 (0.00, 0.99)	0.30 (-2.12, 3.08)	0.33 (0.03, 4.89)	0.40 (0.00, 1.00)	0.41 (-3.25, 3.45)	0.46 (0.05, 6.02)
傾き (全時点)	0.87 (0.00, 1.00)	0.86 (-2.66, 5.36)	0.90 (0.11, 9.76)	0.83 (0.01, 1.00)	0.80 (-2.11, 5.05)	0.84 (0.10, 8.80)	0.85 (0.00, 1.00)	0.83 (-2.37, 5.34)	0.87 (0.11, 9.28)
傾き (0-48 週時点)	0.35 (0.00, 1.00)	0.38 (-2.18, 5.98)	0.47 (0.02, 8.34)	0.07 (0.00, 0.97)	0.18 (-1.14, 2.84)	0.24 (0.01, 4.43)	0.23 (0.00, 1.00)	0.29 (-1.71, 4.35)	0.36 (0.02, 5.90)
ACE 阻害薬併用									
12 週時点	0.14 (0.00, 1.00)	0.04 (-5.19, 4.34)	0.42 (0.02, 9.15)	0.05 (0.00, 1.00)	0.02 (-2.99, 3.09)	0.25 (0.01, 5.49)	0.10 (0.00, 1.00)	0.03 (-3.75, 3.94)	0.34 (0.02, 8.04)
24 週時点	0.14 (0.00, 1.00)	0.08 (-4.42, 3.17)	0.41 (0.01, 7.61)	0.02 (0.00, 1.00)	0.03 (-2.05, 1.58)	0.15 (0.00, 3.67)	0.09 (0.00, 1.00)	0.07 (-2.87, 2.34)	0.29 (0.01, 5.51)
48 週時点	0.17 (0.00, 1.00)	0.13 (-3.18, 4.72)	0.45 (0.02, 6.77)	0.21 (0.00, 1.00)	0.15 (-3.48, 4.53)	0.54 (0.03, 6.89)	0.20 (0.00, 1.00)	0.14 (-3.15, 4.92)	0.51 (0.02, 6.93)
12 週時点*48	0.22 (0.00, 1.00)	0.12 (-4.15, 5.35)	0.50 (0.02, 7.40)	0.22 (0.00, 1.00)	0.12 (-4.71, 5.06)	0.55 (0.02, 7.29)	0.22 (0.00, 1.00)	0.12 (-4.43, 5.20)	0.53 (0.02, 7.16)
傾き (全時点)	0.47 (0.00, 1.00)	0.48 (-4.41, 8.08)	0.88 (0.03, 15.32)	0.47 (0.00, 1.00)	0.45 (-4.18, 7.72)	0.87 (0.03, 13.56)	0.47 (0.00, 1.00)	0.46 (-4.30, 7.99)	0.88 (0.03, 14.71)
傾き (0-48 週時点)	0.33 (0.00, 1.00)	0.23 (-5.91, 10.05)	0.86 (0.05, 15.98)	0.28 (0.00, 1.00)	0.20 (-4.65, 7.39)	0.66 (0.04, 11.74)	0.30 (0.00, 1.00)	0.22 (-5.14, 8.03)	0.76 (0.04, 12.49)
全例									
12 週時点	0.11 (0.00, 1.00)	-0.21 (-4.19, 5.33)	0.46 (0.02, 8.34)	0.08 (0.00, 1.00)	-0.17 (-3.00, 4.70)	0.37 (0.02, 5.40)	0.10 (0.00, 1.00)	-0.19 (-3.41, 4.67)	0.44 (0.03, 6.35)
24 週時点	0.10 (0.00, 1.00)	0.05 (-2.76, 5.54)	0.35 (0.01, 7.27)	0.04 (0.00, 1.00)	0.04 (-1.69, 2.34)	0.20 (0.01, 3.18)	0.08 (0.00, 1.00)	0.05 (-1.92, 3.31)	0.29 (0.02, 5.17)
48 週時点	0.12 (0.00, 1.00)	0.16 (-2.81, 4.70)	0.34 (0.01, 6.63)	0.09 (0.00, 1.00)	0.03 (-2.78, 3.92)	0.33 (0.01, 7.22)	0.11 (0.00, 1.00)	0.09 (-2.56, 4.12)	0.34 (0.03, 6.48)
12 週時点*48	0.15 (0.00, 1.00)	0.19 (-3.25, 5.39)	0.37 (0.03, 7.75)	0.08 (0.00, 1.00)	0.07 (-3.07, 4.18)	0.32 (0.02, 6.92)	0.14 (0.00, 1.00)	0.14 (-3.01, 4.92)	0.36 (0.05, 6.52)
傾き (全時点)	0.79 (0.00, 1.00)	0.99 (-7.49, 6.64)	1.21 (0.15, 12.48)	0.81 (0.00, 1.00)	0.95 (-6.81, 6.89)	1.14 (0.09, 11.00)	0.80 (0.00, 1.00)	0.96 (-6.86, 6.65)	1.18 (0.12, 10.85)
傾き (0-48 週時点)	0.60 (0.00, 1.00)	0.62 (-10.00, 8.20)	0.90 (0.07, 13.59)	0.44 (0.00, 1.00)	0.44 (-5.55, 5.52)	0.60 (0.05, 9.72)	0.52 (0.00, 1.00)	0.53 (-6.71, 6.20)	0.75 (0.09, 12.34)

ACE 阻害薬非併用集団での TE=0.12 (-0.05, 0.28); ACE 阻害薬併用集団での TE=0 (-0.1, 0.11); 全例での TE=0.04 (-0.06, 0.12)

4.5.5 解析結果 3 (治療期間平均を用いた検討)

4.5.2 節の事前検討 1 に記載したとおり、血圧と血清 Cr 値の逆数に対する治療効果は、試験開始 1 年以内では十分ではなく、期間の後半になって確認されている。そこで、試験期間の後半の治療効果が代替性評価に与える影響を確認するため、治療期間すべての値を用いた検討を実施する。

治療期間平均を用いた SBP/DBP の代替性評価の結果を表 4.6 に、尿蛋白変化率の結果を表 4.7 に、そして血清 Cr 値の逆数の結果を表 4.8 に示す。

血圧の代替性評価について、本節の治療期間平均を用いた結果 (表 4.6) を前節の解析結果 2 (表 4.3) と比べると、明確に、期間平均を用いた SBP/DBPの方が C-PCS の中央値が高くなっていることがわかる (例えば、全例の $\alpha = \text{pr}(x_1)$ を用いた C-PCS は、表 4.3 では、0.05 以下であったのに対し、表 4.6 では、0.44 以下である)。このことは、ACE 阻害薬の非併用時に顕著であり、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS は、最大で 0.52 (平均 (全時点), <135/<85) であった。さらに、 $\alpha = 1$ であれば、平均 (全時点), <140/<90 での C-PCS の中央値は 0.81 と高い値であった。一方で、ACE 阻害薬の併用時は、前節までの結果と大差はなかった。このことは、ACE 阻害薬の併用時のオルメサルタンの血圧への治療効果が弱いことから理解できる。つまり、前節にくらべて、血圧の代替性評価に対する ACE 阻害薬の併用・非併用による影響が存在する可能性があり、ACE 阻害薬の非併用時であれば、腎複合イベントに対する治療効果の血圧の代替性は存在する可能性がある。今後もより規模の大きい試験でこのことを確認するべきであると考えられる。

続いて、表 4.7 の期間平均を用いた尿蛋白変化率の結果から、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると、全例を対象としたとき、平均 (全時点) と平均 (2~48 週時点) のいずれにおいても、C-PCS の中央値は前節の結果に比べ高かった (0.75~0.84 [表 4.7]; 0.28~0.71, [表 4.4])。この傾向は、ACE 阻害薬の併用時 (0.29~0.59 [表 4.7]; 0.05~0.32 [表 4.4])、および非併用時 (0.78~0.88 [表 4.7]; 0.43~0.79 [表 4.4]) においても同様であった。また、 $\alpha = 0, 1$ におい

でも、上記と同様の結果が得られている。一方で、前節と同様に PI と R_G は一貫した結果が得られていないため、これらの代替性評価尺度を用いた評価は限定的である。なお、前節と同様に、いずれの代替性評価尺度もその 2.5% 点, 97.5 % 点に基づく範囲は非常に広く、信頼性を伴った判断をすることは困難である。

最後に、表 4.8 の期間平均を用いた血清 Cr 値の逆数の結果から、 $\alpha = \text{pr}(x_1)$ を用いた C-PCS によると、全例の平均 (全時点) と平均 (2~48 週時点) のいずれにおいても、C-PCS の中央値は低かった (0.14 以下)。この傾向は、ACE 阻害薬の併用時も同様であった (0.20 以下)。一方で、ACE 阻害薬の非併用時では、平均 (全時点) の >-0.15 の結果でのみ高い値であった (0.70)。また、 $\alpha = 0, 1$ においても、上記と同様の結果が得られている。これらの結果を、前節の表 4.5 の結果における傾き (全時点) や治療開始後 12 と 48 週時点の組み合わせによる結果をふまえて解釈すると、つぎのような考察が得られる。血清 Cr 値の逆数を用いた代替性について、各時点および平均 (全時点) を使った評価では十分な代替性は確認できなかった。一方で、傾き (全時点) と治療開始後 12 と 48 週時点の組み合わせを使った評価では適度な代替性を示しており、今後の再検討を必要とする結果と考えられる。このことは、オルメサルタンのような ARB では治療初期に一時的に血清 Cr 値の逆数が減少し、その後減少の程度が小さく (つまり、減少しにくく) なり、治療後期にはプラセボ群との間に差がみられる現象が確認されており (Stevens et al., 2006; Holtkamp et al., 2011), その現象と符合するものと考えられる。

表 4.6: ORIENT; 血圧 (SBP/DBP) の代替性評価 (期間平均を用いた検討)

SBP/DBP	NIE(x_0), NDE(x_0) による評価 ($\alpha = 0$)			NIE(x_1), NDE(x_1) による評価 ($\alpha = 1$)			中央値 (2.5%点, 97.5%点) による評価 ($\alpha = \text{pr}(x_1)$)		
	C-PCS	PI	R_G	C-PCS	PI	R_G	C-PCS	PI	R_G
ACE 阻害薬非併用	平均 (2~48 週), <140/<90	0.16 (0.00, 1.00)	0.26 (-1.78, 2.39)	0.31 (0.02, 4.57)	0.06 (-0.84, 1.18)	0.11 (0.00, 1.66)	0.10 (0.00, 1.00)	0.18 (-0.85, 1.50)	0.23 (0.02, 3.36)
	平均 (2~48 週), <135/<85	0.30 (0.00, 1.00)	0.37 (-1.89, 3.23)	0.44 (0.04, 6.00)	0.17 (-1.22, 1.81)	0.21 (0.01, 3.40)	0.22 (0.00, 1.00)	0.27 (-1.57, 2.17)	0.33 (0.04, 4.35)
	平均 (2~48 週), <130/<80	0.16 (0.00, 1.00)	0.28 (-1.44, 2.36)	0.32 (0.02, 4.29)	0.23 (-1.36, 1.92)	0.27 (0.02, 3.16)	0.17 (0.00, 1.00)	0.27 (-1.52, 1.89)	0.31 (0.03, 3.70)
	平均 (全時点), <140/<90	0.81 (0.00, 1.00)	0.93 (-6.45, 7.38)	1.04 (0.27, 14.13)	0.09 (-1.85, 2.09)	0.21 (0.01, 3.35)	0.47 (0.00, 1.00)	0.51 (-3.17, 3.55)	0.63 (0.17, 9.02)
	平均 (全時点), <135/<85	0.67 (0.03, 1.00)	0.66 (-3.80, 4.82)	0.76 (0.15, 10.62)	0.43 (-2.96, 3.25)	0.47 (0.04, 6.86)	0.52 (0.00, 1.00)	0.54 (-3.38, 4.08)	0.62 (0.15, 7.77)
	平均 (全時点), <130/<80	0.36 (0.00, 1.00)	0.41 (-2.29, 4.08)	0.46 (0.05, 5.13)	0.31 (-1.90, 2.31)	0.35 (0.02, 4.10)	0.32 (0.00, 1.00)	0.37 (-1.85, 3.27)	0.41 (0.05, 4.92)
ACE 阻害薬併用	平均 (2~48 週), <140/<90	0.03 (0.00, 1.00)	0.05 (-3.70, 3.08)	0.17 (0.01, 5.04)	0.02 (-2.28, 1.81)	0.12 (0.00, 4.38)	0.03 (0.00, 1.00)	0.04 (-2.72, 2.56)	0.15 (0.01, 4.76)
	平均 (2~48 週), <135/<85	0.10 (0.00, 1.00)	0.14 (-6.27, 2.91)	0.30 (0.02, 8.70)	0.14 (-5.87, 2.89)	0.28 (0.02, 8.39)	0.12 (0.00, 1.00)	0.16 (-5.68, 2.83)	0.30 (0.02, 8.77)
	平均 (2~48 週), <130/<80	0.01 (0.00, 1.00)	0.04 (-2.44, 1.66)	0.11 (0.00, 4.54)	0.06 (-2.94, 2.07)	0.17 (0.01, 5.81)	0.03 (0.00, 1.00)	0.05 (-3.09, 1.77)	0.15 (0.01, 4.55)
	平均 (全時点), <140/<90	0.08 (0.00, 1.00)	0.08 (-6.09, 2.85)	0.27 (0.02, 7.71)	0.04 (-4.31, 2.49)	0.18 (0.01, 5.54)	0.06 (0.00, 1.00)	0.07 (-5.19, 2.42)	0.23 (0.01, 7.15)
	平均 (全時点), <135/<85	0.09 (0.00, 1.00)	0.12 (-8.22, 2.64)	0.27 (0.02, 10.28)	0.13 (-7.35, 2.32)	0.29 (0.01, 9.23)	0.09 (0.00, 1.00)	0.13 (-7.84, 2.45)	0.29 (0.02, 9.56)
	平均 (全時点), <130/<80	0.07 (0.00, 1.00)	0.10 (-5.19, 2.79)	0.26 (0.01, 7.19)	0.17 (-4.97, 3.72)	0.32 (0.02, 10.20)	0.11 (0.00, 1.00)	0.15 (-5.08, 2.90)	0.32 (0.04, 8.44)
全例	平均 (2~48 週), <140/<90	0.09 (0.00, 1.00)	0.21 (-2.39, 2.55)	0.25 (0.01, 5.21)	0.06 (-0.63, 1.15)	0.10 (0.00, 2.82)	0.06 (0.00, 1.00)	0.14 (-1.44, 1.66)	0.18 (0.02, 4.44)
	平均 (2~48 週), <135/<85	0.31 (0.00, 1.00)	0.38 (-3.49, 3.74)	0.43 (0.05, 9.08)	0.23 (-2.73, 3.14)	0.28 (0.02, 5.94)	0.24 (0.00, 1.00)	0.31 (-3.43, 3.25)	0.37 (0.07, 7.84)
	平均 (2~48 週), <130/<80	0.09 (0.00, 1.00)	0.20 (-3.36, 2.89)	0.25 (0.02, 6.39)	0.20 (-2.21, 2.71)	0.24 (0.01, 5.71)	0.11 (0.00, 1.00)	0.21 (-2.72, 2.80)	0.26 (0.03, 6.23)
	平均 (全時点), <140/<90	0.62 (0.00, 1.00)	0.59 (-7.32, 5.01)	0.69 (0.13, 12.25)	0.09 (-1.44, 2.11)	0.16 (0.01, 4.81)	0.38 (0.00, 1.00)	0.34 (-3.21, 3.18)	0.43 (0.10, 8.30)
	平均 (全時点), <135/<85	0.49 (0.00, 1.00)	0.50 (-4.25, 5.23)	0.59 (0.10, 10.64)	0.36 (-3.60, 4.73)	0.43 (0.03, 8.54)	0.44 (0.00, 1.00)	0.44 (-4.48, 5.20)	0.53 (0.10, 10.39)
	平均 (全時点), <130/<80	0.28 (0.00, 1.00)	0.35 (-4.23, 3.84)	0.42 (0.06, 9.79)	0.35 (-4.66, 4.59)	0.40 (0.05, 9.01)	0.31 (0.00, 1.00)	0.36 (-4.20, 4.17)	0.42 (0.10, 8.08)

ACE 阻害薬併用集団での TE=0.12 (-0.05, 0.28); ACE 阻害薬併用集団での TE=0 (-0.1, 0.11); 全例での TE=0.04 (-0.06, 0.12)

表 4.7: ORIENT; 尿蛋白変化率の代替性評価 (期間平均を用いた検討)

尿蛋白変化率	NIE(x_0), NDE(x_0) による評価 ($\alpha = 0$)				NIE(x_1), NDE(x_1) による評価 ($\alpha = 1$)				中央値 (2.5%点, 97.5%点) による評価 ($\alpha = \text{pr}(x_1)$)			
	C-PCS	PI	R_G	R_G	C-PCS	PI	R_G	R_G	C-PCS	PI	R_G	R_G
ACE 阻害薬非併用	平均 (2~48 週), <0	0.80 (0.00, 1.00)	1.25 (-7.33, 8.06)	1.38 (0.24, 16.26)	0.79 (0.00, 1.00)	1.49 (-10.81, 9.56)	1.71 (0.52, 26.22)	0.78 (0.00, 1.00)	1.40 (-8.45, 8.18)	1.57 (0.49, 23.72)		
	平均 (2~48 週), <20	0.80 (0.00, 1.00)	1.15 (-6.75, 7.02)	1.31 (0.32, 18.08)	0.79 (0.00, 1.00)	1.34 (-7.39, 7.62)	1.55 (0.13, 17.56)	0.78 (0.00, 1.00)	1.27 (-7.55, 7.63)	1.41 (0.42, 16.66)		
	平均 (2~48 週), <40	0.78 (0.00, 1.00)	0.82 (-4.84, 6.31)	0.92 (0.30, 15.03)	0.86 (0.00, 1.00)	0.86 (-3.85, 5.36)	0.94 (0.41, 11.27)	0.82 (0.00, 1.00)	0.84 (-4.31, 5.81)	0.93 (0.37, 13.55)		
	平均 (全時点), <0	0.83 (0.00, 1.00)	1.15 (-6.16, 6.71)	1.30 (0.29, 18.23)	0.81 (0.00, 1.00)	1.60 (-8.69, 9.12)	1.78 (0.58, 20.26)	0.81 (0.00, 1.00)	1.38 (-6.76, 7.45)	1.54 (0.55, 19.08)		
	平均 (全時点), <20	0.85 (0.00, 1.00)	1.34 (-6.85, 8.80)	1.51 (0.56, 19.80)	0.86 (0.00, 1.00)	1.19 (-4.96, 6.78)	1.31 (0.44, 14.78)	0.85 (0.00, 1.00)	1.27 (-5.68, 7.72)	1.40 (0.58, 17.07)		
	平均 (全時点), <40	0.85 (0.00, 1.00)	1.16 (-6.87, 7.98)	1.33 (0.46, 13.36)	0.90 (0.00, 1.00)	1.13 (-4.82, 6.45)	1.21 (0.57, 10.60)	0.88 (0.00, 1.00)	1.14 (-5.70, 7.19)	1.27 (0.52, 11.84)		
ACE 阻害薬併用	平均 (2~48 週), <0	0.53 (0.00, 1.00)	0.55 (-17.92, 9.73)	1.09 (0.15, 35.04)	0.47 (0.00, 1.00)	0.43 (-16.99, 7.29)	0.82 (0.06, 24.54)	0.49 (0.00, 1.00)	0.50 (-17.72, 7.99)	0.95 (0.24, 28.37)		
	平均 (2~48 週), <20	0.54 (0.00, 1.00)	0.59 (-18.73, 12.29)	1.14 (0.21, 32.91)	0.30 (0.00, 1.00)	0.23 (-8.69, 7.18)	0.61 (0.03, 15.92)	0.47 (0.00, 1.00)	0.44 (-13.96, 10.73)	0.86 (0.22, 24.16)		
	平均 (2~48 週), <40	0.46 (0.00, 1.00)	0.44 (-16.40, 9.13)	0.86 (0.17, 20.33)	0.33 (0.00, 1.00)	0.26 (-15.69, 6.68)	0.68 (0.03, 21.25)	0.42 (0.00, 1.00)	0.38 (-15.10, 8.33)	0.77 (0.17, 19.93)		
	平均 (全時点), <0	0.34 (0.00, 1.00)	0.31 (-11.45, 8.28)	0.68 (0.06, 26.15)	0.14 (0.00, 1.00)	0.15 (-5.23, 3.96)	0.35 (0.02, 11.33)	0.29 (0.00, 1.00)	0.26 (-7.99, 5.44)	0.53 (0.08, 17.80)		
	平均 (全時点), <20	0.46 (0.00, 1.00)	0.42 (-16.78, 9.79)	0.94 (0.16, 25.09)	0.07 (0.00, 1.00)	0.02 (-5.63, 2.33)	0.29 (0.01, 12.18)	0.36 (0.00, 1.00)	0.21 (-9.24, 4.63)	0.62 (0.13, 18.40)		
	平均 (全時点), <40	0.60 (0.00, 1.00)	0.68 (-21.72, 17.37)	1.26 (0.27, 32.29)	0.54 (0.00, 1.00)	0.56 (-16.29, 11.01)	1.10 (0.17, 29.01)	0.59 (0.00, 1.00)	0.67 (-17.90, 14.05)	1.23 (0.27, 31.15)		
全例	平均 (2~48 週), <0	0.80 (0.00, 1.00)	1.17 (-13.62, 11.39)	1.45 (0.41, 27.63)	0.82 (0.00, 1.00)	1.27 (-12.40, 14.69)	1.47 (0.45, 26.92)	0.80 (0.00, 1.00)	1.21 (-14.93, 13.79)	1.47 (0.56, 26.44)		
	平均 (2~48 週), <20	0.81 (0.00, 1.00)	1.19 (-14.52, 12.31)	1.36 (0.46, 28.13)	0.78 (0.00, 1.00)	1.19 (-11.42, 12.54)	1.19 (0.13, 28.97)	0.79 (0.00, 1.00)	1.09 (-12.57, 11.94)	1.31 (0.43, 29.39)		
	平均 (2~48 週), <40	0.78 (0.00, 1.00)	0.92 (-11.88, 8.56)	1.06 (0.33, 21.41)	0.76 (0.00, 1.00)	0.78 (-6.69, 8.11)	0.90 (0.16, 15.96)	0.76 (0.00, 1.00)	0.84 (-8.87, 8.28)	0.99 (0.34, 19.33)		
	平均 (全時点), <0	0.79 (0.00, 1.00)	0.97 (-10.95, 8.65)	1.10 (0.30, 24.02)	0.82 (0.00, 1.00)	1.03 (-10.26, 9.97)	1.25 (0.41, 28.01)	0.80 (0.00, 1.00)	1.00 (-10.95, 10.08)	1.21 (0.42, 26.56)		
	平均 (全時点), <20	0.83 (0.00, 1.00)	1.21 (-13.31, 11.91)	1.43 (0.49, 32.29)	0.71 (0.00, 1.00)	0.70 (-5.50, 8.67)	0.82 (0.16, 19.10)	0.75 (0.00, 1.00)	0.98 (-9.61, 10.13)	1.12 (0.41, 27.40)		
	平均 (全時点), <40	0.83 (0.00, 1.00)	1.24 (-15.72, 12.23)	1.52 (0.50, 27.12)	0.86 (0.00, 1.00)	1.16 (-13.70, 10.82)	1.33 (0.50, 22.53)	0.84 (0.00, 1.00)	1.20 (-14.92, 11.49)	1.40 (0.51, 25.61)		

ACE 阻害薬併用集団での TE=0.12 (-0.05, 0.28); ACE 阻害薬併用集団での TE=0 (-0.1, 0.11); 全例での TE=0.04 (-0.06, 0.12)

表 4.8: ORIENT; 血清 Cr 値の逆数の代替性評価 (期間平均を用いた検討)

血清 Cr 値の逆数	NIE(x_{t0}), NDE(x_{t0}) による評価 ($\alpha = 0$)				NIE(x_{t1}), NDE(x_{t1}) による評価 ($\alpha = 1$)				中央値 (2.5%点, 97.5%点) による評価 ($\alpha = \text{pr}(x_{t1})$)			
	C-PCS	PI	R _G	R _C	C-PCS	PI	R _G	R _C	C-PCS	PI	R _G	R _C
ACE 阻害薬非併用	平均 (2~48 週), >-0.05	0.02 (0.00, 1.00)	-0.04 (-1.54, 1.20)	0.14 (0.00, 2.40)	0.02 (0.00, 1.00)	-0.03 (-1.46, 1.06)	0.13 (0.00, 2.63)	0.02 (0.00, 1.00)	0.02 (0.00, 1.00)	-0.04 (-1.50, 1.10)	0.14 (0.00, 2.59)	0.13 (0.01, 2.01)
	平均 (2~48 週), >-0.10	0.01 (0.00, 1.00)	0.00 (-0.99, 1.02)	0.13 (0.00, 1.98)	0.01 (0.00, 1.00)	0.00 (-1.24, 1.07)	0.12 (0.01, 2.34)	0.02 (0.00, 1.00)	0.02 (0.00, 1.00)	0.00 (-1.05, 1.06)	0.13 (0.01, 2.01)	0.13 (0.01, 2.01)
	平均 (2~48 週), >-0.15	0.00 (0.00, 0.47)	-0.01 (-0.44, 0.49)	0.06 (0.00, 1.08)	0.01 (0.00, 0.61)	0.03 (-0.62, 0.78)	0.09 (0.00, 2.00)	0.01 (0.00, 0.54)	0.01 (0.00, 0.54)	0.00 (-0.31, 0.30)	0.08 (0.00, 1.73)	0.08 (0.00, 1.73)
	平均 (全時点), >-0.05	0.01 (0.00, 1.00)	0.00 (-0.97, 1.19)	0.11 (0.00, 1.67)	0.06 (0.00, 1.00)	-0.02 (-2.58, 2.32)	0.27 (0.01, 4.66)	0.04 (0.00, 1.00)	0.04 (0.00, 1.00)	-0.01 (-1.70, 1.67)	0.20 (0.01, 3.43)	0.20 (0.01, 3.43)
	平均 (全時点), >-0.10	0.25 (0.00, 1.00)	0.35 (-0.99, 2.79)	0.40 (0.01, 5.53)	0.41 (0.00, 1.00)	0.45 (-1.29, 2.76)	0.50 (0.02, 5.85)	0.35 (0.00, 1.00)	0.35 (0.00, 1.00)	0.41 (-1.09, 2.79)	0.46 (0.02, 5.68)	0.46 (0.02, 5.68)
	平均 (全時点), >-0.15	0.79 (0.00, 1.00)	0.77 (-2.13, 4.33)	0.83 (0.09, 6.04)	0.60 (0.00, 1.00)	0.56 (-2.39, 3.54)	0.61 (0.06, 5.30)	0.70 (0.00, 1.00)	0.70 (0.00, 1.00)	0.67 (-2.40, 3.97)	0.73 (0.08, 5.83)	0.73 (0.08, 5.83)
ACE 阻害薬併用	平均 (2~48 週), >-0.05	0.06 (0.00, 1.00)	0.00 (-2.30, 4.15)	0.26 (0.01, 5.47)	0.11 (0.00, 1.00)	0.01 (-3.59, 5.17)	0.38 (0.01, 6.41)	0.09 (0.00, 1.00)	0.09 (0.00, 1.00)	0.01 (-2.98, 4.62)	0.33 (0.01, 6.78)	0.33 (0.01, 6.78)
	平均 (2~48 週), >-0.10	0.02 (0.00, 1.00)	-0.02 (-2.06, 2.42)	0.15 (0.00, 4.17)	0.12 (0.00, 1.00)	-0.10 (-4.20, 5.81)	0.41 (0.01, 10.12)	0.08 (0.00, 1.00)	0.08 (0.00, 1.00)	-0.07 (-3.13, 4.42)	0.30 (0.01, 7.58)	0.30 (0.01, 7.58)
	平均 (2~48 週), >-0.15	0.06 (0.00, 1.00)	-0.08 (-2.84, 3.71)	0.28 (0.01, 7.19)	0.15 (0.00, 1.00)	-0.21 (-5.59, 8.23)	0.59 (0.02, 14.26)	0.10 (0.00, 1.00)	0.10 (0.00, 1.00)	-0.15 (-3.82, 5.49)	0.44 (0.02, 10.83)	0.44 (0.02, 10.83)
	平均 (全時点), >-0.05	0.13 (0.00, 1.00)	0.15 (-8.50, 3.92)	0.33 (0.02, 14.53)	0.06 (0.00, 1.00)	0.11 (-5.56, 3.28)	0.23 (0.01, 10.32)	0.10 (0.00, 1.00)	0.10 (0.00, 1.00)	0.14 (-7.00, 3.78)	0.29 (0.02, 13.01)	0.29 (0.02, 13.01)
	平均 (全時点), >-0.10	0.17 (0.00, 1.00)	0.07 (-4.53, 6.07)	0.46 (0.02, 12.10)	0.13 (0.00, 1.00)	0.07 (-4.24, 5.95)	0.36 (0.01, 10.03)	0.16 (0.00, 1.00)	0.16 (0.00, 1.00)	0.07 (-4.19, 5.97)	0.41 (0.02, 9.97)	0.41 (0.02, 9.97)
	平均 (全時点), >-0.15	0.17 (0.00, 1.00)	-0.09 (-8.75, 11.53)	0.60 (0.02, 23.81)	0.21 (0.00, 1.00)	-0.09 (-10.10, 11.03)	0.67 (0.02, 29.16)	0.20 (0.00, 1.00)	0.20 (0.00, 1.00)	-0.09 (-9.19, 11.70)	0.64 (0.02, 26.40)	0.64 (0.02, 26.40)
全例	平均 (2~48 週), >-0.05	0.03 (0.00, 1.00)	-0.07 (-2.12, 2.01)	0.19 (0.01, 3.57)	0.04 (0.00, 1.00)	-0.08 (-2.47, 2.71)	0.23 (0.01, 7.27)	0.04 (0.00, 1.00)	0.04 (0.00, 1.00)	-0.08 (-2.37, 2.26)	0.21 (0.02, 6.22)	0.21 (0.02, 6.22)
	平均 (2~48 週), >-0.10	0.01 (0.00, 1.00)	-0.06 (-1.60, 1.14)	0.12 (0.01, 2.58)	0.04 (0.00, 1.00)	-0.13 (-3.71, 2.16)	0.23 (0.01, 5.66)	0.03 (0.00, 1.00)	0.03 (0.00, 1.00)	-0.08 (-3.25, 1.38)	0.18 (0.01, 4.30)	0.18 (0.01, 4.30)
	平均 (2~48 週), >-0.15	0.02 (0.00, 1.00)	-0.12 (-1.97, 1.91)	0.17 (0.01, 4.36)	0.05 (0.00, 1.00)	-0.21 (-2.44, 4.57)	0.29 (0.01, 6.51)	0.04 (0.00, 1.00)	0.04 (0.00, 1.00)	-0.16 (-2.21, 3.55)	0.24 (0.01, 5.68)	0.24 (0.01, 5.68)
	平均 (全時点), >-0.05	0.05 (0.00, 1.00)	0.13 (-1.28, 1.98)	0.21 (0.01, 3.19)	0.05 (0.00, 1.00)	0.09 (-1.45, 2.61)	0.22 (0.01, 3.38)	0.06 (0.00, 1.00)	0.06 (0.00, 1.00)	0.12 (-1.25, 2.19)	0.21 (0.02, 2.89)	0.21 (0.02, 2.89)
	平均 (全時点), >-0.10	0.12 (0.00, 1.00)	0.20 (-2.72, 3.94)	0.32 (0.01, 7.20)	0.14 (0.00, 1.00)	0.23 (-2.52, 3.72)	0.33 (0.02, 5.84)	0.14 (0.00, 1.00)	0.14 (0.00, 1.00)	0.23 (-2.17, 3.51)	0.33 (0.03, 6.87)	0.33 (0.03, 6.87)
	平均 (全時点), >-0.15	0.16 (0.00, 1.00)	0.24 (-3.14, 3.10)	0.34 (0.02, 7.87)	0.08 (0.00, 1.00)	0.12 (-3.48, 2.95)	0.27 (0.01, 8.59)	0.13 (0.00, 1.00)	0.13 (0.00, 1.00)	0.18 (-3.29, 2.81)	0.31 (0.04, 7.12)	0.31 (0.04, 7.12)

ACE 阻害薬非併用集団での TE=0.12 (-0.05, 0.28); ACE 阻害薬併用集団での TE=0 (-0.1, 0.11); 全例での TE=0.04 (-0.06, 0.12)

4.6 まとめ

本章では、第3章で提案した代替性評価尺度を統計的因果推論の観点から再定義した尺度として、自然な因果効果を用いた代替性評価尺度 **C-PCS** を提案し、これが既存の代替性評価尺度の問題点 (特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとる (Taylor et al., 2005; Albert, 2008)) を解決していることを確認した。また、**C-PCS** を **ORIENT** (Imai et al., 2011) に適用し、腎複合イベントに対する代替エンドポイントの候補について、従来の報告では代替性がないことが示唆されていたのに対し (Imai et al., 2011, 2013a,b), 本章の **C-PCS** を用いた検討では、代替性の評価対象時期や背景因子の影響を考慮した結果、いくつかの候補については、中程度の代替性があることが示唆された。ただ、この代替性評価尺度の 2.5% 点、97.5 % 点に基づく範囲は非常に広く、代替性の正確な評価のためには、より規模の大きい試験でこれらのことを確認するべきである。

本章の結果を用いることで、治療、代替エンドポイント、そして臨床エンドポイントを取り巻く因果メカニズムを視覚的に表現した上で、代替エンドポイントの候補によってどの程度治療効果を捕捉できるかを評価可能となることが期待され、統計的因果推論の議論をふまえた代替性の評価をするべきという臨床家からの要求 (Stevens et al., 2006; Manns et al., 2006) に応えることが可能と考えられる。

第5章 総括

5.1 本論文で取り組んだ課題と提案手法の特徴

本論文で取り組んだ課題は, 既存の統計的関連性を用いた代替性評価尺度の問題点を解決し, かつ臨床家でも理解しやすい性質を備えた代替性の評価尺度を提案すること, および統計的因果推論に関する基礎的な概念を確認したうえで, Pearl 流の統計的因果推論を用いた代替性の評価尺度を提案することであった. これらの課題に対し, 第2章では, 統計的因果推論を用いた代替性評価尺度を提案する準備として, 統計的因果推論に関する基礎的な背景(黒木・小林, 2012)を述べた. 続く第3章では, 新たな統計的関連性を用いた代替性評価尺度を提案し, 併せて最頻値を用いた代替性の評価方法も提案した (Kobayashi and Kuroki, 2014a). そして, 第4章では, 第3章で提案した代替性評価尺度を, 統計的因果推論の観点から再定義し, 自然な因果効果を用いた代替性評価尺度を提案した (Kobayashi and Kuroki, 2014b).

それぞれの課題に対して本論文で提案した手法の特徴を以下に記述する.

課題の一つ目に対して, 既存の統計的関連性を用いた代替性評価尺度の問題点を解決し, かつ臨床家でも理解しやすい性質を備えた代替性の評価尺度を提案した. 単一の無作為化比較試験における PTE, PE や PIG などの統計的関連性を用いた代替性の評価尺度を扱う研究の多くにおいて, 生物統計分野の研究者や実務家は, 三つの問題点を指摘していた (Lin et al., 1997; De Gruttola et al., 1997; Li et al., 2001; Wang and Taylor, 2002; Qu and Case, 2007). 一つ目の問題点は, 代替性評価尺度の多くは 特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとること, 二つ目として, 既存の代替性評価尺度は, 代替性の水準を判断するため

の適切なカットオフ値を定義できないこと,そして三つ目は,代替性評価尺度の多くは,そのばらつきが大きく,そのため信頼区間が許容できないほどに広いことである.一つ目の問題を解決するため,本論文では,統計的関連性を用いた新たな代替性評価尺度 **PCS** を,治療効果のうち代替エンドポイントの候補が捕捉する部分と,捕捉しない部分への分解に基づいて導入した.そして,二つ目の問題を解決するため,**PCS** の微分を用いて代替性の水準を判断するためのカットオフ値を与えた.ここで,最終的な代替性の判断は,このカットオフ値によるものだけではなく,臨床的かつ生物学的な観点に基づいてなされるべきであることを強調しておく.最後に,三つ目の問題を解決するため,推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法を提案した.

課題の二つ目に対しては,統計的因果推論に関する基礎的な概念を確認したうえで, Pearl 流の統計的因果推論を用いた代替性の評価尺度を提案した.統計的因果推論の議論をふまえた代替性の評価は,代替性を議論する際の根本的な要求であると考えられる.統計的因果推論に基づく既存の代替性評価尺度 (Taylor et al., 2005; Albert, 2008) は,特定の仮定をおかない限り範囲 $[0, 1]$ 外の値をとるという問題点があった.この問題を解決するため,本論文では自然な直接効果および間接効果を用いた代替性の評価尺度 **C-PCS** を提案した.

5.2 本論文の貢献

5.2.1 臨床的観点からみた本論文の貢献

無作為化比較試験における治療効果を判断するための評価項目(エンドポイント)の設定にあたっては,治療効果を判断するための感度があること,および臨床的な妥当性があることが重要である (Fleming, 1996).そして,臨床エンドポイントの代わりに評価項目として,代替エンドポイントが必要とされており,そこに求められる実用上の性質として,臨床エンドポイントより早期に,より簡便に測定できることが挙げられている (Molenberghs

et al., 2005).

Prentice (1989) に端を発する Freedman et al. (1992) の PTE などの代替性の評価尺度は、比較的簡単に算出できるため、近年になっても臨床家に広く利用されている (DePrimo et al., 2009; Boekholdt et al., 2012). しかし、前述の既存の代替性の評価尺度に対して、多くの問題点が指摘されているのも事実である (Lin et al., 1997; De Gruttola et al., 1997; Li et al., 2001; Wang and Taylor, 2002; Qu and Case, 2007). また、複数の無作為化比較試験データを用いたメタアナリシスアプローチ (Buyse et al., 2000) も有効な手法ではあるが、適切なデータ収集のための基盤が十分に整備されていないという問題点が残っている (Buyse et al., 2010; Sargent and Mandrekar, 2013). 加えて、メタアナリシスを用いて代替性を評価する場合は、次の二つの問題点が生じる. 一つ目は、メタアナリシスの対象となった複数の試験で用いられた薬剤に基づく代替性の結果を、それらとは機序の異なる薬剤の効果判定に適用可能であるか不明確である、という問題点がある (Prentice, 2009). 二つ目の問題点として、より早期に・より簡便に測定できる代替エンドポイントを探索する際に、候補となる測定値の測定時点や評価方法が対象となった複数の試験間で不均一である場合は、妥当な評価ができない、という問題点も考えられる (Molenberghs et al., 2005). これらのような問題点を考慮すると、利用可能な情報に制限が加わるメタアナリシスより、詳細な検討を加えることが可能な一つの試験に基づく代替性の検討は、依然として医学的に重要な意味を持つものと考えられる.

本論文で提案した代替性の評価尺度 PCS は、このような一つの試験に基づく代替性の評価を念頭においている. そして、PCS は既存の統計的関連性を用いた代替性評価尺度の問題点を解決し、かつ臨床家でも理解しやすい性質を備えていることから、臨床的観点からみて有用な尺度であると考えられる. 加えて、既存の代替性評価尺度の信頼区間が許容できないほどに広く代替性の評価が十分に実施できなかった問題についても、最頻値を用いた代替性の評価方法により解決されており、本論文で提案した手法を活用することで今後は代替性の定量的な検討が促進されると考えられる.

また, Pearl 流の統計的因果推論に基づき, 治療, 代替エンドポイント, そして臨床エンドポイントを取り巻く因果メカニズムを視覚的に表現した上で代替性を評価することにより, 臨床家と統計家が代替性の議論を綿密に行うことが期待される. この場面において, 本論文で提案した自然な直接効果および間接効果を用いた代替性の評価尺度 **C-PCS** は有用な尺度であると考えられる. 具体的には, **C-PCS** は **PCS** が持つ性質 (具体的には, 範囲 $[0, 1]$ 内の値をとる性質) を引き継いでおり, 臨床家でも理解しやすい代替性の評価尺度となっている. これらのことが, 臨床的観点からみた本論文の貢献である.

5.2.2 統計的観点からみた本論文の貢献

本論文で提案した代替性の評価尺度 **PCS** は, 範囲 $[0, 1]$ 内の値をとること, 代替性の水準を判断するための適切なカットオフ値を定義できることが特徴であった. 前者の範囲 $[0, 1]$ 内の値をとることについては, この性質を満足する尺度を統計家が提供するのとは自然な考えである. また, 後者の代替性の水準を判断するための適切なカットオフ値を定義できる特徴については, そのカットオフ値が **PCS** の微分に基づいて与えられていることから, ここに統計的な貢献があると考えられる. ただし, 第3章でも述べたとおり, 与えられたカットオフ値は, 代替性のベンチマークとして利用可能ではあるが, 最終的な代替性の判断は, このカットオフ値によるものだけではなく, 臨床的かつ生物学的な観点に基づいてなされるべきである.

次に本論文では, 既存の代替性評価尺度の多くが持つ“信頼区間が許容できないほど広い”という問題を解決するために, 推定された代替性評価尺度のブートストラップ分布の最頻値を用いた代替性の評価方法を提案した. ここで, この問題に対する本論文の解決の意味について言及しておきたい. この問題に対する統計的観点からみた素直な解決方法は, 既存の代替性評価尺度そのものの推定精度向上を狙った方法の提案である. しかし, 本論文で提案した最頻値を用いた代替性の評価方法は, 尺度そのものの推定精度を向上させて

いるわけではなく、多重共線性の問題の回避を狙って、推定された尺度の中で最も頻繁に出現するという意味で適切な値を推定値としていることを強調しておく。なお、提案した代替性の評価方法における最頻値は興味ある尺度の一致推定量となることは担保されていない。例えば、興味ある尺度の分布が左右対称な分布ではなかった場合には、一致推定量とはならないであろう。しかしながら、本論文で提案した最頻値を用いた代替性の評価方法では、代替性評価尺度の点推定値ではなく、信頼区間の上限と下限を利用する方針をとっている。この手続きにしたがえば、点推定値として一致推定量が得られない状況であっても、第3章の数値実験や無作為化比較試験への適用例で確認したように、代替性の妥当な評価が可能となっていることがわかる。

5.3 今後の展開および課題

今後の展開として、メタアナリシスの設定 (Molenberghs et al., 2005) に、提案した代替性評価尺度を適応することが考えられる。例えば、複数の試験における PCS の結果を用いて、統合的な PCS を算出することができる。その際、PCS の単純な平均を用いるのか、各試験の例数や分散などの重み付け平均を用いるかなどの検討が必要である。また、統合された PCS の分散および信頼区間の算出方法は明確ではなく、代替性の水準の評価方法を考慮した議論が必要であり、この点も今後の検討課題である。なお、第3章で述べたように PCS に PE との関係式 (3.9) が存在するので、PE を用いて報告された既存研究も、統合的な PCS を算出に利用することができる。

もう一つの展開として、代替エンドポイントの候補、および臨床エンドポイントに欠測値が存在する場面において、欠測値の影響を考慮した代替性の評価も重要である。本論文では検討する代替エンドポイントの候補、および臨床エンドポイントには欠測値がないことを前提としていたが、仮に背後に特定の欠測メカニズムが存在していた場合には、代替性の評価にメカニズムに応じたバイアスが混入する可能性がある。Fleming (2005) は、代替

エンドポイントの候補, および臨床エンドポイントの測定値が存在していない集団での検討がバイアスを導く可能性があるとして指摘している. Tsiatis et al. (1995) は経時測定データ解析の枠組みの中で欠測パターンを考慮した検討を行っているが, そこでは代替性の評価への影響は十分に検討されていない. PCS および C-PCS の利用においても, この問題を十分に考慮する必要があるが, その詳細は今後の重要な課題の一つと考えている.

第4章の自然な因果効果を用いた代替性評価尺度 C-PCS の定式化における, 二つの項それぞれの重要性を表すパラメータの設定方法についても臨床的な観点を考慮した検討が必要である. 具体的には, パラメータ α の値は, 興味ある処理群の選択に応じて, 統計家と臨床家の議論を踏まえて決めるべきである. 第4章の事例として, “ $X = x'$ が比較の基準値であり, それに対する $X = x$ の間接効果を検討する場合には $\alpha = 0$ とする”ことを述べていたが, この設定は, プラセボ群を比較の基準値とし, 代替エンドポイントの候補における実薬群とプラセボ群の差がどの程度の代替性をもつかを考える状況に利用価値があると考えられる.

最後に, 第4章において提案した C-PCS の算出においては, 代替エンドポイントの候補と臨床エンドポイントの間の交絡因子が同定されていることを想定していたが, 現実的にはこの想定は困難を伴うことがある. 例えば, 交絡因子同定のための十分な変数集合を観測できない場合である. このような場合には, 因果効果の存在範囲 (Cai et al., 2008; Sjölander, 2009) や, 感度分析 (VanderWeele, 2010) を用いた検討が有効であると考えられる. これらの実際の適用については, 今後の重要な検討課題である.

5.4 まとめ

本論文の冒頭で示した問い“患者にとって価値のある医薬品を, 1日でも早く提供するにはどうすれば良いか”に対して, 代替エンドポイントの利用が一つの有用なアプローチである. その一方で, 統計的関連性を用いた既存の代替性評価尺度は, 代替性の尺度の求めら

れる性質を満たしていなかったため、定量的な代替性の評価は困難であった。この問題点を解決した統計的関連性を用いた代替性評価尺度 **PCS** と最頻値に基づく代替性の評価方法を用いることで、今後は代替性の定量的な検討が可能になるであろう。また、高橋 (1967) が示した“病気の治癒のしくみについての的を射たものであるかどうかがよく吟味されていなければならない”に対しても、本論文で提案した統計的因果推論の考え方を用いた自然な直接効果および間接効果を用いた代替性評価尺度 **C-PCS** により、統計的因果推論に基づく解釈を含めた検討が可能となる。以上、本論文で提案した手法を活用することで、今後は、定量的な代替エンドポイントの検討が可能になり、本論文の冒頭に示した問いに科学的に答えるための一助になると考えられる。

引用文献

- Albert, J. M. (2008) Mediation analysis via potential outcomes models. *Statistics in medicine*, 27 (8), 1282–1304.
- Albert, J. M., Nelson, S. (2011) Generalized causal mediation analysis. *Biometrics*, 67 (3), 1028–1038.
- Balke, A., Pearl, J. (1997) Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92 (439), 1171–1176.
- Barter, P. J., Caulfield, M., Eriksson, M., Grundy, S. M., Kastelein, J. J., Komajda, M., Lopez-Sendon, J., Mosca, L., Tardif, J. C., Waters, D. D., Shear, C. L., Revkin, J. H., Buhr, K. A., Fisher, M. R., Tall, A. R., Brewer, B. (2007) Effects of torcetrapib in patients at high risk for coronary events. *New England Journal of Medicine*, 357 (21), 2109–2122.
- Berk, R. A. (1987) Causal inference as a prediction problem. *Crime & Just.*, 9, 183.
- Bickel, D. R. (2002) Robust estimators of the mode and skewness of continuous data. *Computational Statistics and Data Analysis*, 39 (2), 153–163.
- Bickel, D. R., Frühwirth, R. (2006) On a fast, robust estimator of the mode: Comparisons to other robust estimators with applications. *Computational Statistics and Data Analysis*, 50 (12), 3500–3530.
- Boekholdt, S. M., Arsenault, B. J., Mora, S., Pedersen, T. R., LaRosa, J. C., Nestel, P. J., Simes, R. J., Durrington, P., Hitman, G. A., Welch, K. M., DeMicco, D. A., Zwinderman,

- A. H., Clearfield, M. B., Downs, J. R., Tonkin, A. M., Colhoun, H. M., Gotto, A. M., Ridker, P. M., Kastelein, J. J. (2012) Association of LDL cholesterol, non-HDL cholesterol, and apolipoprotein B levels with risk of cardiovascular events among patients treated with statins: a meta-analysis. *Journal of the American Medical Association*, 307 (12), 1302–1309.
- Bollen, K. A. (1989) *Structural Equations with Latent Variables*, Wiley-Interscience, 1st edition.
- Bowden, R. J., Turkington, D. A. (1984) *Instrumental Variables*, Cambridge University Press.
- Brenner, B. M., Cooper, M. E., de Zeeuw, D., Keane, W. F., Mitch, W. E., Parving, H.-H., Remuzzi, G., Snapinn, S. M., Zhang, Z., Shahinfar, S. (2001) Effects of losartan on renal and cardiovascular outcomes in patients with type 2 diabetes and nephropathy. *New England Journal of Medicine*, 345 (12), 861–869.
- Buyse, M., Molenberghs, G. (1998) Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*, 54 (3), 1014–1029.
- Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., Geys, H. (2000) The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, 1 (1), 49–67.
- Buyse, M., Sargent, D. J., Grothey, A., Matheson, A., De Gramont, A. (2010) Biomarkers and surrogate end points—the challenge of statistical validation. *Nature reviews Clinical oncology*, 7 (6), 309–317.
- Cai, Z., Kuroki, M. (2007) Evaluating treatment effects with instrumental variables. *The 9th Japan- China Joint Symposium on Statistics*, 17–20.
- Cai, Z., Kuroki, M. (2008) On identifying total effects in the presence of latent variables and selection bias. *The 24th Conference on Uncertainty in Artificial Intelligence*, 62–69.

- Cai, Z., Kuroki, M., Sato, T. (2007) Non-parametric bounds on treatment effects with non-compliance by covariate adjustment. *Statistics in medicine*, 26 (16), 3188–3204.
- Cai, Z., Kuroki, M., Pearl, J., Tian, J. (2008) Bounds on direct effects in the presence of confounded intermediate variables. *Biometrics*, 64 (3), 695–701.
- Chen, H., Geng, Z., Jia, J. (2007) Criteria for surrogate end points. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69 (5), 919–932.
- Cole, S. R., Frangakis, C. E. (2009) The consistency statement in causal inference: a definition or an assumption? *Epidemiology*, 20 (1), 3–5.
- Cox, D. R. (1958) *Planning of Experiments*, John Wiley & Sons.
- Csaky, K. G., Richman, E. A., Ferris, F. L. (2008) Report from the NEI/FDA Ophthalmic Clinical Trial Design and Endpoints Symposium. *Investigative Ophthalmology and Visual Science*, 49 (2), 479–489.
- Davison, A., Hinkley, D. (1997) *Bootstrap Methods and Their Application*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- Dawid, A. P. (2000) Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95 (450), 407–424.
- De Gruttola, V., Fleming, T., Lin, D. Y., Coombs, R. (1997) Perspective: validating surrogate markers—are we being naive?. *Journal of Infectious Diseases*, 175 (2), 237–246.
- DePrimo, S. E., Huang, X., Blackstein, M. E., Garrett, C. R., Harmon, C. S., Schöffski, P., Shah, M. H., Verweij, J., Baum, C. M., Demetri, G. D. (2009) Circulating levels of soluble kit serve as a biomarker for clinical outcome in gastrointestinal stromal tumor patients receiving sunitinib following imatinib failure. *Clinical Cancer Research*, 15 (18), 5869–5877.

- Edwards, D. (2000) *Introduction to Graphical Modelling*, Springer, 2nd edition.
- Efron, B., Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*, New York, Chapman & Hall.
- Ellerbeck, E. F., Jencks, S. F., Radford, M. J., Timothy, F., Craig, A. S., Gold, J. A., Krumholz, H. M., Vogel, R. A. (1995) Quality of care for medicare patients with acute myocardial infarction: a four-state pilot study from the cooperative cardiovascular project. *Journal of the American Medical Association*, 273 (19), 1509–1514.
- Fisher, R. A. (1925) *Statistical methods for research workers*, Hafner, (遠藤健児・鍋谷清治 訳, 「研究者のための統計的方法」, 森北出版, 1970年) .
- Fleming, T. R. (1996) Surrogate endpoints in clinical trials. *Drug Information Journal*, 30 (2), 545–551.
- Fleming, T. R. (2005) Objective response rate as a surrogate end point: a commentary. *Journal of clinical oncology*, 23 (22), 4845–4846.
- Frangakis, C. E., Rubin, D. B. (2002) Principal stratification in causal inference. *Biometrics*, 58 (1), 21–29.
- Freedman, L. S., Graubard, B. I., Schatzkin, A. (1992) Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, 11 (2), 167–178.
- Galles, D., Pearl, J. (1997) Axioms of causal relevance. *Artificial Intelligence*, 97 (1), 9–43.
- Galles, D., Pearl, J. (1998) An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3 (1), 151–182.

- Gan, S. C., Beaver, S. K., Houck, P. M., MacLehose, R. F., Lawson, H. W., Chan, L. (2000) Treatment of acute myocardial infarction and 30-day mortality among women and men. *New England Journal of Medicine*, 343 (1), 8–15.
- Geng, Z., Li, G. (2002) Conditions for non-confounding and collapsibility without knowledge of completely constructed causal diagrams. *Scandinavian journal of statistics*, 29 (1), 169–181.
- Greenland, S. (2000) An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology*, 29 (4), 722–729.
- Greenland, S. (2003) Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, 14 (3), 300–306.
- Greenland, S., Pearl, J., Robins, J. M. (1999) Causal diagrams for epidemiologic research. *Epidemiology*, 10 (1), 37–48.
- Guo, S. Y., Fraser, M. W. (2009) *Propensity Score Analysis: Statistical Methods and Applications*, SAGE Publications, Inc.
- Holland, P. W. (1986) Statistics and causal inference. *Journal of the American statistical Association*, 81 (396), 945–960.
- Holtkamp, F. A., de Zeeuw, D., Thomas, M. C., Cooper, M. E., de Graeff, P. A., Hillege, H. J., Parving, H.-H., Brenner, B. M., Shahinfar, S., Heerspink, H. J. L. (2011) An acute fall in estimated glomerular filtration rate during treatment with losartan predicts a slower decrease in long-term renal function. *Kidney international*, 80 (3), 282–287.
- Huang, Y., Valtorta, M. (2006) Pearl’s calculus of intervention is complete. *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, 13–16.

- Huang, Y., Valtorta, M. (2008) On the completeness of an identifiability algorithm for semi-markovian models. *Annals of Mathematics and Artificial Intelligence*, 54 (4), 363–408.
- Imai, E., Chan, J., Ito, S. f., Yamasaki, T., Kobayashi, F., Haneda, M., Makino, H. (2011) Effects of olmesartan on renal and cardiovascular outcomes in type 2 diabetes with overt nephropathy: a multicentre, randomised, placebo-controlled study. *Diabetologia*, 54 (12), 2978–2986.
- Imai, E., Haneda, M., Yamasaki, T., Kobayashi, F., Harada, A., Ito, S., Chan, J. C., Makino, H. (2013a) Effects of dual blockade of the renin–angiotensin system on renal and cardiovascular outcomes in type 2 diabetes with overt nephropathy and hypertension in the orient: a post-hoc analysis (orient–hypertension). *Hypertension Research*, 36 (12), 1051–1059.
- Imai, E., Haneda, M., Chan, J. C., Yamasaki, T., Kobayashi, F., Ito, S., Makino, H. (2013b) Reduction and residual proteinuria are therapeutic targets in type 2 diabetes with overt nephropathy: a post hoc analysis (orient-proteinuria). *Nephrology Dialysis Transplantation*, 28 (10), 2526–2534.
- Ingelsson, E., Schaefer, E. J., Contois, J. H., McNamara, J. R., Sullivan, L., Keyes, M. J., Pencina, M. J., Schoonmaker, C., Wilson, P. W., D’Agostino, R. B., Vasan, R. S. (2007) Clinical utility of different lipid measures for prediction of coronary heart disease in men and women. *Journal of the American Medical Association*, 298 (7), 776–785.
- Joffe, M. M., Greene, T. (2009) Related causal frameworks for surrogate outcomes. *Biometrics*, 65 (2), 530–538.
- Ju, C., Geng, Z. (2010) Criteria for surrogate end points based on causal distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72 (1), 129–142.

- Kobayashi, F., Kuroki, M. (2014a) A new proportion measure of the treatment effect captured by candidate surrogate endpoints. *Statistics in medicine*, 33 (19), 3338–3353.
- Kobayashi, F., Kuroki, M. (2014b) Causal measures of the treatment effect captured by candidate surrogate endpoints. submitted.
- Kuroki, M. (2007) Graphical identifiability criteria for causal effects in studies with an unobserved treatment/response variable. *Biometrika*, 94 (1), 37–47.
- Kuroki, M. (2012) Optimizing a control plan using a structural equation model with an application to statistical process analysis. *Journal of Applied Statistics*, 39 (3), 673–694.
- Kuroki, M., Cai, Z. (2008) Formulating tightest bounds on causal effects in studies with unmeasured confounders. *Statistics in medicine*, 27 (30), 6597–6611.
- Kuroki, M., Cai, Z. (2011) Statistical Analysis of ‘Probabilities of Causation’ Using Co-variate Information, *Scandinavian Journal of Statistics*, 38 (3), 564–577.
- Kuroki, M., Cai, Z., Geng, Z. (2010) Sharp bounds on causal effects in case-control and cohort studies. *Biometrika*, 97 (1), 123–132.
- Kuroki, M., Miyakawa, M. (1999) Identifiability criteria for causal effects of joint interventions. *Journal of the Japanese Statistical Society*, 29, 105–117.
- Kuroki, M., Miyakawa, M. (2003) Covariate selection for estimating the causal effect of control plans by using causal diagrams. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (1), 209–222.
- Kuroki, M., Pearl, J. (2014) Measurement bias and effect restoration in causal inference. *Biometrika*, 101 (2), 423–437.

- Landis, J. R., Koch, G. G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159–174.
- Lauritzen, S. L. (2004) Discussion on causality. *Scandinavian Journal of Statistics*, 31 (2), 189–193.
- Lewis, D. K. (1973) *Counterfactuals*, Wiley-Blackwell, (吉満昭宏訳, 「反事実条件法」, 勁草書房, 2007年) .
- Li, Z., Meredith, M. P., Hoseyni, M. S. (2001) A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in Medicine*, 20 (21), 3175–3188.
- Lin, D. Y., Fleming, T. R., De Gruttola, V. (1997) Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine*, 16 (13), 1515–1527.
- Little, R. J. A., Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, Wiley-Interscience, 2nd edition.
- Liu, J., Sempos, C. T., Donahue, R. P., Dorn, J., Trevisan, M., Grundy, S. M. (2006) Non-high-density lipoprotein and very-low-density lipoprotein cholesterol and their risk predictive values in coronary heart disease. *American Journal of Cardiology*, 98 (10), 1363–1368.
- MacLehose, R. F., Kaufman, S., Kaufman, J. S., Poole, C. (2005) Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology*, 16 (4), 548–555.
- Manns, B., Owen Jr, W. F., Winkelmayr, W. C., Devereaux, P., Tonelli, M. (2006) Surrogate markers in clinical studies: problems solved or created? *American journal of kidney diseases*, 48 (1), 159–166.
- Molenberghs, G., Burzykowski, T., Buyse, M. (2005) *The Evaluation of Surrogate Endpoints*, Statistics for Biology and Health, Springer.

- Murphy, S. A. (2003) Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65 (2), 331–355.
- Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., Joffe, M. M., Glynn, R. J. (2011a) Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American journal of epidemiology*, 174 (11), 1213–1222.
- Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., Glynn, R. J. (2011b) Myers et al. Respond to “Understanding Bias Amplification”, *American journal of epidemiology*, 174 (11), 1228–1229.
- Nakamura, H., Arakawa, K., Itakura, H., Kitabatake, A., Goto, Y., Toyota, T., Nakaya, N., Nishimoto, S., Muranaka, M., Yamamoto, A., Mizuno, K., Ohashi, Y. (2006) Primary prevention of cardiovascular disease with pravastatin in Japan (MEGA Study): a prospective randomised controlled trial. *Lancet*, 368 (9542), 1155–1163.
- Neyman, J. (1935) Statistical problems in agricultural experimentation. *Journal of the Royal Statistical Society*, 2, 107–154.
- Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann.
- Pearl, J. (1994) From bayesian networks to causal networks. In Gammerman, A. ed. *Bayesian Networks and Probabilistic Reasoning*, 1–31.
- Pearl, J. (1995) Causal diagrams for empirical research. *Biometrika*, 82 (4), 669–688.
- Pearl, J. (2009a) *Causality: Models, Reasoning and Inference*, Cambridge University Press, 2nd edition, (黒木学訳, 「統計的因果推論-モデル・推論・推測-」, 共立出版, 2009年) .

- Pearl, J. (2009b) Myth, confusion, and science in causal analysis. *UCLA Cognitive Systems Laboratory, Technical Report (R-348)*.
- Pearl, J. (2009c) Letter to the editor: Remarks on the method of propensity score. *Statistics in medicine*, 28 (9), 1415–1416.
- Pearl, J. (2010a) On the consistency rule in causal inference: axiom, definition, assumption, or theorem? *Epidemiology*, 21 (6), 872–875.
- Pearl, J. (2010b) An introduction to causal inference. *The International Journal of Biostatistics*, 6 (2).
- Pearl, J. (2010c) On a class of bias-amplifying covariates that endanger effect estimates. In Grunwald, P., Spirtes, P. eds. *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, 417–424.
- Pearl, J. (2011) Invited commentary: understanding bias amplification. *American journal of epidemiology*, 174 (11), 1223–1227.
- Pharmacological therapy for macular degeneration study (1997) Interferon alfa-2a is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration. Results of a prospective randomized placebo-controlled clinical trial. Pharmacological Therapy for Macular Degeneration Study Group. *Archives of Ophthalmology*, 115 (7), 865–872.
- Prentice, R. L. (1989) Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine*, 8 (4), 431–440.
- Prentice, R. L. (2009) Surrogate and mediating endpoints: current status and future directions. *Journal of the National Cancer Institute*.

- Psaty, B. M., Lumley, T. (2008) Surrogate end points and FDA approval: a tale of 2 lipid-altering drugs. *Journal of the American Medical Association*, 299 (12), 1474–1476.
- Qu, Y., Case, M. (2006) Quantifying the indirect treatment effect via surrogate markers. *Statistics in Medicine*, 25 (2), 223–231.
- Qu, Y., Case, M. (2007) Quantifying the effect of the surrogate marker by information gain. *Biometrics*, 63 (3), 958–962.
- Robins, J. (1986) A new approach to causal inference in mortality studies with a sustained exposure period? application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7 (9), 1393–1512.
- Rosenbaum, P. R. (2002) *Observational Studies*, Springer, 2nd edition.
- Rosenbaum, P. R., Rubin, D. B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70 (1), 41–55.
- Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies.. *Journal of educational Psychology*, 66 (5), 688–701.
- Rubin, D. B. (1978) Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, 6 (1), 34–58.
- Rubin, D. B. (1980) Comment: Discussion of “randomization analysis of experimental data in the fisher randomization test” by d. basu. *Journal of the American Statistical Association*, 75 (371), 591–593.
- Rubin, D. B. (1986) Comment: Which ifs have causal answers. *Journal of the American Statistical Association*, 81 (396), 961–962.

- Rubin, D. B. (2005) Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100 (469), 322–331.
- Rubin, D. B. (2006) *Matched Sampling for Causal Effects*, Cambridge University Press.
- Rubin, D. B. (2008) Author’s reply (to ian shrier’s letter to the editor). *Statistics in Medicine*, 27, 2741–2742.
- Sargent, D. J., Mandrekar, S. J. (2013) Statistical issues in the validation of prognostic, predictive, and surrogate biomarkers. *Clinical Trials*, 10 (5), 647–652.
- Shpitser, I., Pearl, J. (2006a) Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence*, 1219–1226.
- Shpitser, I., Pearl, J. (2006b) Identification of conditional interventional distributions. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, 437–444.
- Shrier, I. (2008) Letter to the editor. *Statistics in Medicine*, 27, 2740–2741.
- Shrier, I. (2009) Letter to the editor: Propensity scores. *Statistics in Medicine*, 28 (8), 1317–1318.
- Shrier, I., Platt, R. W. (2008) Reducing bias through directed acyclic graphs. *BMC medical research methodology*, 8 (1), 70.
- Sjölander, A. (2009) Bounds on natural direct effects in the presence of confounded intermediate variables. *Statistics in Medicine*, 28 (4), 558–571.
- Spirtes, P., Glymour, C., Scheines, R. (2001) *Causation, Prediction and Search*, The MIT Press, 2nd edition.

- Stevens, L. A., Greene, T., Levey, A. S. (2006) Surrogate end points for clinical trials of kidney disease progression. *Clinical Journal of the American Society of Nephrology*, 1 (4), 874–884.
- Taylor, J. M., Wang, Y., Thiébaud, R. (2005) Counterfactual links to the proportion of treatment effect explained by a surrogate marker. *Biometrics*, 61 (4), 1102–1111.
- Teramoto, T., Nakaya, N., Yokoyama, S., Ohashi, Y., Mizuno, K., Nakamura, H. (2010) Association between lowering low-density lipoprotein cholesterol with pravastatin and primary prevention of cardiovascular disease in mild to moderate hypercholesterolemic Japanese. *Journal of Atherosclerosis and Thrombosis*, 17 (8), 879–887.
- The ONTARGET Investigators (2008) Telmisartan, ramipril, or both in patients at high risk for vascular events. *The New England Journal of Medicine*, 358, 1547–1559.
- Tian, J., Pearl, J. (2000) Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28, 287–313.
- Tian, J., Pearl, J. (2002) A general identification condition for causal effects. In *Proceedings of the 18th National Conference on Artificial Intelligence*, 567–573.
- Tsiatis, A. A., Degruetola, V., Wulfsohn, M. S. (1995) Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90 (429), 27-37.
- Van Ryzin, J. (1973) A histogram method of density estimation. *Communications in Statistics*, 2 (6), 493-506.
- VanderWeele, T. J. (2011) Causal mediation analysis with survival data. *Epidemiology*, 22 (4), 582–585.

- VanderWeele, T. J. (2009a) Concerning the consistency assumption in causal inference. *Epidemiology*, 20 (6), 880–883.
- VanderWeele, T. J. (2009b) Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20 (1), 18–26.
- VanderWeele, T. J. (2010) Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, 21 (4), 540–551.
- VanderWeele, T. J. (2013) Surrogate measures and consistent surrogates. *Biometrics*, 69 (3), 561–565.
- VanderWeele, T. J., Mukherjee, B., Chen, J. (2012) Sensitivity analysis for interactions under unmeasured confounding. *Statistics in medicine*, 31 (22), 2552–2564.
- VanderWeele, T. J., Hernán, M. A. (2011) Causal inference under multiple versions of treatment.
- VanderWeele, T. J., Shpitser, I. (2011) A new criterion for confounder selection. *Biometrics*, 67 (4), 1406–1413.
- Verteporfin in photodynamic therapy study group (2001) Verteporfin therapy of subfoveal choroidal neovascularization in age-related macular degeneration: two-year results of a randomized clinical trial including lesions with occult with no classic choroidal neovascularization—verteporfin in photodynamic therapy report 2. *American Journal of Ophthalmology*, 131 (5), 541–560.
- Wang, Y., Taylor, J. M. (2002) A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics*, 58 (4), 803–812.

Weir, C. J., Walley, R. J. (2006) Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Statistics in Medicine*, 25 (2), 183–203.

Wold, H. (1954) Causality and econometrics. *Econometrica*, 162–177.

Wright, S. (1923) The theory of path coefficients a reply to niles's criticism. *Genetics*, 8 (3), 239–255.

Wright, S. (1934) The method of path coefficients. *The Annals of Mathematical Statistics*, 5 (3), 161–215.

宮川雅巳 (2004) 「統計的因果推論-回帰分析の新しい枠組み-」, 朝倉書店.

高橋暁正 (1967) 「現代医学概論」, 東京大学出版会.

黒木学 (2007a) 統計的因果推論—因果的効果の識別可能問題におけるベイジアンネットワークの役割—, 「人工知能学会誌」, 22, 328–335.

黒木学 (2007b) 未観測交絡因子が存在する場合の総合効果の識別可能条件, 「応用統計学」, 36, 71–85.

黒木学 (2008) 分散に対する因果効果の定量的評価と工程解析への応用, 「品質」, 38, 87–98.

黒木学・小林史明 (2012) 構造的因果モデルについて, 「計量生物学」, 32, 119–144.

黒木学・陳希 (2010) 線形構造方程式モデルに基づく因果効果のグラフィカル識別可能条件, 「人工知能学会誌」, 25, 819–825.

佐藤俊哉 (2006) ランダム化にもとづく解析, 丹後俊郎・上坂浩之 (編) 「臨床試験ハンドブック」, 朝倉書店, 535–556.

佐藤俊哉・松山裕 (2002) 疫学・臨床研究における因果推論, 「統計科学のフロンティア 5 「多変量解析の展開-隠れた構造と因果を推理する」」, 岩波書店, 133–175.

田中鮎夢 (2014) 自由貿易協定 (FTA) の効果, 「経済産業研究所国際貿易と貿易政策研究メモ (RIETI International Trade and Trade Policy: Research Notes)」, 22, 1-6.

田中司朗・大庭幸治・吉村健一・手良向聡 (2010) 代替エンドポイントの評価のための統計的基準とその適用事例, 「計量生物学」, 31 (1), 23-48.

野口晴子・茅野真男・清水谷諭・川渕孝一 (2003) 急性心筋梗塞疾患患者への PTCA 施行を用いた医療評価の方法とプロセスの研究, *ESRI Discussion Paper* (58).

謝辞

本論文の審査にあたり、統計数理研究所の椿広計教授、黒木学准教授、逸見昌之准教授、船渡川伊久子准教授、成蹊大学の岩崎学教授、名古屋大学の松井茂之教授には、大変お忙しい中審査を引き受けていただいた。審査中にいただいた多くの有益な意見や指摘のおかげで、本論文は大きな進展を遂げることができた。

主査の椿広計教授には、私の入学時の受け入れ教官として大変お世話になり、さらに統計的因果推論に関して多くの指導をいただいた。また、主任指導教官の黒木学准教授には、私が掲げた(なかば無謀ともいえる)高い目標に対し、これを達成できるよう非常に多くの指導をいただいた。特に、社会人学生特有の時間的制約を踏まえながらも、目標達成に向けて適切な指摘と激励をいただいた。これらのおかげで、本研究は当初の目標を達成できたと言っても過言ではない。心から御礼を申し上げたい。

所属先である第一三共株式会社の諸先輩方には、私がか社での職務と並行して学位を取得することへの理解と支援をいただいた。さらに、困難な状況に陥った際にも多くの叱咤激励をいただいた。彼らから得た科学的な考え方や情熱は、間違いなく本研究活動の土台となっている。また本研究は、会社が主体となり実施した臨床試験 (MEGA study, ORIENT) のデータを用いて代替性を検討している。このようなデータに巡り合えたことは本当に幸運なことであった。これらの試験に参加してくださった患者や試験の計画実行に関った多くの人達の努力と貢献に感謝したい。

最後に、4年に渡る研究生生活を支えてくれた妻の小林理恵子には、言葉では表現できない程の心労をかけた。彼女の存在なしではこの学位取得は考えられなかったであろう。彼女の献身的な貢献に心から感謝の言葉を贈りたい。