

標的試験エミュレーション：観察研究における因果推論を強化するためのフレームワーク

深澤 俊貴^{*1}, 篠崎 智大^{*2}

^{*1} 京都大学大学院医学研究科社会健康医学系専攻薬剤疫学分野

^{*2} 東京理科大学工学部情報工学科

＜抄録＞ランダム化比較試験（RCT）が実施不可能、非倫理的、あるいは時宜を逸するような状況においては、観察データに基づく因果推論が科学的・臨床的意思決定への有効な選択肢となりうる。しかしながら、観察研究はランダム化の欠如による交絡に加え、不適切な研究デザインがもたらす選択バイアスや不死時間など、効果推定を根本的に歪める脆弱性を内包している。こうした課題に対処するための体系的なフレームワークとして、「標的試験エミュレーション（target trial emulation）」が近年注目を集めている。このアプローチは、関心のある因果的な問いに答えうる仮想のプラグマティック RCT（標的試験）のプロトコルを精緻に設計し、それを利用可能な観察データを用いて明示的に模倣（エミュレート）する形で研究を実施する２段階プロセスから成る。その最大の貢献は、観察研究における因果的な問いの曖昧さを排し、明確に定義された因果的な推定対象へと昇華させることにある。本稿では、標的試験エミュレーションの概念的基盤を整理するとともに、実装における考慮事項を詳述する。また、このフレームワークの適用事例として、骨粗鬆症を併存する維持透析患者を対象にデノスマブと経口ビスホスホネート製剤の心血管安全性および骨折予防効果を比較した観察研究を紹介する。標的試験フレームワークは、デザインに起因するバイアスを回避するとともに、観察データに内在する限界を浮き彫りにし、因果的な問いに向き合う疫学者をより妥当な推論へと導くであろう。

（薬剤疫学 2025 ; 30(2) : 53-73）

キーワード：因果推論, 疫学, 観察研究, 標的試験エミュレーション, ランダム化比較試験

1. はじめに

医療や薬事規制に携わる意思決定者は、患者の健康状態を改善する治療と、悪影響をもたらす可能性のある治療を適切に見極めるために、確かな因果推論を必要としている。この目的において、ランダム化比較試験（RCT）が標準的なアプローチとして位置づけられてきた。理想的な RCT では、明確に定義された因果的な問いに基づき、被験者をランダムに治療戦略へ割り付け、完全な治

療アドヒアランスと追跡を実現することで、因果効果の精確な推定を可能にする。しかしながら、倫理的配慮、高額な費用、運営上の障壁、あるいは長期追跡の必要性といった諸事情から、RCT の実施が非現実的な状況は少なくない¹⁾。その結果、多くの意思決定が実験的エビデンスを欠く中で行われざるをえない。RCT の枠組みでは対応しきれない複雑な臨床疑問や、迅速な判断を要する公衆衛生上の課題に取り組むために、観察研究の重要

性は近年ますます高まっている。

一方で、観察研究はランダム化の欠如による交絡に加え、不適切な研究デザインがもたらす選択バイアスや不死時間 (immortal time) など、効果推定を容易に歪めかねない脆弱性を内包している²⁾³⁾。このため、観察研究は伝統的に因果推論のための最適手段とは見なされてこなかった。こうした課題に対処するための有効なフレームワークとして、「標的試験エミュレーション (target trial emulation)」が近年注目を集めている⁴⁾。これは、関心のある因果的な問いに答えうる仮想のプラグマティック RCT (標的試験) を、利用可能な観察データを用いて明示的に模倣 (エミュレート) するアプローチを指す。このパラダイムを導入することで観察研究における因果推論を強化しようとする取り組みは、国際的な研究コミュニティにおいて急速に普及しつつある⁵⁾。

本稿ではまず第2章において、標的試験エミュレーションの概念的基盤を整理する。続く第3章では、このアプローチの方法論的側面と実装上の考慮事項を詳細に検討する。最後に第4章では、標的試験エミュレーションの適用事例を紹介する⁶⁾。本稿を通じて、標的試験エミュレーションの最大の貢献が、因果的な問いに内在する曖昧さを大幅に解消し、観察研究を介した効果推定の妥当性を高める点にあることを論証する。

2. 標的試験エミュレーションの概念的基盤

標的試験エミュレーションは、因果推論における反事実理論 (counterfactual theory) に立脚している。反事実理論における因果効果は、同一集団が異なる介入を受けた場合の潜在アウトカム (potential outcomes) の比較として定義される。つまり、ある集団がある介入を受けた場合の結果と、同一集団が別の介入を受けた場合の結果の差として因果効果が概念化される。適切に設計・実施された RCT においては、ランダム割り付けによって介入群と対照群の間の交換可能性 (exchangeability) が保証され、観察された群間差を因果効果として解釈することが正当化される。

反事実理論に基づく因果効果の概念化と推論は

観察研究の文脈にも拡張される。観察研究では対象者に対する実験的な介入を伴わないため、ランダム割り付けが保証する比較群間の交換可能性を一般に期待することはできない。さらに、実験的環境で観察が行われるわけではないことから、必要な情報が必要な精度で測定されるときも限らない。それにもかかわらず、標的試験エミュレーションは観察研究の目的を実験研究 (人を対象とする場合には臨床試験) の目的と揃えることを原理とし、また後述するようにデータの扱い方も実験研究に整合させることで、因果推論を目的とした実験研究での結果を疑似的に再現することを目指している。このような標的試験エミュレーション、あるいはそれと理論的に等価な概念は、20世紀半ばの統計学や計量経済学に起源を持ち、その後の因果推論の発展において重要な役割を果たしてきた⁷⁾。1986年、反事実理論は James Robins によって持続的治療戦略 (sustained treatment strategies) に拡張される形で一般的な定式化がなされ、現在では “g-methods” として知られる統計推測理論の基本が完成した⁸⁾。標的試験エミュレーションはこれらの理論を骨格としながら、Robins のまとめた複雑な数学理論を巧妙に回避する形で、2016年に Miguel Hernán と Robins によって観察研究のための体系的なフレームワークとして提唱された⁴⁾。

標的試験エミュレーションは、2段階のプロセスから構成される⁴⁾⁷⁾。第1段階では、関心のある因果的な問いに答えるための標的試験プロトコルを設計する。続く第2段階では、このプロトコルを利用可能な観察データを用いて明示的にエミュレートする形で研究を実施する。

第1段階で設計する標的試験プロトコルは、以下の3つのセクションから構成される (表1)⁷⁾⁹⁾：

(1) 因果的な推定対象 (causal estimands)

ここでは、次の6つの要素を定義することによって、因果的な推定対象を明確化する：

- (i) 適格基準、(ii) 治療戦略、(iii) 治療割り付け、
- (iv) アウトカム、(v) 追跡開始と終了、(vi) 因果的な対比。

観察研究に対する批判は、しばしば交絡やその他のバイアスに集中しがちである

表 1 標的試験の特定とエミュレーション

	標的試験の特定 ^a	標的試験のエミュレーション ^b
因果的な 推定対象	適格基準 各基準を記述する。	適格基準 観察データにおいて各基準をどのように定義するかを記述する。
	治療戦略 各治療戦略を記述する。	治療戦略 観察データにおいて各治療戦略をどのように定義するかを記述する。
	治療割り付け 適格者が非盲検下で治療戦略にランダム割り付けされることを明示する。	治療割り付け 観察データにおいてランダム割り付けをどのように仮想的に達成するか、そのアプローチを記述する。
	アウトカム 各アウトカムを記述する。	アウトカム 観察データにおいて各アウトカムをどのように定義するかを記述する。
	追跡開始と終了 追跡が治療割り付け時点で開始されることを明示するとともに、追跡終了の条件を記述する。	追跡開始と終了 追跡が治療割り付け時点で開始されることを明示するとともに、観察データにおいて追跡終了の条件をどのように定義するかを記述する。
	因果的な対比 関心のある因果的な対比（例えば、intention-to-treat 効果や per-protocol 効果）を効果指標とともに記述する。	因果的な対比 関心のある因果的な対比の観察研究におけるアナロジーを効果指標とともに記述する。
識別仮定	各々の因果的な推定対象の識別仮定を記述する。これらの仮定に関与する変数があれば、それについても記述する。	ベースライン交絡の調整に関する仮定を含め、各々の因果的な推定対象の識別仮定を記述する。これらの仮定に関与する変数があれば、観察データにおいてそれをどのように定義するかも記述する。
推定量	各々の因果的な推定対象に対する解析手法について、そのモデリング仮定とともに記述する。	各々の因果的な推定対象に対する解析手法について、そのモデリング仮定とともに記述する。各種エミュレーションおよび識別仮定やモデリング仮定について、潜在的な逸脱がある場合に実施する、結果の頑健性を評価するための感度解析についても記述する。

^a：因果的な問いに答える標的試験のプロトコルの要素を特定する。

^b：観察データから標的試験のプロトコルをどのようにエミュレートするかを記述する。ここには、すべての変数をどのように測定または確認（直接測定できない場合）するかも含む。

が、それとは別次元の根本的な問題として、そもそも何を推定対象とするのかという問い自体の曖昧さを看過してはならない¹⁰⁾¹¹⁾。問いが不明瞭なままでは、得られた推定値の解釈が困難となり、意思決定への活用を正当化できなくなる。標的試験プロトコルのフレームワークは、推定対象の要素を一つひとつ意識的に設定することを通じて、反事実アプローチの数学的記法や概念に精通していない研究者に対しても、曖昧さのない因果的な問

いを策定するよう導く体系的な方法論といえる。

(2)識別仮定 (identifying assumptions)

因果効果を識別する（観察データの確率分布で反事実的な量を表現する）ために必要な仮定を整理し、観察データでの蓋然性を評価する。単回の治療または介入に対しては、条件付き交換可能性 (conditional exchangeability)、因果一致性 (causal consistency)、条件付き正值性 (conditional positivity) の3つが

代表的な仮定のセットとして挙げられる¹²⁾。ランダム化が行われていない観察研究においては特に条件付き交換可能性の担保が主要な課題となるが、観察研究では原理的に介入プロトコル自体が設定されていないことから、一致性と正值性についても介入研究に比べて蓋然性は低下する。持続・継続される治療(sustained treatments)や時間とともに変更される治療(time-varying treatments)に対しては、治療時点ごとの逐次的な交換可能性(sequential exchangeability)とそれに対応した一貫性・正值性を要する。

(3)推定量(estimators)

識別仮定および追加のモデリング仮定のもとで、標的試験が実施された場合に得られうる因果的な推定対象の推定値(estimates)を導出するための解析手法(データを代入する公式や計算アルゴリズム)を提示する。

標的試験エミュレーションの第2段階では、第1段階で設計したプロトコルの3つのセクションを、実際の観察データを用いて明示的にエミュレートする。標的試験は「理想としてのRCT」であると同時に、「利用可能な観察データを用いて合理的にエミュレート可能な試験」でなければならないため、第1段階のプロトコル設計と第2段階のエミュレーションは相互に行き来する反復的なプロセスを経て洗練されていく¹³⁾。この過程を通じて、「どこまで理想的なRCTに近づくのか」と「どこまで観察データに起因する制約を許容するか」の間にある妥協点が明確化される。こうした構造化されたアプローチこそが標的試験エミュレーションの核心をなし、RCTが本来備える望ましい特長—すなわち因果的な問いを厳密に定義したうえで効果推定を行う枠組み—を観察研究においても最大限維持するための方法論的基盤となるのである。なお、観察データを前提とする以上、プラセボ対照や盲検化、現実には存在しない介入を設定することは不可能である。また、患者のモニタリング頻度も、利用可能なデータの粒度に合わせる必要がある。

3. 標的試験エミュレーションの方法論的側面と実装上の考慮事項

第2章で示した概念的基盤を踏まえ、本章では標的試験プロトコルをエミュレートする際に直面する課題と考慮事項について論じる。標的試験エミュレーションの実践においては、理想と現実のバランスを取りながら、プロトコルの各構成要素を利用可能な観察データにいかに忠実に反映させるかが鍵となる。

エミュレーションは、標的試験プロトコルの各要素—適格基準、治療戦略、治療割り付け、アウトカム、追跡開始と終了、因果的な対比—を、利用可能な観察データでどのように定義するか検討することから始まる⁷⁾。このプロセスの質が、エミュレーション全体の妥当性を左右する。すなわち、これらの要素を定義するデータ項目が、仮に標的試験を実施した場合に生成されうるデータをどれだけ近似しているかが、エミュレーションの成否を決定づける。このプロセスを適切に実施するには、研究に用いる観察データの構造的特性、収集方法、および限界についての深い理解が求められる。表1のエミュレーションの欄には、因果的な推定対象の各要素を観察データからどのように定義するかを明示するとともに、一部の要素がエミュレート不可能または不完全である場合には、その逸脱内容を記述することが求められる⁷⁾⁹⁾。このようなプロトコル要素の逸脱に対しては、感度解析を行い、因果効果の推定に与える影響を多角的に検討することが推奨される¹³⁾。感度解析は標的試験プロトコルの一種のバリエーションと見なすことができ、研究結果の頑健性を評価するうえで不可欠である。なお、標的試験プロトコルに含まれる識別仮定は、治療戦略や治療割り付け、解析手法に応じて異なりうるため、これらのエミュレーションに関する議論全般にわたって言及されることを申し添えておく。標的試験エミュレーションの各要素は独立したものではなく、相互に関連し影響し合う有機的な全体を形成していることを理解することが、このアプローチの本質を把握するうえで重要である。

3.1 適格基準のエミュレーション

適格基準のエミュレーションは、標的試験プロトコルで規定された選択・除外基準を観察データに適用する過程である。この段階では、各種基準を十分な精度で反映できる観察データを、その特性と限界を綿密に検討したうえで選択することが肝要である。適格基準の適切な設計と適用は、エミュレーションの内的妥当性と外的妥当性の均衡を保持するための礎石となる。

レセプトデータや電子カルテデータなどの二次利用データを用いる場合、診断コードの妥当性や臨床情報の完全性といったデータの質が適格性判定の精度に直接影響する。特定の基準に関するバリデーション研究が存在する場合は¹⁴⁾、その知見を参照することで判定精度を向上させることができる。例えば、特定の疾患の診断コードに関するバリデーション研究から得られる感度、特異度、陽性的中率、陰性的中率の情報は、適格基準の構築において貴重な指針となる。しかしながら、実際の観察データにおいては、標的試験の適格基準を完全にエミュレートできないことも少なくない。例えば、禁忌や重篤な併存疾患に関する除外基準は、既存の観察データで利用可能な変数のみでは完全に表現できない場合がある。このような状況では、利用可能なデータを最大限に活用しつつも、基準の不完全な適用が最終的な結果に与える影響を評価する必要がある。さらに、研究によっては適格性判定に特定の検査値や詳細な臨床評価が必要となるが、これらの測定値が観察データに含まれていない、あるいは選択的に欠測しているケースも珍しくない。こうしたデータ上の制約に直面した場合、研究者は以下のような選択肢から適切なアプローチを選択しなければならない：検査値や臨床所見以外の代替基準を用いる、外部データとのリンケージによって不足情報を補完する¹³⁾、あるいは追加の情報収集のための資源を投入する。

適格基準の設定は、得られた結果の一般化可能性 (generalizability) や移送可能性 (transportability) の概念にも密接に関連する¹⁵⁾。一般化可能性は研究対象集団が標的集団の厳密な部分集合であ

る場合に関係し、移送可能性は研究対象集団と標的集団が部分的に重なるまたは完全に異なる場合に関係する。いずれの場合も、効果修飾因子の分布が研究対象集団と標的集団で異なっていると、推定された効果は標的集団での効果としてはバイアスを含む推定値となってしまう。したがって、適格基準を設計する際には、研究対象集団と最終的な標的集団の効果修飾因子の分布の類似性を考慮するか、効果修飾因子のデータを用いて統計的に対処することが重要である。

観察研究では往々にして、最適とはいえない選択肢の中から選択・除外の判断を下さざるをえない状況に直面する。適格基準を厳密に設定して研究対象集団を限定すれば内的妥当性は向上するものの、その代償として、除外された集団に対する結果の外的妥当性は低下する恐れがある。しかし、薬剤疫学研究が医療や薬事規制の意思決定に与える影響の重大性を考慮すれば、たとえ外的妥当性のある程度犠牲にしても、まずは内的妥当性を確保することが望ましいであろう¹⁶⁾。研究者はこうしたトレードオフを十分に認識したうえで、最終的な選択とその根拠を明示する必要がある。

3.2 治療戦略のエミュレーション

治療戦略のエミュレーションの成否は、因果効果の識別仮定の1つである（因果）一致性の成立に依拠している。一致性は2つの要素から構成される：(i)十分に明確化された介入 (sufficiently well-defined interventions) と (ii)介入とデータのリンケージ (linkage between interventions and the data) である¹⁷⁾。第一の構成要素「十分に明確化された介入」とは、その介入を受けた場合と受けなかった場合それぞれの潜在アウトカムが一意に定義されるよう、介入が詳細に指定されていることを意味する。例えば「スタチンの効果」という表現を考えてみよう。この表現は曖昧である——スタチンの種類、用法、用量、使用期間によって治療戦略に複数のバージョン (multiple versions of treatment) が存在し、それぞれが異なる「スタチンを投与された場合に観察されるであろうアウトカム」と対応することで、求めたい

効果が一意に定まらないからである。治療戦略に複数のバージョンが存在する状況では、それらを単一の治療戦略として統合するか、あるいは区別して扱うかという判断を求められる。バージョン間の相違が定義される効果に実質的な違いを与えないという仮定 (treatment-variation irrelevance assumption) が成立する場合に限り¹⁸⁾、複数のバージョンを同一の治療戦略として扱うことが正当化される。最終的に、何が「十分に明確化された介入」であるかという判断は、当該領域の専門家間の合意と科学的知見に基づくものとなる。

一致性の第二の構成要素「介入とデータのリンケージ」は、第一の構成要素「十分に明確化された介入」が実際の観察データに存在することを要求する。逆に言えば、「十分に明確化された介入」が定義できたとしても、その介入自体が観察データに実在しない場合、当然ではあるが、推定された効果の解釈は困難になる。例えば、「5年間にわたり1日10mgのプラバスタチンを2回に分けて経口投与する効果」を推定したいという状況を想定しよう。しかし、解析対象のデータセットに「5年間にわたり1日10mgのプラバスタチンを1回で経口投与」した記録しか存在しない場合、得られた推定値を関心のある介入の効果として解釈することは不可能である。

治療戦略のエミュレーションにおいては、さらに治療の時間的特性に着目し、時間固定治療 (time-fixed treatments) と時間依存性治療 (time-varying treatments) の2つに大別して考える必要がある。時間固定治療は一時的介入 (point interventions) とも称され、単回のワクチン接種や外科的処置のように、研究期間内の単一時点で実施される介入を指す。対照的に、時間依存性治療は持続的治療戦略 (上述) とも呼ばれ、薬剤の定期的服用など、時間経過に伴う複数回の連続的介入からなる。時間依存性治療はさらに、静的治療戦略 (static treatment strategies) と動的治療戦略 (dynamic treatment strategies) に分類される。静的治療戦略は、治療開始後の個人の臨床経過にかかわらず、対象集団全体に対して一律に適用される治療方針を指す (例えば、治療継続、治

療非実施、一定期間のみ治療実施など)。しかし、このような治療戦略は「治療開始以降の副作用発現後も治療を継続する」といった非現実的な設定を含むため、必ずしも臨床的な関心を集めるとは限らない。一方、動的治療戦略は、治療開始後に時間とともに変化する個人の特性や臨床経過に応じて治療内容を変更する治療方針またはルールであり、臨床実践により即した設定といえる。例えば、「副作用発現時には治療中止、それ以外は継続」という治療戦略では、副作用のために治療を中止した個人は、この治療戦略からの逸脱とは見なされないことに留意すべきである¹⁹⁾。

静的・動的を問わず、時間依存性治療のエミュレーションは時間固定治療と比較してより複雑である。この複雑性は、時間経過に伴う治療へのアドヒアランスの変動に対処するために、時間依存性交絡変数 (time-varying confounders) の調整が不可欠になることに起因する。特に重要なのは、過去の治療が現在の交絡変数に影響を及ぼし、その交絡変数がさらに未来の治療選択にも影響するという、治療・交絡変数フィードバック (treatment-confounder feedback) の存在である。このように複雑な因果構造を適切に扱うには、識別仮定を拡張し (治療時点ごとの「逐次」交換可能性および対応する一致性・正值性)、“g-methods”と総称される解析手法が必要になる (第3.7節参照)。もちろん、治療アドヒアランスに関連する時間依存性交絡変数に関する十分に詳細な情報が観察データから得られない場合、g-methodsを用いたとしても時間依存性治療を含む標的試験の妥当なエミュレーションは困難になる。また、時間依存性治療におけるアドヒアランス (開始、継続、中止を含む²⁰⁾) の操作的定義にも不確実性が伴う。レセプトデータや電子カルテデータから得られる処方・調剤記録は比較的安全性が高いものの、患者の実際の服薬行動との対応関係を検証することはできない。このような不確実性に対処するために、治療中止を定義する猶予期間 (grace periods) を変更するなど、複数の感度解析を実施することが推奨される²¹⁾²²⁾。

3.3 治療割り付けのエミュレーション

RCTと観察研究のデータを本質的に区別するものは主に2つである。RCTでは(i)ランダム化によりベースライン交絡が平均的には存在しないことを理論的に期待できること、(ii)治療割り付け確率が既知であることである。しかし、観察研究において条件付き交換可能性または逐次交換可能性を仮定することは(i)十分な共変量セット(sufficient set of covariates)を条件付けた中で、実質的に(潜在アウトカムとは独立に)ランダムに治療が決まっていると宣言することにほかならず、また(ii)その共変量セットごとの治療割り付け確率は正確に特定されたモデルから推定することができる。この2点においてそれぞれ仮定(条件付き/逐次交換可能性と治療確率モデルの正しい特定)が必要になるものの、観察研究において標的試験のランダム化の要素をエミュレートすることが可能になる¹²⁾。

十分な共変量セットを特定するには、研究対象に関する専門的知見(subject-matter knowledge)が不可欠である。因果有向非巡回グラフ(causal directed acyclic graph, causal DAG)は、変数間の因果関係に関する知見や仮定を矢線によって可視化するツールであり²³⁾⁻²⁵⁾、因果効果の推定において調整すべき変数と調整を避けるべき変数を体系的に検討することを可能にする。理論的には、バックドア基準(backdoor criterion)は、因果DAGにおける治療戦略間の交換可能性を成立させるのに十分な共変量セットを特定するためのゴールドスタンダードとされる。しかし実践においては、関連するすべての変数間の因果関係を網羅した完全な因果DAGを構築することは容易ではなく、バックドア基準の適用に限界がある場合がある。この理論と実践のギャップを埋めるために提案された共変量選択基準が、modified disjunctive cause criterionである²⁶⁾²⁷⁾。この基準では、治療に先行して測定され、治療、アウトカム、またはその両方に影響を与える任意の共変量(disjunctive causes)を調整対象とする。ただし、治療にのみ関連するような操作変数(instrumental variables)は調整から除外し、治療とアウトカ

ムの共通原因との相関の大きな代理変数(proxy variables)は含めるという「修正」された指針を与える。このアプローチは、詳細な因果DAGの構築が困難な状況においても、十分な共変量セットを確からしく選定できる可能性の高い実用的な戦略として評価されている。

このように選択した共変量セットを用いて、共変量の組み合わせごとの治療確率やアウトカムの平均をモデル化することで標的試験での群間比較をエミュレートする。サンプルサイズが無限大であれば、モデルに関する仮定を最小限に治療割り付け確率を推定できるが、実際のサンプルサイズは常に有限であるため、バイアス・分散トレードオフ(bias-variance tradeoff)を考慮する必要がある。複雑なモデルは効果推定量のバイアスを減少させる可能性がある一方で、分散を増加させることも少なくない。また、高次項や交差項などを含む柔軟なモデルの推定値は、スパースデータバイアス(統計的には有限標本バイアスの一種)の影響を受けるかもしれない。したがって、サンプルサイズに比して共変量の数が十分少ない場合以外は、modified disjunctive cause criterionを適用した後に、治療確率モデルまたはアウトカム回帰モデルに含む項を制限したり、共変量の数自体を統計的観点から削減したりする必要があるかもしれない。そのような状況における共変量選択の汎用的な統計的アプローチはいまだ確立されていないが、論点は引用文献を参照されたい²⁶⁾²⁷⁾。

観察研究における交絡調整には、層別、回帰モデリング、共変量マッチング、傾向スコアモデリング、g-methods(第3.7節参照)などさまざまな手法が利用可能である。これらの手法はいずれも、同一の識別仮定(条件付き/逐次交換可能性、一致性、正值性)および(層別以外は)相異なるモデリング仮定に依拠している(表2)。識別仮定を直接検証することは不可能であるが、モデルの誤特定がない(no model misspecification)という仮定については、モデル診断やモデル比較を通じて部分的に評価することが可能である。交絡調整手法の選択によって異なる推定対象が定義されることがあるため、研究者は手法を選択する前に関

表 2 主な交絡調整手法

交絡調整手法	時間固定治療/ 時間依存性治療	識別仮定	モデリング仮定	効果ごとの推定方法の例	
				条件付き効果 ^a	周辺効果 ^b
層別 (サブグループ解析, Mantel-Haenszel 推定量, 標準化)	時間固定治療	- 条件付き交換可能性 - 因果一致性 - 条件付き正値性	- 不要 (サブグループ解析および標準化) - 効果修飾の不存在 (Mantel-Haenszel 推定量)	- 層ごとに個別に効果指標を計算 (サブグループ解析) - 層間で効果指標を統合 (Mantel-Haenszel 推定量)	- 標準化 (リスク差・比の指標に限っては効果修飾が存在しない仮定下で, 条件付き効果推定量で代替可)
回帰モデリング (回帰係数の推定, 回帰モデルに基づく標準化)	時間固定治療	上に同じ	- 治療・交絡変数を与えたアウトカムの条件付き期待値を正しい関数形で特定 - 特に効果修飾の不存在または制約	回帰係数 (効果修飾の不存在を仮定したモデル) または回帰係数の組み合わせ (交差項などで制約された効果修飾を許容したモデル) の推定値	治療を特定の値に置き換えた回帰モデルの予測値を全員で平均をとることで, 交絡変数の経験分布で標準化 (時間固定治療に対する parametric g-formula に等しい)
共変量マッチング (正確マッチング, カテゴリ化された正確マッチング, マハラノビス距離マッチングなど)	時間固定治療	上に同じ	正確にマッチできる場合は不要 (マッチされる変数の粒度が荒い場合, マッチ後の解析で用いる手法に応じて, 回帰モデルや Mantel-Haenszel 推定量と同様の仮定が必要)	- 交絡変数を調整した回帰モデルの回帰係数推定値 - マッチされた層を固定効果とした回帰モデル (例えば, 条件付きロジスティックモデル) または Mantel-Haenszel 推定量	- マッチされた集団での単純比較 - 層別/回帰モデルと同様の周辺効果推定値 (標的集団はマッチされた集団)
傾向スコアモデリング (傾向スコア層別・回帰調整, 傾向スコアマッチング, 逆確率重み付け)	時間固定治療	上に同じ	- 交絡変数を与えた条件付き治療確率 (傾向スコア) を正しい関数形で特定 (傾向スコアを回帰モデルで調整する場合) 傾向スコアと治療を与えたアウトカムの条件付き期待値の正しい特定	- 層ごとの効果指標推定値またはそれらの統合 (傾向スコアの層別)^c - 回帰係数推定値 (回帰モデルによる傾向スコアの調整)^c - 傾向スコアマッチ集団における交絡変数を調整した回帰モデルの回帰係数推定値 - 交絡変数を調整した回帰モデルの逆確率重み付け推定値	- 標準化 (傾向スコアの層別または回帰モデル調整) - マッチされた集団での単純比較^d - 逆確率重み付け後の単純比較
Parametric g-formula	時間依存性治療	- 各時点の治療に対する逐次条件付き交換可能性 - 治療戦略の因果一致性 - 各時点の治療に対する条件付き正値性	- それまでの治療・交絡変数履歴を与えたアウトカム (イベント型の場合は各時点のアウトカム有無) の条件付き期待値の正しい特定 - それまでの治療・交絡変数履歴を与えた各時点の交絡変数分布の正しい特定	Parametric g-formula 推定値 (モデル推定値) に基づいて, 特定の治療戦略開始後の交絡変数分布のモンテカルロ・シミュレーション近似と, シミュレーション値に基づくアウトカム予測値を用いた標準化)	

表 2 (つづき)

効果ごとの推定方法の例				
交絡調整手法	時間固定治療/ 時間依存性治療	識別仮定	モデリング仮定	効果ごとの推定方法の例
				条件付き効果 ^a 周辺効果 ^b
逆確率重み付け (必要に応じて周辺構造モデル)	時間依存性治療	上に同じ	• それまでの治療・交絡変数履歴を与えた各時点の治療確率 (時間依存性傾向スコア) の正しい特定 • (必要に応じて) 周辺構造モデルによる治療効果または治療戦略効果の正しい制約	ベースライン交絡変数を条件付けた周辺構造モデルパラメータ推定 (各時点の傾向スコア推定値に基づく逆確率重みをアウトカム評価時点まで累積して解析に用いる逆確率重みを作成し、周辺構造モデルと同じ形状の帰帰モデルを逆確率重み付け推定) • ベースライン交絡変数を条件付けない周辺構造モデルのパラメータ推定 (周辺構造モデルと同じ形状の帰帰モデルを逆確率重み付け推定)
構造ネストモデルの g-estimation	時間依存性治療	条件付き正値性を除いて上に同じ	• 構造ネストモデルの正しい特定 • それまでの治療・交絡変数履歴を与えた各時点の治療確率 (時間依存性傾向スコア) またはそれまでの治療・交絡変数履歴を与えた (以降の治療を除いた介入下での) 潜在アウトカムの条件付き期待値の正しい特定 (二重ロバスト性) • (周辺効果を求める場合) それまでの治療・交絡変数履歴を与えた各時点の交絡変数分布の正しい特定	G-estimation による構造ネストモデルのパラメータ推定値を得た後 (blip-down ステップ), parametric g-formula 同様に各モデル推定値から治療戦略開始後の交絡変数およびアウトカム分布をモンテカルロ・シミュレーションにより推定し (blip-up ステップ) 標準化

^a : 手法によって条件付けける変数が異なることによって別の条件付き効果を表すことがある (特に断らない限りはすべてのベースライン交絡変数による条件付け)

^b : 本表では特に断らない限り解析対象集団全体を標的集団とした周辺効果を著すが、各手法とも標的集団を変えることができる (治療群、非治療群、効果の一般化および移送)

^c : 交絡変数ではなく傾向スコアで条件付けた効果

^d : マッチされた集団を標的集団とした周辺効果

心のある因果効果を明確にする必要がある。ここで、条件付き効果 (conditional effect) と周辺効果 (marginal effect) の区別が重要となる。条件付き効果は特定の共変量レベルにおける因果効果を指し、特定の患者サブグループに対する治療効果の理解に有用である。一方、周辺効果は対象集団全体における平均因果効果に関心がある場合に適している。ただし、これらは相対的な定義であるため、条件付きか周辺かの名称よりも「どの集団における、どのような仮定を含む効果か」を明確にすべきである。

他に注意すべきこととしては、標的試験エミュレーションを適用したとしても、交絡を調整するための情報が欠如しているといった、データ自体の限界から生じるバイアスを除去することはできないことが挙げられる。そのため、バイアス解析 (bias analysis) の実施が重要となる。治療割り付けのエミュレーションの質を評価するアプローチの1つとして、ネガティブコントロール (negative controls) の活用が挙げられる²⁸⁾。ネガティブコントロールには主に2種類ある。1つ目のネガティブコントロール「治療」はアウトカムと因果関係を持たないが、未測定交絡変数が本来の治療と類似している変数である。もう1つのネガティブコントロール「アウトカム」は治療と因果関係を持たないが、未測定交絡変数が本来のアウトカムと類似している変数である。ネガティブコントロール解析において見かけ上の関連が検出された場合、未測定交絡によりランダム化をエミュレートできていない可能性が示唆される。しかし、同時に未測定交絡によるバイアスの程度を推定することもできるため、得られた因果効果の推定値に対する一種のストレステストとして見ることもできる。

3.4 アウトカムのエミュレーション

観察データ、特にレセプトデータや電子カルテデータ等からアウトカムをエミュレートする際、研究者はアウトカム同定のためのアルゴリズムを駆使する。こうしたアルゴリズムは、臨床ワークフローの知見と利用可能なデータに基づき、診断、診療行為、医薬品処方、臨床検査結果、医療

提供環境などから1つ以上の要素を組み合わせて構築される²⁹⁾。しかし、このようなアルゴリズムの妥当性は、患者の受診行動パターン、医療アクセスの格差、医療の質的相違、鑑別診断の複雑性、あるいは除外診断のための暫定的コーディングなど、諸要因に影響を受ける³⁰⁾。これらの要因が複雑に交錯するため、アルゴリズムが関心のあるアウトカムを正確に捕捉できる保証はない。したがって、バリデーション研究を通じてアルゴリズムの診断性能 (感度、特異度、陽性的中率、陰性的中率など) を適切に評価することが重要となる。

これらの評価の重点をどこに置くかは、推定したい効果指標のスケールに依存することがある。特に、アルゴリズムが関心のあるアウトカムの絶対リスクやリスク差の推定に使用される場合、低感度のアルゴリズムはこれらの推定値にバイアスをもたらす。アウトカムが治療関連有害事象である場合、不適切な安全性保証や誤った警告を避けるために、高感度のアルゴリズムが概して望ましい。リスク比に関しては、理論的には、アウトカム測定の特異度が完全で、感度が比較群間で非差異的であり、さらに曝露の測定誤差が他の誤差と独立している場合に限り、推定値にバイアスが生じない³¹⁾³²⁾。この理論的背景から、リスク比の推定においては、特異度を優先すべきという見解が存在する。しかし、実際の観察研究においてこれらの条件が完全に満たされることは稀であり、その検証も極めて困難である。したがって、リスク比が主要な関心事であるという理由だけで、高特異度を確保する代償として感度を意図的に低下させることは推奨されない¹⁴⁾。

既に検証されたアルゴリズムであっても、その検証が行われた集団が、アルゴリズム適用対象集団の部分集団に過ぎない場合、あるいは本質的に異なる特性を持つ集団である場合には、アルゴリズムの移送可能性を慎重に吟味する必要がある²⁹⁾。さらに、基盤となる医療システム、診療ガイドライン、データ生成メカニズム等の相違により、ある特定のデータソースにおいて妥当性が確認されたアルゴリズムが、異なるデータソースにおいても同等の性能を発揮する保証はない³³⁾⁻³⁵⁾。

同様に、特定の期間において検証されたアルゴリズムが、異なる期間において同等の診断性能を維持するという仮定も、医療実践の進化や診断基準の変遷を考慮すれば、常に成立するとは限らない¹⁴⁾²⁹⁾。

3.5 追跡開始と終了のエミュレーション

標的試験エミュレーションにおいて、アウトカムの追跡ルールの適切な設定は結果の妥当性に大きな影響を与える。本節では、追跡開始時点であるタイムゼロ (time zero) の設定と追跡終了の決定に関する考慮事項を、関連する重要概念とともに整理する。

タイムゼロを設定するうえで最も重要な原則は、追跡開始時点、研究への適格性の判定および治療割り付けのタイミングと同期させることである²⁾。RCT では、被験者の適格性判定とほぼ同時に治療割り付け (ランダム化) が行われ、割り付け群間の比較可能性が保たれているランダム化時点を追跡開始時点とすることで、3つの時点を意図的に同期させる。観察研究においても同様の原則を適用し、タイムゼロを定義する必要がある。しかし、因果推論を目的とした観察研究では、タイムゼロのエミュレーションの失敗が頻発する。これらの失敗は主に次の2つの問題から生じる：(i)タイムゼロが一意に定まらないこと、(ii)タイムゼロで治療戦略を一意に割り付けできないこと。タイムゼロの不適切な設定は、選択バイアスや不死時間を誘発する²⁾³⁾。これらが生じるメカニズムと、その回避のためのアプローチである逐次試験エミュレーション (sequential trial emulation) および clone-censor-weight アプローチについては、『薬剤疫学』31巻1号に掲載予定の別稿で詳述する。

追跡終了は、関心のある因果的な対比 (第3.6節参照) に応じて異なる形で決定される。Intention-to-treat 効果を推定する場合、追跡終了は関心のあるアウトカムの発生、競合イベント (competing events) の発生、追跡不能 (loss to follow-up)、または追跡の管理的終了 (administrative end of follow-up) によって決まる。Per-protocol 効果を推定する場合には、これらに加えて個

人が割り付けられた治療戦略から逸脱した時点も追跡終了とすることがある。

特に注意を要するのが競合イベントの取り扱いである。競合イベントとは、関心のあるアウトカムの発生を不可能にするイベントを指す。例えば、心血管死をアウトカムとする研究では、他の原因による死亡は競合イベントとなる。競合イベントの取り扱いによって異なる推定対象が定義されるため、研究者は関心のある効果を明示する必要がある。多くの観察研究における推定値の妥当性に関する議論では、推定対象を十分に検討せずに解析手法の議論に終始しているが、推定対象の選択なくして適切な解析手法を決定することはできない。競合イベントの存在下における主な推定対象は次の3つに整理することができる：(i)複合アウトカムに対する効果、(ii)総合効果 (total effect)、(iii)制御された直接効果 (controlled direct effect)³⁶⁾。1つ目の複合アウトカムは、競合イベントをアウトカムの構成要素として包含するアプローチであり、データの取り扱いは単純で分かりやすい。しかし、この方法では当初の因果的な構成要素 (第3.4節) そのものを変質させ、推定値は関心のあるアウトカムへの効果よりも競合イベントの発生パターンに左右される可能性がある。2つ目の総合効果は、競合イベントをそれ以降のアウトカム発生を妨げるイベントとして解釈し、競合イベントを経験した個人はアウトカム発生なしとして管理的終了まで追跡される。この方法では、治療戦略からアウトカムに至るまでのすべての因果経路 (競合イベントを経由してアウトカム発生を「予防」する間接経路を含む) を捉えていると見ることができる。これは集団全体が異なる治療戦略を受けた場合のアウトカム分布の比較となり、明確に定義された (well-defined) 因果効果として解釈可能である。ただし、治療戦略が競合イベントを増加させる場合、アウトカムへの総合効果は保護的になりうるなど、因果効果としては明確に定義されていたとしても解釈に注意を要する場合がある。3つ目の制御された直接効果は、競合イベントをアウトカム発生ではなくアウトカム観察を妨げる「打ち切り」イベントとして解釈

する。競合イベントを打ち切りとして扱うことは、競合イベントを排除するような介入により、競合イベントが発生しない状態に制御した下での、治療戦略からアウトカムへの直接効果と見ることができる。この推定対象はすべての競合イベントを防ぐという介入を想定するため、現実には達成困難な仮想的介入に依存している可能性がある。その場合には、この制御された直接効果は明確に定義されない (ill-defined) 因果効果となる。各推定対象の臨床的意義に加えて、それぞれの妥当性を判断するには研究対象に関する深い専門的知見が不可欠である。

3.6 因果的な対比のエミュレーション

因果推論の研究では、それがRCTであれ観察研究であれ、異なる治療戦略間の因果的な対比を定量化することを目的とする。標的試験エミュレーションにおいては、intention-to-treat 効果と per-protocol 効果という2つの対比がよく取り上げられる。本節では、これらの因果的な対比の概念をRCTの文脈で整理し、観察研究においてどのようにエミュレートできるかを論じる。

RCTにおいて、intention-to-treat 効果は治療戦略への割り付け効果であり、その後の実際の治療アドヒアランスとは無関係に定義される。この効果の解釈には複数の論点が存在する。第一に、二重盲検下ではどの対象者に対しても真の治療効果が存在しなければ intention-to-treat 効果も存在しないことが期待されるが、非盲検下では治療割り付けそのものが医師や患者の行動変容を通じてアウトカムに影響を与えうするため、真の治療効果が存在することが intention-to-treat 効果が存在することに含意されない可能性がある。第二に、intention-to-treat 効果は一般に per-protocol 効果よりも保守的 (効果が小さい) と考えられているが、これは必ずしも正確ではない。治療に非単調効果が存在する場合 (つまり一部の患者には有益だが他の患者には有害である場合) intention-to-treat 効果は per-protocol 効果より大きくなったり逆方向になったりすることがある¹²⁾。結果として、介入を伴う2つの治療戦略を比較するオープンラベルの試験では、各群におけるアドヒアラン

スパターンの違いにより、intention-to-treat 効果は治療戦略に従った場合の per-protocol 効果を過大または過小評価する可能性がある。例えば、治療戦略Aと治療戦略Bの比較において、副作用のために治療戦略Bへのアドヒアランスが低い場合、intention-to-treat 効果は治療戦略Aに従った場合の効果を過大評価するだろう。逆に、治療戦略Aが安全性のアウトカムではない有害事象を引き起こし、被験者の早期脱落を招く場合、intention-to-treat 効果は治療戦略Aに従った場合の安全性アウトカムを過小評価しかねない。第三に、intention-to-treat 効果は不完全な治療アドヒアランスを組み込んでいるため、実臨床での治療効果を反映すると主張されるが、その根拠は弱い。特定のRCTで観察されたアドヒアランスパターンは、他のRCTや臨床現場のそれとは異なる可能性が高い¹⁹⁾。以上の理由から、RCTであっても特に安全性試験や非劣性試験において intention-to-treat 効果のみに依拠することは、リスクを過小評価したり臨床的に重要な差異を覆い隠したりする可能性があるため、正当化しにくい³⁷⁾。一方で、per-protocol 効果は、割り付けられた治療戦略を遵守した場合の効果を意味し、治療戦略の実際の効果を知りたい医師や患者にとって、より直接的な関心の対象となる。しかしながら、治療・交絡変数フィードバックの存在下において per-protocol 効果をバイアスなく精確に推定するには、RCTであっても治療アドヒアランスを予測する共変量のデータを収集し、g-methodsを原理にもつ適切な解析手法が必要となる (第3.7節参照)³⁸⁾。

観察研究では特定の治療戦略へ実際に割り付けているわけではないため、intention-to-treat 効果をエミュレートできることは稀である。代わりに、intention-to-treat 効果の観察研究におけるアナロジー (observational analog) として、「異なる治療戦略を開始すること」の効果が定義される¹²⁾。これはベースライン後の治療アドヒアランスパターンを問わない点で intention-to-treat 効果の概念的特性を継承している。また、per-protocol 効果の観察研究におけるアナロジーは、開始され

た治療戦略を遵守した場合の効果として定義される。観察研究における intention-to-treat 効果の推定には治療開始とアウトカムに影響を与える共変量の調整が、それに加えて per-protocol 効果の推定には RCT と同様に、治療アドヒアランスに影響を与える共変量の適切な測定と調整を必要とする。

因果効果の解釈においては、理想的な実験条件下の効果 (efficacy) と現実的条件下での効果 (effectiveness) という二分法がしばしば用いられるが、この分類は一貫性を欠くことが多い¹²⁾。同一の RCT 内で intention-to-treat 効果を effectiveness, per-protocol 効果を efficacy と呼ぶ場合がある一方で、不完全な治療アドヒアランスを伴う RCT の intention-to-treat 効果を efficacy, アドヒアランスを予測する共変量を適切に調整した観察研究の per-protocol 効果を effectiveness とする解釈も存在する。特に後者においては、RCT の intention-to-treat 効果と観察研究の per-protocol 効果のどちらがより efficacy と呼ぶにふさわしいかを判断するのは多くの条件が異なるために困難である。このような efficacy/effectiveness という単純化された用語の二分法よりも、すべての因果効果が何らかの制約を課された現実的条件下での effectiveness の連続体上に位置づけられると理解し、具体的な治療戦略とともに因果効果を明示的に定義の方が意思決定者にとって有用である³⁸⁾。

3.7 解析手法のエミュレーション

標的試験エミュレーションにおいて、適切な解析手法の選択は効果推定の妥当性を左右する。本節では、まず RCT における intention-to-treat 解析と per-protocol 解析を振り返り、第 3.6 節で見た intention-to-treat 効果および per-protocol 効果との関係性を整理する。次に観察研究においても、ベースライン交絡の調整という差異を除けば、これらの効果を求めるための解析方針を RCT と本質的に同じように実装できることを説明する。最後に「アウトカム発生までの時間に対する治療効果」を対象とした因果的な問いに対して、従来の生存時間解析に内在する課題とその代替となるアプローチについても考察する。

RCT において、実際の治療アドヒアランスにかかわらず、ランダム割り付けされた治療戦略に基づいてすべての対象者を含めて比較することを intention-to-treat 解析と呼ぶ。この方法は、すべての対象者を追跡できる状況では、intention-to-treat 効果に対してバイアスのない推定値を提供する。ただし、追跡不能により解析対象から脱落する個人が生じ、この追跡不能が社会的・臨床的要因（ランダム化後まで含めたアウトカムの予測変数）によって非ランダムに、かつ割り付け群間で異なるパターンで発生する場合、打ち切られた個人と追跡が継続された個人との間でアウトカムの交換可能性が崩れることから選択バイアスが発生しうる。RCT の intention-to-treat 解析においても、このバイアスに対処しながら intention-to-treat 効果を求めるためには、追跡不能に影響を与えるランダム化前とランダム化後の共変量を、逆確率重み付け (inverse probability weighting) によって適切に調整する必要がある¹²⁾。

RCT における per-protocol 解析は、割り付けられた治療戦略を遵守した per-protocol 集団のみに限定した解析を指す。このような per-protocol 解析 (“naïve” per-protocol 解析) からは妥当な per-protocol 効果の推定値を得られないことがあり、その原因は 2 つの潜在的なバイアスに分類できる¹⁹⁾。第一に、非ランダムな追跡不能による選択バイアスである。上記のような “naïve” per-protocol 解析は、実のところ per-protocol 集団に限定した intention-to-treat 解析と見なせる。したがって、intention-to-treat 解析が必ずしも intention-to-treat 効果をバイアスなく求められないのと同様に、追跡不能に影響を与えるランダム化前後の共変量は適切な方法で調整される必要がある。第二に、治療戦略への不完全なアドヒアランスによる交絡である。アドヒアランスも追跡不能と同様に、ランダム化後の諸要因に影響される可能性があり、さらにこれらの要因は治療の影響を受けうることから、ランダム化後の交絡 (post-randomization confounding) と呼ばれる現象が生じる。このような不完全なアドヒアランスに対処して per-protocol 効果を求めるためには、まず個人が

治療戦略から逸脱した時点で追跡を打ち切り（動的治疗戦略では、副作用などの臨床的理由で治療を中止した場合は逸脱と見なさず、追跡を打ち切らないこともある）³⁹⁾、次にアドヒアランスを予測するランダム化前とランダム化後の共変量を適切に調整する必要がある。ランダム化後の共変量が過去の治療の影響を受けない場合は、層別、回帰モデリング、共変量マッチング、傾向スコアモデリングなどの従来の統計手法で対応できる。ランダム化後の共変量が割り付け後の治療の影響を受ける場合には（多くの状況でこの可能性は否定できないはずである）、治療・交絡変数フィードバックを調整するために g-methods が必要となる¹²⁾。G-methods には (i) parametric g-formula, (ii) 逆確率重み付け（必要に応じて周辺構造モデルの特定）、(iii) 構造ネストモデルの g-estimation, および (iv) それらを拡張した二重ロバスト推定法が含まれる¹²⁾⁴⁰⁾。これらの手法は実質的に同じ識別仮定に依拠するが（g-estimation のみ正值性の仮定を一部緩和できる）、互いに異なるモデリング仮定を必要とする（表 2）。実際の適用においては、推定対象および利用可能なデータの性質を考慮したうえで、各手法の特性を踏まえて適切な方法を選択するか、複数手法を適用してモデリング仮定の頑健性を確認すべきである。G-methods の実装にあたっては、R や SAS などの標準的な統計ソフトウェアで利用可能な専用パッケージがある^{41)~43)}。なお当然ではあるが、いずれの手法を用いるとしても、測定された共変量の調整が効果推定値から未測定変数によるすべてのバイアスを除去する保証はなく、これは RCT から per-protocol 効果を求める場合でも例外ではない。

ベースラインのランダム化を、治療開始時の交絡の適切な調整に必要な仮定（交絡変数がすべて測定され、解析に用いるモデルが正しく特定されている）に置き換えることを除けば、RCT と観察研究とで解析方針に本質的な違いはない。観察研究における intention-to-treat 解析では、条件付き交換可能性および正值性の仮定とベースライン共変量の調整を通じて層内でのランダム化をエミュレートする。追跡不能が非ランダムである場合、

それに影響を与えるベースラインおよびベースライン後の共変量を逆確率重み付けにより調整して選択バイアスに対処する必要性は、RCT と観察研究の両方に共通する。同様に、RCT と観察研究のいずれにおいても、per-protocol 効果を推定するには、追跡不能による選択バイアスへの対処に加え、治療戦略からの逸脱を予測するベースラインおよびベースライン後の共変量を g-methods により適切に調整することで、不完全なアドヒアランスによる交絡に対処する必要がある。その他、交絡変数の明示的な調整を必要としない方法として、操作変数法（instrumental variable methods）や回帰不連続デザイン（regression discontinuity design）が挙げられるが、その適用範囲は時間固定治療の比較的単純な因果的な問いに限られることが多く、また大半の状況で正当化の難しい別の識別仮定が必要となる。

多くの因果的な問いは「アウトカム発生までの時間に対する治療効果」に関わるため、生存時間解析が必要となる。Cox 比例ハザードモデルは長年にわたり RCT と観察研究での標準的手法とされてきたが、このモデルから得られるハザード比の推定値には解釈上の限界が存在し、因果的な対比として十全ではない。ハザード比に関する第一の問題は、それが時間依存的に変化することである⁴⁴⁾⁴⁵⁾。比例ハザードモデルを用いた研究で報告される単一のハザード比は、実際に各時点でのハザード比が変化する（比例ハザード性が成立しない）場合であっても時間変化するハザード比の時点を通した（ある種の）加重平均として解釈できるが、追跡期間の設定や（本来興味の対象外である）打ち切り時間分布によって異なる値をとりうる。第二の問題は、ハザード比にはハザードに内在する選択バイアスが構造的に組み込まれていることである（built-in selection bias）⁴⁴⁾。時点 t ($t = 0, 1, 2, \dots$ は任意の単位の時間区間) におけるハザード比は、時点 $t-1$ まで生存した条件付き集団での比較となる。治療がアウトカムに効果を持つ場合、治療初期にイベントを経験しやすい (susceptible) 対象者が選択的に除外されることで、治療群ごとのハザードには選択バイアスが生じ

(depletion of susceptibles), 因果的な解釈が難しくなる。第三の問題は、周辺ハザード比が条件付きハザード比の重み付け平均として表現できない併合不能性 (non-collapsibility) である⁴⁶⁾。この現象は交絡の有無や効果修飾の存在とは無関係に生じるため、交絡変数でない共変量を調整した場合でもハザード比の推定値は変化する (しかも第一の問題と組み合わせり、そのハザード比の推定値は「平均」ハザード比に過ぎない)。

これらの課題に対処するために、絶対リスク (累積発生率) に基づく生存時間分布の比較が推奨される⁴⁷⁾。この手法は、ハザード比が追跡期間全体を通じて一定であるという仮定を要さず、また因果的な解釈も可能な妥当な対比となる。時間固定治療の研究では、Kaplan-Meier 法や、競合リスク存在下の総合効果 (第 3.5 節) を求める場合は累積発生関数 (cumulative incidence function) により、打ち切りがなかった場合の累積発生率を推定できる。非ランダムな追跡不能 (RCT と観察研究) と交絡 (観察研究) の調整には、それぞれ打ち切られない条件付き確率と条件付き治療確率に基づく逆確率重み付けを利用することができ、重み付け Kaplan-Meier 推定量や累積発生関数によって打ち切りのない標的試験で得られるであろう累積発生率曲線をエミュレートできる。また、離散時間に対する Cox 比例ハザードモデルとほぼ同等のパラメトリックモデルである pooled logistic 回帰モデルによる標準化 (parametric g-formula) は、時間固定および時間依存性治療の双方に適用可能なアプローチである¹²⁾。この方法では、データを人時間形式 (person-time format) に構成し、治療と共変量で条件付けた離散時間ハザード (discrete-time hazards) を時点ごとにパラメトリック推定し、その累積から個人ごとに異なる共変量の値ごとに生存曲線を算出する (この際、回帰モデルの説明変数に治療と時点の交差項を含むことで比例ハザード性の制約を緩和することができるが、これは同じ交差項を時間依存性共変量として用いる時間依存性 Cox モデルとほぼ同等のモデルとなる)。最後に、生存確率 (または累積発生率) を時点ごとに集団全体で平均するこ

とで、調整生存曲線 (または調整累積発生率曲線) を得る。時間依存性治療の場合には、離散時間ハザードに加えて時間依存性共変量のモデリングと、推定したモデルからのモンテカルロ・シミュレーションによる数値近似が必要となるが、パッケージで簡便に実行できる^{41)~43)}。または、時点ごとの治療確率に基づく逆確率重みを累積した重み付け pooled logistic 回帰モデルを用いることで、周辺構造 Cox モデルの近似的な推定を通してモンテカルロ・シミュレーションを経由せずに反事実の生存確率を指定することもできる。

このように、よく用いられる交絡調整手法は、アウトカム回帰 (治療・共変量ごとのアウトカム平均) のモデリングに基づく標準化と、傾向スコア (共変量ごとの治療確率) のモデリングに基づく逆確率重み付けの 2 つのアプローチに大別できる。両者とも「追跡が完全な標的試験での累積発生率曲線」を共通の推定対象とする調整累積発生率曲線を提供するが、異なるモデリング仮定に依拠している。両方のモデルを用いるが、少なくとも一方のモデリングが正しければバイアスなく累積発生率曲線の推定値を得られる方法として、二重ロバスト推定量を用いることもできる¹²⁾。

4. 標的試験エミュレーションの適用事例

標的試験エミュレーションのフレームワークに基づいた研究の事例を紹介する。題材は、骨粗鬆症を併存する維持透析患者を対象に、デノスマブと経口ビスホスホネート製剤の心血管安全性および骨折予防効果を比較した観察研究である⁶⁾。DeSC ヘルスケア株式会社が提供するレセプトデータを用い⁴⁸⁾、2014 年 4 月 1 日から 2022 年 10 月 31 日までのデータを解析対象とした。標的試験の特定とエミュレーションの対応表は表 3 のとおりである。

4.1 適格基準のエミュレーション

維持透析導入から 90 日以上が経過した 50 歳以上の骨粗鬆症患者を対象とした。この集団から、骨粗鬆症の治療目的でデノスマブ 60 mg の皮下投与または経口ビスホスホネート製剤を新規に開始した患者を特定した。いずれか最も早い調剤日を

表 3 維持透析患者の骨粗鬆症に対するデノスマブと経口ビスホスホネート製剤を比較する標的試験の特定とエミュレーション

	標的試験の特定	標的試験のエミュレーション
因果的な 推定対象	適格基準 選択基準： (1) 2015年4月1日から2021年10月31日の期間に、日本国内で骨粗鬆症と診断された50歳以上の患者 (2) 90日以上維持透析歴を有する患者 除外基準： (1) 追跡開始以前に悪性腫瘍、骨巨細胞腫、または骨パジェット病の既往を有する患者 (2) 追跡開始以前に腎移植の既往を有する患者 (3) 追跡開始の前日までにデノスマブまたは経口・静注ビスホスホネート製剤の調剤記録を有する患者 (4) 追跡開始日にデノスマブと経口ビスホスホネート製剤の両方の調剤記録を有する患者 (5) 追跡開始以前の90日間に急性心筋梗塞、脳卒中、または心不全で入院した患者 治療戦略 (1) デノスマブ 60 mg の皮下投与の開始 (2) 経口ビスホスホネート製剤の開始 開始後の治療期間の決定は、臨床医の裁量に委ねる 治療割り付け 患者は非盲検下でベースライン時にいずれかの治療戦略にランダム割り付けられる アウトカム 安全性のアウトカム： MACE（急性心筋梗塞、脳卒中、入院を伴う心不全、または心血管死） 有効性のアウトカム： 骨折（椎体、股関節、骨盤、大腿、下腿、足首、肩、前腕、手首の骨折を含む） 追跡開始と終了 追跡は治療割り付け時点で開始され、関心のあるアウトカムの発生、追跡不能、追跡の管理的終了（追跡開始から3年時点または2022年10月31日）のいずれか早い日に終了する 因果的な対比 Intention-to-treat 効果： 効果指標は、3年リスク、リスク差、リスク比とする	適格基準 選択基準： (1)–(2) 標的試験と同じ 除外基準： (1)–(3) 標的試験と同じ（データベース内で適及可能な限りの期間において判定） (4)–(5) 標的試験と同じ (6) 追跡開始以前のデータベースへの登録期間が365日未満の患者 治療戦略 標的試験と同じ（調剤記録に基づき投与開始を定義した） 治療割り付け 患者は観察されたデータと一致する治療戦略に分類され、ベースライン共変量の層内でのランダム化が想定された アウトカム 安全性のアウトカム： 標的試験と同じ（診断コードと医薬品または診療行為コードの組み合わせにより操作的に定義した） 有効性のアウトカム： 標的試験と同じ（診断コードと診療行為コードの組み合わせにより操作的に定義した） 追跡開始と終了 標的試験と同じ（追跡不能は、データベースからの脱落として定義した） 因果的な対比 Intention-to-treat 効果の観察研究におけるアナロジー： 効果指標は、標的試験と同じ
識別仮定	Intention-to-treat 効果： ランダム割り付けによる交換可能性	Intention-to-treat 効果の観察研究におけるアナロジー： ベースライン共変量の調整による条件付き交換可能性 ベースライン共変量は以下の通り：人口統計学的特性、追跡開始年、骨粗鬆症の初回診断からの期間、初回適格性（維持透析と骨粗鬆症の両方）を満たしてからの期間、骨折歴および骨折手術歴、併存疾患の既往歴、骨粗鬆症治療薬の使用歴、他の薬剤の使用歴、医療利用、透析の種類およびデータベース上の初回透析からの期間
推定量	Intention-to-treat 解析： 各治療戦略における累積発生率曲線から3年リスク（累積発生率）、リスク差、およびリスク比を推定する	Intention-to-treat 解析： 標的試験と同じ

MACE, 主要心血管イベント

追跡開始時点（タイムゼロ）と定義し、2015年4月1日から2021年10月31日の期間に限定した。経口ビスホスホネート製剤を比較対照薬として選択したのは、日本では維持透析患者への使用が承認されており、腎疾患患者に広く用いられているためである。

以下のいずれかに該当する患者は研究対象から除外した：(i)追跡開始以前に悪性腫瘍、骨巨細胞腫、または骨パジェット病の既往を有する患者（デノスマブおよびビスホスホネート製剤が骨粗鬆症に対して投与されていることを確実にするため）；(ii)追跡開始以前に腎移植の既往を有する患者；(iii)追跡開始の前日までにデノスマブまたは経口・静注ビスホスホネート製剤の調剤記録を有する患者；(iv)追跡開始日にデノスマブと経口ビスホスホネート製剤の両方の調剤記録を有する患者；(v)追跡開始以前の90日間に急性心筋梗塞、脳卒中、または心不全で入院した患者；(vi)追跡開始以前のデータベースへの登録期間が365日未満の患者。

4.2 治療戦略のエミュレーション

治療戦略は時間固定治療であり、(i)デノスマブの投与を開始する、または(ii)経口ビスホスホネート製剤の投与を開始する、の2つであった。開始後の治療期間の決定は、臨床医の裁量に委ねられた。

4.3 治療割り付けのエミュレーション

交絡調整のためのベースライン共変量として、既報や専門的知見に基づき、modified disjunctive cause criterion を満たす以下の変数を選択した：人口統計学的特性、追跡開始年、骨粗鬆症の初回診断からの期間、初回適格性（維持透析と骨粗鬆症の両方）を満たしてから期間、骨折歴および骨折手術歴、併存疾患の既往歴、骨粗鬆症治療薬の使用歴、他の薬剤の使用歴、医療利用、維持透析の種類およびデータベース上の初回維持透析からの期間。

交絡変数によるベースラインの不均衡を調整するために、逆確率重み付けを適用した。この方法により、患者は観察されたデータと一致する治療戦略に分類され、ベースライン共変量の層内での

ランダム化が仮想的に達成された。

4.4 アウトカムのエミュレーション

安全性のアウトカムは、急性心筋梗塞、脳卒中、入院を伴う心不全、または心血管死を含む主要心血管イベント（major adverse cardiovascular events, MACE）とした。冠動脈疾患が心不全入院の最も一般的な原因であるため、MACEには心不全も含めた。MACEは、循環器専門医の臨床的知見に基づき、診断コードと医薬品または診療行為コードの組み合わせによりイベント別に定義した。日本でのバリデーション研究に照らし合わせると⁴⁹⁾、これらの定義の陽性的中率は100%に近いことが期待された（ただし、心血管死についてはバリデーション研究は実施されていない）。

有効性のアウトカムは骨折（椎体、股関節、骨盤、大腿、下腿、足首、肩、前腕、手首の骨折を含む）とした。骨折は、米国のバリデーション研究を参照しつつ⁵⁰⁾、整形外科専門医の臨床的知見から日本の実態に即すように、骨折部位の診断コードおよび特異的な診療行為コードの組み合わせによって定義し、陽性的中率は90%を超えることが期待された。

4.5 追跡開始と終了のエミュレーション

追跡はデノスマブまたは経口ビスホスホネート製剤のいずれか最も早い調剤日に開始し、関心のあるアウトカムの発生、追跡不能（データベースからの脱落）、追跡の管理的終了（追跡開始から3年時点または2022年10月31日）のいずれか早い日に終了した。推定対象は総合効果であったため、競合イベントである死亡（MACEに対する解析では他の原因による死亡）では追跡を打ち切らず、アウトカム発生なしとして管理的終了まで追跡を継続した³⁶⁾。

4.6 因果的な対比のエミュレーション

エミュレーションは、intention-to-treat 効果の観察研究におけるアナロジーに焦点を当て、デノスマブまたは経口ビスホスホネート製剤の治療戦略を開始することの効果の推定した。治療戦略へのアドヒアランスを合理的に評価できなかったため、当該原稿のMethods EditorであったHernánの助言もあり、per-protocol 効果の観察研究にお

けるアナロジーは推定しなかった。

4.7 解析手法のエミュレーション

Intention-to-treat 解析では、各治療戦略における累積発生率曲線から3年リスク(累積発生率)、リスク差、およびリスク比を推定した。95%信頼区間は、500個のブートストラップサンプルからのパーセンタイルブートストラップ法により算出した。

結果の頑健性を確認するために、以下の5つの感度解析を実施した：(i)日本でのみ承認されているミノドロン酸を経口ビスホスホネート製剤から除外；(ii)維持透析を血液透析に限定；(iii)患者の追跡を競合イベントの発生時に打ち切る；(iv)追跡開始以前の365日間に急性心筋梗塞、脳卒中、または心不全で入院した患者を除外；(v)研究対象者を骨粗鬆症治療薬の使用歴のない患者に限定。また、効果修飾を評価するために、性別、年齢(80歳未満 vs. 80歳以上)、MACEの既往歴、骨折歴、初回適格性(維持透析と骨粗鬆症の両方)を満たしてからの期間(18カ月未満 vs. 18カ月以上)で層別したサブグループ解析も実施した。

5. おわりに

本稿では、標的試験エミュレーションの概念を整理するとともに、方法論的側面と実装上の考慮事項を詳細に検討し、さらにはその応用事例を紹介した。標的試験エミュレーションの最大の貢献は、因果的な問いに内在する曖昧さを大幅に解消することにある。問いが不明瞭なままでは、解釈困難な効果推定値を生み出すだけであり、その推定値から科学的・臨床的意思決定への活用を正当化することはできない。標的試験エミュレーションは、研究デザインに起因するバイアスを防止することはできるが、交絡および選択バイアスを調整するための情報の欠如や測定誤差といったデータそのものの限界から生じるバイアスを排除するものではない。しかし、デザインバイアスを回避することで、標的試験フレームワークは、研究者にデータバイアスに焦点を当てさせ、それらの制約が将来の研究においていかに緩和できるかを検討する機会を提供する。このように、標的試験フ

レームワークは研究デザインおよびデータ解析における優れた実践を促すためのアプローチと見なすことができる。また、このアプローチは研究成果を完全かつ透明性をもって報告するうえでも大きな価値を持つと考えられるが、本稿執筆時点において、標的試験エミュレーションの報告ガイドラインは開発段階にある⁵¹⁾。因果的な問いに真摯に向き合う日本の疫学者が、標的試験の明示を通してその問いを洗練された形で具体化し、観察研究に基づく因果推論の質を高め、より確かな科学的根拠に立脚した意思決定へと進むことを期待したい。

謝辞

本論文の作成にあたり医学・薬学的見地からご助言いただいた、京都大学大学院医学研究科社会健康医学系専攻薬剤疫学分野の安納崇之氏、嶋田博樹氏、榊田崇一郎氏、松本恵実氏に深謝する。

利益相反

深澤俊貴は、エーザイ株式会社および協和キリン株式会社との産学共同講座として京都大学大学院医学研究科に設置されているデジタルヘルス学講座に所属していた。また、アストラゼネカ株式会社から研究資金の提供を受けるとともに、旭化成ファーマ株式会社、株式会社JMDC、およびリアルワールドデータ株式会社からコンサルタント料を受領している。加えて、深澤俊貴の配偶者はMSD株式会社の社員である。篠崎智大は、参天製薬株式会社およびファイザー株式会社からコンサルタント料を受領している。

著者貢献

深澤俊貴は、本論文を構想し、起草した。篠崎智大は、本論文の構想に貢献し、重要な知的内容について批評的な推敲を行った。著者全員が出版原稿の最終承認を行った。

文 献

- 1) Dahabreh IJ, Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. *JAMA* 2024 ; **331** : 1845-53.
- 2) Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol* 2016 ; **79** : 70-5.
- 3) Hernán MA, Sterne JAC, Higgins JPT, Shrier I, Hernández-Díaz S. A structural description of biases that generate immortal time. *Epidemiology*

- 2025 : **36** : 107–14.
- 4) Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016 ; **183** : 758–64.
- 5) Hansford HJ, Cashin AG, Jones MD, et al. Reporting of Observational Studies explicitly aiming to emulate randomized trials : a systematic review. *JAMA Netw Open* 2023 ; **6** : e2336023.
- 6) Masuda S, Fukasawa T, Matsuda S, Kawakami K. Cardiovascular safety and fracture prevention effectiveness of denosumab versus oral bisphosphonates in patients receiving dialysis : a target trial emulation. *Ann Intern Med* 2025 ; **178** : 167–76.
- 7) Hernán MA, Dahabreh IJ, Dickerman BA, Swanson SA. The target trial framework for causal inference from observational data : Why and when is it helpful? *Ann Intern Med* 2025 ; **178** : 402–7.
- 8) Robins J. A new approach to causal inference in mortality studies with a sustained exposure period — application to control of the healthy worker survivor effect. *Math Model* 1986 ; **7** : 1393–512.
- 9) McAuley J. SYM16 : Reporting comparative effectiveness studies using the Target Trial Framework — the TARGET guidelines. Presented at : World Congress of Epidemiology 2024 : September 24–27, 2024 : Cape Town, South Africa. [<https://www.wce2024.org/wp-content/uploads/2024/10/James-SYM16-Reporting-comparative-effectiveness-studies-using-the-Target-Trial-Framework-%E2%80%93-the-TARGET-guidelines.pdf> (accessed 2025-4-4)]
- 10) Hernán MA. Invited commentary : hypothetical interventions to define causal effects — afterthought or prerequisite? *Am J Epidemiol* 2005 ; **162** : 618–20 ; discussion 621–2.
- 11) Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology* 2011 ; **22** : 368–77.
- 12) Hernán MA, Robins JM. *Causal Inference : What If*. Chapman & Hall/CRC, 2020.
- 13) Desai RJ, Wang SV, Sreedhara SK, et al. Process guide for inferential studies using healthcare data from routine clinical practice to evaluate causal effects of drugs (PRINCIPLED) : considerations from the FDA Sentinel Innovation Center. *BMJ* 2024 ; **384** : e076460.
- 14) Ehrenstein V, Hellfritsch M, Kahlert J, et al. Validation of algorithms in studies based on routinely collected health data : general principles. *Am J Epidemiol* 2024 ; **193** : 1612–24.
- 15) Degtiar I, Rose S. A review of generalizability and transportability. *Annu Rev Stat Appl* 2023 ; **10** : 501–24.
- 16) Schneeweiss S, Patorno E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocr Rev* 2021 ; **42** : 658–90.
- 17) Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol* 2016 ; **26** : 674–80.
- 18) VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology* 2009 ; **20** : 880–3.
- 19) Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med* 2017 ; **377** : 1391–8.
- 20) Vrijens B, De Geest S, Hughes DA, et al. A new taxonomy for describing and defining adherence to medications. *Br J Clin Pharmacol* 2012 ; **73** : 691–705.
- 21) Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ* 2018 ; **363** : k3532.
- 22) Fukasawa T, Iwagami M, Hara A, Urushihara H. Design Diagram : A framework for visualizing study designs using real-world data and improving study reproducibility. *Jpn J Pharmacoepidemiol* 2023 ; **28** : 39–55.
- 23) Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology* 1999 ; **10** : 37–48.
- 24) Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams : Pitfalls and tips. *J Epidemiol* 2020 ; **30** : 153–62.
- 25) Feeney T, Hartwig FP, Davies NM. How to use directed acyclic graphs : guide for clinical researchers. *BMJ* 2025 ; **388** : e078226.
- 26) VanderWeele TJ. Principles of confounder selection. *Eur J Epidemiol* 2019 ; **34** : 211–9.
- 27) Inoue K, Sakamaki K, Komukai S, Ito Y, Goto A, Shinozaki T. Methodological tutorial series for epidemiological studies : Confounder selection and sensitivity analyses to unmeasured confounding from epidemiological and statistical perspectives. *J Epidemiol* 2025 ; **35** : 3–10.
- 28) Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls : a tool for detecting confounding and bias in observational studies. *Epidemiology* 2010 ; **21** : 383–8.
- 29) Weinstein EJ, Ritchey ME, Lo Re V 3rd. Core concepts in pharmacoepidemiology : Validation of health outcomes of interest within real-world healthcare databases. *Pharmacoepidemiol Drug Saf* 2023 ; **32** : 1–8.

- 30) Lanes S, Beachler DC. Validation to correct for outcome misclassification bias. *Pharmacoepidemiol Drug Saf* 2023 ; **32** : 700-3.
- 31) Greenland S, Gustafson P. Accounting for independent nondifferential misclassification does not increase certainty that an observed association is in the correct direction. *Am J Epidemiol* 2006 ; **164** : 63-8.
- 32) Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. *Modern epidemiology*, 4th ed. Philadelphia, PA : Lippincott Williams and Wilkins, 2021.
- 33) Avillach P, Coloma PM, Gini R, et al. Harmonization process for the identification of medical events in eight European healthcare databases : the experience from the EU-ADR project. *J Am Med Inform Assoc* 2013 ; **20** : 184-92.
- 34) Gini R, Schuemie M, Brown J, et al. Data extraction and management in networks of observational health care databases for scientific research : A comparison of EU-ADR, OMOP, Mini-Sentinel and MATRICE strategies. *EGEMS(Wash DC)* 2016 ; **4** : 1189.
- 35) Arena PJ, Huang K, Löfling L, et al. Validation of safety outcomes in routinely collected data : Lessons learned from a multinational postapproval safety study. *Pharmacoepidemiol Drug Saf* 2023 ; **32** : 592-6.
- 36) Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med* 2020 ; **39** : 1199-236.
- 37) Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012 ; **9** : 48-55.
- 38) Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed as observational studies. *Ann Intern Med* 2013 ; **159** : 560-2.
- 39) Robins JM, Finkelstein DM. Correcting for non-compliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000 ; **56** : 779-88.
- 40) van der Laan MJ, Rose S. *Targeted Learning*. New York, NY : Springer, 2011.
- 41) Hernán MA. The code used in the *What If* book. Miguel Hernan's Faculty Website. [<https://miguelhernan.org/whatifbook> (accessed 2025-4-4)]
- 42) CAUSALab. Software. CAUSALab — A Center to Learn What Works. [<https://causalab.hsph.harvard.edu/software/> (accessed 2025-4-4)]
- 43) CAUSALab. Open Source Software Tools for Causal Inference. GitHub. [<https://github.com/CausalInference> (accessed 2025-4-4)]
- 44) Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010 ; **21** : 13-5.
- 45) Stensrud MJ, Hernán MA. Why test for proportional hazards? *JAMA* 2020 ; **323** : 1401-2.
- 46) Didelez V, Stensrud MJ. On the logic of collapsibility for causal effect measures. *Biom J* 2022 ; **64** : 235-42.
- 47) Stensrud MJ, Hernán MA. Invited Commentary : Why use methods that require proportional hazards? *Am J Epidemiol* 2025. doi : 10.1093/aje/kwae361.
- 48) Yasunaga H. DeSC database. *Ann Clin Epidemiol* 2025 ; **7** : 46-9.
- 49) Fujihara K, Yamada-Harada M, Matsubayashi Y, et al. Accuracy of Japanese claims data in identifying diabetes-related complications. *Pharmacoepidemiol Drug Saf* 2021 ; **30** : 594-601.
- 50) Wright NC, Daigle SG, Melton ME, Delzell ES, Balasubramanian A, Curtis JR. The design and validation of a new algorithm to identify incident fractures in administrative claims data. *J Bone Miner Res* 2019 ; **34** : 1798-807.
- 51) Hansford HJ, Cashin AG, Jones MD, et al. Development of the TrAnsparent ReportinG of observational studies Emulating a Target trial (TARGET) guideline. *BMJ Open* 2023 ; **13** : e074626.

〔早期公開 : 2025 年 6 月 27 日〕

●Special Issue on “Target Trial Emulation : A New Paradigm for Causal Inference in Observational Studies”

Target Trial Emulation : A Framework for Strengthening Causal Inference in Observational Studies

Toshiki FUKASAWA^{*1}, Tomohiro SHINOZAKI^{*2}

^{*1} Department of Pharmacoepidemiology, Graduate School of Medicine and Public Health, Kyoto University, Kyoto, Japan

^{*2} Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science, Tokyo, Japan

〈Abstract〉

When a randomized controlled trial (RCT) is infeasible, unethical, or untimely, causal inference from observational data can serve as an effective alternative for scientific and clinical decision-making. However, observational studies harbor methodological vulnerabilities—not only confounding due to lack of randomization, but also selection bias or immortal time, arising from flawed study designs—that can fundamentally distort effect estimates. Target trial emulation has gained prominence as a framework for addressing these challenges. This approach has two steps : (1) specifying the protocol of a hypothetical pragmatic RCT (the target trial) that would answer the causal question of interest, and (2) explicitly emulating that trial with existing observational data. Its greatest contribution is the elimination of ambiguous causal questions in observational studies, transforming them into well-defined causal estimands. In this article, we synthesize the conceptual foundations of target trial emulation and detail methodological considerations for its implementation. As an illustrative example, we describe an observational study that compared denosumab with oral bisphosphonates for cardiovascular safety and fracture-prevention effectiveness in maintenance dialysis patients with osteoporosis. The target trial framework offers a structured approach that prevents design-induced biases and clarifies the limitations inherent in observational data, thereby enabling epidemiologists who grapple with causal questions to draw more valid inferences.

(*Jpn J Pharmacoepidemiol* 2025; 30(2) : 53–73)

Keywords: causal inference, epidemiology, observational studies, target trial emulation, randomized controlled trials