

# 「もしも」の科学—医療・健康科学における因果推論

篠崎 智大

もしも、あの時にあっていたら——という思いは、人間である以上は日常的に心に思い浮かべるものである。私の「フィールド」である医学・健康科学分野において、このような素朴な空想は例えば「もしこの対象者に異なる治療介入を与えていたら」という臨床的な問いに置き換わる。このような問いに、定式化（モデル）と科学的な議論の土台（議論の条件と方法の整理）を与えるのが因果推論（causal inference）という分野である。

医学研究の世界では長らくランダム化比較試験（RCT）が研究デザインのゴールドスタンダードと見なされてきた。RCTの重要性は、実臨床の中で半自動的に蓄積される診療報酬明細や電子カルテ情報を統合したリアルワールドデータ（RWD）全盛の昨今においても変わらない（篠崎, 2022a）。一方で、人の健康や疾患を対象とする研究では倫理的・実務的な制約からRCTが実施できない場面が少なくなく、RWDの利活用が医療、公衆衛生、産業界から求められることも事実である。興味深いのは、医学分野においてRCTとRWDという大きな2つの潮流がある中で両者が二項対立に陥るのではなく、(1) RCTでは実臨床というリアルワールドに近い状況を志向した動きが広がっており、(2) RWDを用いたデータベース研究ではRCTを参照するようなデータ解析を目指す

アプローチが主流になりつつあることである。これらの流れの接点にあるのが因果推論であり、冒頭に挙げた「もしも…」という反事実（counterfactual）どうしの比較が、リサーチクエストンとして科学的な共通言語として広がっていることを表している。

RCTをはじめとする臨床試験は、人を対象とした実験的な研究である。ここで「実験的」というのは、「人の健康に関する事象に影響を与える要因を制御する介入」を含むことを指す。この介入がランダムに割り当てられるとき、特定の介入群A（例えば試験治療群）と別の介入群B（例えば「通常治療群」）は同じような集団構成となっていることが期待され、介入によって変化させた要因について比較可能性が担保されることは直観的にもお分かりいただけるだろう。因果推論のモデルではこの状態を「交換可能性」が成り立っているといい、介入群AとBの介入内容を入れ替えたとしても（上の例ではAを「通常治療群」、Bを「試験治療群」としても）期待的に同じ結果が得られることを意味する。もちろん介入群をランダムに分ける際の偶然性によって結果は確率的に変動するが、これは統計的に評価できる。

介入をランダム化しようがしまいが（言い換えれば、交換可能性が成り立っていようがまいが）、統計学や機械学習でとらえられるのは

「異なる状態にある異なる集団の比較」という関連 (association) に過ぎない。つまりわれわれは、異なる介入内容の、A と B という異なる群で結果が異なるかどうかを見ることしかできない。しかし、交換可能性が成り立つならば、この関連は「群 A・B をまとめた集団全体が試験治療を受けた場合」と「同じ集団全体が試験治療でなく通常治療を受けた場合」の比較と見なすことができる。これが現代の因果推論で広く受け入れられている因果関係 (causation) の定式化である。より広く言い換えると「同じ集団の異なる状態の比較」を因果関係とよぶ。

しかし、現実の臨床試験では介入が割り付け通りに行われるとは限らない。それは単に患者の不注意や不遵守によるかもしれないし、患者の状態 (副作用等) に応じて治療介入を変更しなければならないのかもしれない。特に高血圧や糖尿病、がんなどの慢性疾患に対する医療の多くは、一度きりの介入ではなく治療の継続や経時的な組み合わせが必要となる。このような治療ではその「単体での効果」は一意には定義しにくく、「治療の方針 (ルール)」の考え方を含めて拡張する必要がある (篠崎, 2022b)。さらに、患者の途中経過によっては比較されるアウトカムの測定ができなくなる場合がある。極端な例として、進行性のがん患者の QOL (生活の質) を薬剤間で比べる臨床試験では、死亡以降の QOL については、測定はおろか定義すらできなくなってしまう。このような状況で得られたデータだけを比較すると、反直観的な結果が得られることが知られている。

こうした問題意識から医薬品開発の文脈で国際的に策定されたガイドライン文書が ICH E9

(R1) である。簡単に背景を説明すると、ICH とは医薬品規制調和国際会議 (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) の略称であり、医薬品規制当局 (米国食品医薬品局、欧州医薬品庁、日本では厚生労働省・医薬品医療機器総合機構など) と製薬業界の代表者による、医薬品開発の臨床試験と規制を想定したガイドラインを策定するための国際活動である。「有効性 (efficacy)」に関する ICH のガイドラインの一つとして統計的な原則を定めたのが 1998 年の「E9」文書であり、2019 年にその補遺文書である「E9 (R1)」の最終稿が合意された。この E9 (R1) は、上に挙げたような、「治療変更」や「アウトカムの存在性に影響する事象」が生じた臨床試験で、製薬会社と規制側で合意しうる「効果」を明確化し、透明性ある意思決定を目指した文書である。因果推論の定式化によって、試験計画と解析結果の解釈および意思決定の曖昧さを排除することが意図されている (篠崎・菅波, 2022)。「もしも」の観点が医薬品開発・承認という現実的な意思決定の場面に活かされていると言える。

医学研究は臨床試験だけではない (というより、臨床試験は人を対象とした研究ではごく限られた状況で実施できる特殊な形態である)。介入のランダム割り付けを伴わない「観察研究」では、群間の交換可能性が明らかに成り立たず (そもそも特定の医療を受ける人はその必要があって受けるものであり、それを受けていない人たちとは質的に異なるのが常である)、このとき交絡 (confounding) が生じていると

いう。しかし因果推論の枠組みによって、このような観察研究においても介入研究と同様に効果を定義し、交絡を「調整」するための手法（層別解析や回帰モデル・傾向スコアモデルによるその近似）とそのための仮定を整理することができる（Hernán & Robins, 2020）。このような理論が受け入れられていることは、実験的な介入が因果推論に必要でも十分でもなく、RCTという研究デザインが因果推論にお墨付きを与えてくれるという形式的な見方はもはや本質的でないことを示している。むしろ、介入の有無は効果を求めるための仮定の蓋然性を部分的に保証する手順という位置づけであり、観察研究もRCT（臨床試験）も因果推論においては同じ枠組み（モデル）で捉えるべきという見方ができる。この枠組みにより、医療の適切な一般化や個別化に必要な効果の異質性評価など、実験的介入によってかえって損なわれうる側面があることも知られるようになってきた。

このような概念的な発展は、観察研究を仮想的な臨床試験の模倣としてとらえ直す「標的試験エミュレーション（target trial emulation）」の考え方につながる（Hernán & Robins, 2016; 深澤・篠崎, 2025）。標的試験とは、知りたい治療効果を検証できるような仮想的な（実際には実施されない）臨床試験を指し、データ上のエミュレーション（模倣）技法と組み合わせることで、観察研究データから標的試験を疑似的に再現するための統計解析の枠組みを提供する。この枠組みによって、複雑な治療変更を前提とした動的な治療レジメン（dynamic treatment regimes）として定式化されるような因果効果まで含めて、臨床試験デザインとい

う医学研究者にとっての日常用語に翻訳することが可能となる。こうした共通言語によって、必ずしも因果推論を専門としない研究者（臨床家、疫学者、統計家）の間で足並みをそろえて、直感的かつ透明性をもって観察データの解析計画を立てることができるようになる。技術的にも、正統的な因果推論の理論ではg-formula、構造ネストモデル、周辺構造モデルといった特殊な統計解析方法論（Hernán & Robins, 2020; 篠崎他, 2026）が必要とされていたリサーチクエストションも、「臨床試験の模倣」に落とし込むことで単純明快に効果推定を実装できるようになる。少なくとも、公衆衛生大学院のカリキュラムを修めた程度のスキルと知識があれば、数理統計学の知識は少なくとも理解して使うことができる。

医学研究での因果推論は、もともとは観察研究で比較を行うことの正当化や限界の提示が意図されていた。この意味で、ICH E9 (R1) ガイドラインは、臨床試験から観察研究へ接近する大きな動きと言える。それと時を同じくして、標的試験エミュレーションという形で観察研究から臨床試験に接近する動きが主流になりつつあるのは偶然だろうか。これらの接点には因果推論があり、そして今この時代が、「もしも」を通した定式化が有効であることが広く認識され始めた時期であることを思うと不思議ではないのかもしれない。もちろん、この「接点」において各々の文脈での議論は完全には重ならず、すれ違いを生じることはある。例えば、ICH ガイドラインは医薬品開発・承認のための臨床試験（いわゆる治験）に特化した文書であり、臨床試験一般をカバーするよう意図された

ものではない。治験と治療法確立のための一般の臨床試験とは目的が同じとは限らないので、E9 (R1) で射程外の効果が必要となる臨床試験や、それを模倣するための観察研究は存在する。また、観察研究で興味がある「効果」はすべて臨床試験の介入によって翻訳できるわけではない (Suzuki et al., 2025)。因果にとって介入を要件とするべきかどうかは、因果の存在論

または認識論をめぐる議論につながり、現代の因果推論においてもホットなトピックである (Robins, 2022; 清水, 2025)。しかし、このような臨床試験と観察研究の「すれ違い」は因果推論を軸とする医学分野での議論の未熟さではなく、人を相手とした文脈依存の問いの豊かさを反映しているのである。

### 参考文献

- Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial: when a randomized trial is not available. *American Journal of Epidemiology*, 183 (8), 758-764.
- Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC.
- Robins, J. (2022). Interview with Jamie Robins. *Observational Studies*, 8 (2), 37-76.
- Suzuki, E., Shinozaki, T. and Yamamoto, E. (2025). Natural effects and separable effects: insights into mediation analysis. *Current Epidemiology Reports*, 12, 20.
- 篠崎 智大 (2022a). ビッグデータ論文を読む際の10か条. *薬理と治療*, 50 (2), 187-192.
- 篠崎 智大 (2022b). 治療経過に応じて決まる治療方針の因果効果. *医学のあゆみ*, 280 (5), 508-516, 2022
- 篠崎 智大, 菅波 秀規 (2022). 臨床研究におけるエスティマンド. *医学のあゆみ*, 280 (5), 473-479.
- 篠崎 智大・萩原 康博・田栗 正隆・松山 裕 (2026). *疫学・臨床研究のための因果推論—Robins の g-methods による現実問題へのアプローチ*. 東京: 朝倉書店.
- 清水 雄也 (2026). 因果の正体を哲学する. *数学セミナー*, 771, 12-16.
- 深澤 俊貴, 篠崎 智大 (2025). 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. *薬剤疫学*, 30 (2), 53-73.



篠崎 智大 (しのぎき・ともひろ)

[専門] 生物統計学・疫学理論

[主たる著書・論文]

・坂巻 顕太郎・篠崎 智大 (監修) (2023) 『生物統計学の道標: 研究デザインから論文報告までをより深く理解するための24講』厚生労働統計協会

・篠崎 智大・萩原 康博・田栗 正隆・松山 裕 (2026) 『疫学・臨床研究のための因果推論—Robins の g-methods による現実問題へのアプローチ』朝倉書店

[所属] 大学院情報学環/大学院医学系研究科 (兼任)

[所属学会] 日本計量生物学会, 日本疫学会, ヘルスデータサイエンス学会など