総 説

医学のための因果推論の基礎概念

Fundamental concepts for causal inference in medicine

田中司朗^{*1}
Shiro Tanaka^{*1}

^{*1} 京都大学大学院医学研究科臨床統計学

^{*1} Department of Clinical Biostatistics, Graduate School of Medicine, Kyoto University

e-mail : tanaka.shiro.8n@kyoto-u.ac.jp

A central problem in medical research is how to make inferences about the causal effects of treatments or exposures. In this article, we review fundamental concepts for making such inferences in randomized clinical trials or observational studies. The statistical framework consists of potential outcomes, an assignment mechanism, and probability distributions. Randomization-based and model-based methods of statistical inference are illustrated with a series of extracorporeal membrane oxygenation (ECMO) clinical trials, which are thought-provoking in that each trial used different assignment mechanisms.

Key words: assignment mechanism, collapsibility, confounding, model-based inference, outcome adaptive randomization, propensity score, potential outcomes, randomization-based inference

1. はじめに

医学において、原因と結果の関係について言及するとき、しばしば反事実 (counterfactual) の形をとる。たとえば「遷延性肺高血圧症の新生児に、ECMO (体外式膜型人工肺, extracorporeal membrane oxygenation) を用いて酸素を供給していたら、命が助かったか」というとき、実際には ECMO は用いられなかったのだから、これは事実に反する命題についての推測である。

統計学的推測では、この問題を確率分布で定式化することになるが、それは大きく分けて2つある。1つ目は、観察されるデータの確率分布である。治療 A_i とアウトカム Y_i の条件付き確率分布 $\Pr(Y_i|A_i)$ がわかれば、ECMO を用いたときどのようなアウトカムが生じるかを予測することができる。

もう1つは、観察されないアウトカム (潜在結果変数, potential outcomes) を含めた定式化である (Neyman 1923, Rubin 1974)。ECMO ではなく従来療法を用いたある個人について、ECMO を用いたときのアウトカムは、実際には観察することのできない潜在的な (事実に反する) 変数である。これを導入することで、反事実に基づく因果効果を定義することができる。仮に対象者 i に従来療法を用いたときの潜在結果変数を Y_i^0 , 仮に ECMO を用いたときの潜在結果変数を Y_i^1

と定義する．両者に差があるかどうかは $\tau_i = Y_i^1 - Y_i^0$ によって調べることができる（これを，解析単位レベルの因果効果という）．相関と因果の区別，ランダム化の下での p 値の意味（Fisher 1935），交絡（confounding）や併合可能性（collapsibility）といった観察研究で生じる概念（Greenland, Robins, and Pearl 1999）は，潜在結果変数という概念を導入することで，うまく説明できる．

本論文では，1980～90 年代に行われた ECMO 臨床試験（Bartlett, et al. 1985, O'Rourke, et al. 1989, UK Collaborative ECMO Trial Group 1996）を題材に，因果推論の基礎概念について解説する．この事例は，治療をランダムに割付けることの倫理性が問われたこと，アウトカム適応的ランダム化（outcome adaptive randomization）が採用されたこと，それにもかかわらず交絡が生じたことなど教訓に富む．最近になって，がん臨床試験を中心に，アウトカム適応的ランダム化への関心が高まっている（たとえば Park, et al. 2017）．初期のアウトカム適応的ランダム化試験に再び注目する価値はあるのではないかと思う．倫理的観点からの詳細な議論については，研究者自身による考察と Statistical Science 誌での誌上討論（Ware 1989）を参照されたい．

因果推論の研究テーマは，基礎概念や理論に関係するものと統計手法を扱うものに大別できる．手法の説明は本論文の目的ではない．そのため，ECMO 臨床試験に沿って 2×2 表の推測に議論を限定する．潜在結果変数モデルの起源は，Neyman (1923) とされており，初期の実験計画法のテキストにアイデアの一部が取り入れられている（Fisher 1935, Kempthorne 1952, Cox 1958）．Rubin (1974) は，これらの議論を，潜在結果変数，割付けメカニズム（assignment mechanism），確率分布という 3 つの概念で整理した．Rosenbaum and Rubin (1983) のプロペンシティスコアに関する論文は，観察研究であっても割付けメカニズムの理解が統計解析を行う上で重要ということを指摘した．本論文は，この流れを踏まえて，以下のように構成されている．まず，2 節では ECMO 臨床試験について，割付けメカニズムの違いを中心に記述する．3～5 節では，潜在結果変数，割付けメカニズム，統計的推測に関連する概念（表 1 に一覧を示す）を導入する．6～7 節では，それぞれミシガン試験とハーバード試験を取り上げて，因果推論の基礎概念を具体的に解説する．8 節では ECMO 臨床試験を踏まえて因果効果の推測方式について考察し，9 節で結語を述べて終える．本論文の議論は Rubin (1974) や Imbens and Rubin (2015) のテキストに従ったが，因果推論にはいくつかの学派があり，基礎となる考え方がしばしば異なる．必要に応じて，黒木・小林 (2011)，Greenland and Brumback (2002) のレビューや，Pearl (2009)，VanderWeele (2015) のテキストを参照してほしい．

2. 事 例

2.1 ECMO 臨床試験

新生児遷延性肺高血圧症は，胎児期における肺高血圧の状態が持続する呼吸循環障害の 1 つである．1980 年頃の人工呼吸器による治療成績は，死亡率 80 % 前後であった．ECMO とは，人工肺とポンプを用いた体外循環により，呼吸循環機能の回復まで酸素を供給することにより，重症呼吸不全患者または重症心不全患者に対して行われる生命維持法である．1982 年（Bartlett, et al. 1985），1986 年（O'Rourke, et al. 1989），1993 年（UK Collaborative ECMO Trial Group 1996）に，

表 1. 基礎概念の一覧

潜在結果変数 (potential outcomes)	割付けメカニズム (assignment mechanism)	統計的推測 (statistical inference)	その他
因果効果の分類	割付けメカニズムの分類	ランダムサンプリング	交絡 (confounding) の定義
解析単位レベルの因果効果 (unit-level causal effects)	無視可能 (ignorable)	モデルに基づく頻度論の推測 (model-based frequentist inference)	比較可能性 (comparability, exchangeability)
集団レベルの因果効果 (population-level causal effects)	非交絡 (unconfounded, strongly ignorable)	Bayes 流の推測 (Bayesian inference)	交絡割付けメカニズム (confounded assignment mechanism)
効果の均一性 (between-unit homogeneity of effects)	無視不可能 (nonignorable)	推定方程式を用いる推測 (inference using estimating equations)	併合可能性 (collapsibility)
効果の修飾 (effect modification)	確率的割付け (probabilistic assignment)	無限母集団因果リスク差 (super- population causal risk difference)	
	オーバーラップ (overlap)		
仮定	プロペンシテイスコア (propensity score)	ランダム化 ランダム化に基づく推測 (randomization-based inference)	変数の分類 共変量 (covariate)
Multiple Versions of Treatment		有限集団因果リスク差 (finite- population causal risk difference)	割付因子 (allocation factor)
非干渉性 (no interference between unit)		Fisher 帰無仮説 (Fisher null hypothesis, sharp null hypothesis)	予後因子 (prognostic factor)
Stable Unit Treatment Value Assumption		Neyman 帰無仮説 (Neyman hypothesis)	効果修飾因子 (effect modifier)
			予測マーカー (predictive marker)
			交絡因子 (confounder)
			合流点因子 (collider)
			中間因子 (intermediate)

新生児遷延性肺高血圧症を対象に、ECMO と従来療法の死亡割合を比較する臨床試験 3 件が行われた。試験実施施設は、それぞれミシガン大学、ハーバード大学、英国 5 施設であり、医療環境、適格基準、治療計画は少しずつ異なる。これらの試験を実施する上で問題になったのは、ECMO による成績がかなりよいと一部の医師は考えていたため、ECMO と従来療法をランダムに割付けることが非倫理的ではないかということだった (Ware 1989)。

2.2 ミシガン試験の割付けメカニズム

既に治療を受けた対象者のアウトカムに基づいて、割付けのアルゴリズムを変更することをアウトカム適応的ランダム化という。1982 年のミシガン試験 (Bartlett, et al. 1985) で割付け確率を決めるために用いられたのは、ランダム化プレイザウィナールール (Wei and Durham 1978, Wei, et al. 1990) である。このアルゴリズムは、同数の白と黒のボールが入っている壺に例えられる。はじめに壺には白と黒のボールが 1 個ずつ入っている。壺からボールを 1 つ取り出して、白だったら試験治療群に割付け、黒だったらコントロール群に割付け、ボールを壺に戻す。試験治療群で良い結果が出たか、コントロール群で悪い結果が出た場合には、壺に白いボールを入れる。そうすると、壺には白いボール 2 個と黒いボール 1 個が入っているから、白いボールが出て試験治療群に割付けられる確率は $1/2$ から $2/3$ になる。コントロール群で良い結果が出たか、試験治療群で悪い結果が出た場合には、逆に壺に黒いボールを入れる。このように、対象者のアウトカムが評価される度に、壺の中の白と黒のボールの割合が更新され、次の対象者が成績のよい治療を受ける確率を高めるのが、このルールである。

2.3 ハーバード試験の割付けメカニズム

その後ハーバード大学で行われた試験 (O'Rourke, et al. 1989) では、個人ではなく集団単位で

割付け確率を変える 2 ステージデザインが採用された (Ware 1989)。これは、ランダム化を行う第 1 ステージと、それまでの成績のよい治療を強制的に割付ける第 2 ステージから構成される。第 1 ステージでは、ブロックサイズ 4 の置換ブロック法により、4 人ごとに均等に試験治療群とコントロール群を割付ける。そして、一方の治療で 4 件の死亡が観察されたとき、第 1 ステージを終了する。第 2 ステージでは、4 件の死亡が観察されるか優越性を結論できるまで、成績のよい方の治療を用いて試験を継続する。

ハーバード試験のデザインは、頻度論の立場によるものである。主たる解析には、第 1・第 2 ステージを併合した正確検定が採用された。デザインの根拠を得るため、過去の遷延性肺高血圧症のある新生児についてチャートレビューが行われ、従来療法の死亡割合は 13 人中 11 人という事前情報が得られた。これに基づき、死亡確率 20% 対 80% という対立仮説を用いて、片側有意水準 5% の下で検出力 74% が得られるように、デザイン（たとえば死亡 4 件という第 2 ステージの開始基準）が決められた。

2.4 英国試験の割付けメカニズム

ミシガン試験・ハーバード試験の結果は ECMO の有効性を示していたが、コントロール群の人数が少なすぎるなどの批判があり、医師間でコンセンサスには至らなかった。その後、英国で行われた大規模ランダム化臨床試験 (UK Collaborative ECMO Trial Group 1996) では、原疾患、重症度、施設を用いた最小化法が採用された。

3. 潜在結果変数

3.1 因果リスク差の定義

対象者 i について観察されるアウトカムを、0 または 1 の値をとる 2 値変数 Y_i で表す。また、従来療法と ECMO に対応する潜在結果変数を、それぞれ Y_i^0 と Y_i^1 と定義する。本来なら潜在結果変数モデルで知りたいのは、個人 i の因果効果

$$\tau_i = Y_i^1 - Y_i^0$$

である。これは、同一個人が従来療法を受けたときと ECMO を受けたときの差を表している。現実の臨床試験では、ある個人について、両方の潜在結果変数を同時に観察することはできない。そこで、 N 人からなる有限集団を考える。対象者全員の潜在結果変数は、

$$Y_i^0 = Y_i^1 = 0$$

$$Y_i^0 = 0, Y_i^1 = 1$$

$$Y_i^0 = 1, Y_i^1 = 0$$

$$Y_i^0 = Y_i^1 = 1$$

の 4 パターンに分類できる。これらのパターンの人数を、それぞれ $n_{00}, n_{01}, n_{10}, n_{11}$ で表す。対象者全員に ECMO を用いたときと従来療法を用いたときの比較は、有限集団因果リスク差 (finite-

population causal risk difference)

$$\bar{\tau} = \frac{n_{01} - n_{10}}{N} = \frac{\sum_{i=1}^N Y_i^1}{N} - \frac{\sum_{i=1}^N Y_i^0}{N}$$

により行うこととする。通常のリスク差と異なり、分母が対象者全員である点に注意してほしい。

さらに区別が必要なのは、有限集団と無限母集団の違いである。 N 人全員の潜在結果変数の組 $\{Y_i^0, Y_i^1; i=1, \dots, N\}$ を無限母集団からのランダムサンプルとみなすことがある。そのときの τ_i の期待値は、 $\Pr(Y_i^0=1)=\pi^0$ と $\Pr(Y_i^1=1)=\pi^1$ の差になる。このように定義された

$$\tau = \pi^1 - \pi^0$$

を母集団因果リスク差 (super-population causal risk difference) という。本論文では因果リスク差を主に扱うが、それは話が簡単になるからであって必然ではない。因果リスク比 π^1/π^0 や因果オッズ比 $\{\pi^1/(1-\pi^1)\}/\{\pi^0/(1-\pi^0)\}$ など、他の指標が用いられることもある。

3.2 因果リスク差の推定

次に、ECMO 臨床試験で観察されるデータと潜在結果変数がどのように対応しているかを述べながら記法を整理しよう。患者 i に ECMO を用いたかどうかを A_i ($A_i=1$ なら ECMO, $A_i=0$ なら従来療法)、アウトカムを Y_i ($Y_i=1$ なら死亡, $Y_i=0$ なら生存) で表す。また、治療前に測定された共変量 (covariate) を L_i で表す。3つの試験のデータは、いずれも表2のような 2×2 表に要約できる (簡単のため L_i は省略した)。 S_1, S_0, S はそれぞれ試験治療群、コントロール群、全体の死亡数である。

表2のデータが得られれば、有限集団または母集団因果リスク差は、2群の死亡割合の差

$$\bar{\tau} = \frac{S_1}{N_1} - \frac{S_0}{N_0}$$

により推定できる。

3.1 節では、潜在結果変数は4パターンに分類できると述べた。各パターンの試験治療群の人数を、それぞれ $x_{00}, x_{10}, x_{01}, x_{11}$ で表すと、表3のようにまとめることができる。表3は、 N 組の Y_i^0 と Y_i^1 から構成される有限集団のパターンを表している。つまり、各パターンの人数 $n_{00}, n_{01}, n_{10}, n_{11}$ がわかれば (もちろん $n_{00} + n_{01} + n_{10} + n_{11} = N$ という制約がある)、この有限集団で因果効果があるかが判断できるから、これが推測の対象 (未知パラメータ) である。一方で、治療をランダムに割付けるとき、 $x_{00}, x_{10}, x_{01}, x_{11}$ は確率変数とみなすことができる。

ランダム割付けを仮想的に反復すると、ある個人がどちらの治療に割付けられるかによって、 $x_{00}, x_{10}, x_{01}, x_{11}$ の値がばらつく。観察データ (表2) でいえば、割付け結果によって、 $S_1 = x_{01} + x_{11}$

表2. 観察データの記法

	試験治療 ($A=1$)	コントロール ($A=0$)	合計
生存 ($Y=0$)	$N_1 - S_1$	$N_0 - S_0$	
死亡 ($Y=1$)	S_1	S_0	S
合計	N_1	N_0	N

表 3. 潜在結果変数の記法

	試験治療 ($A=1$)	コントロール ($A=0$)	合計
常に生存 ($Y_i^0=Y_i^1=0$)	x_{00}	$n_{00}-x_{00}$	n_{00}
試験治療でのみ死亡 ($Y_i^0=0, Y_i^1=1$)	x_{01}	$n_{01}-x_{01}$	n_{01}
コントロール治療でのみ死亡 ($Y_i^0=1, Y_i^1=0$)	x_{10}	$n_{10}-x_{10}$	n_{10}
常に死亡 ($Y_i^0=Y_i^1=1$)	x_{11}	$n_{11}-x_{11}$	n_{11}
合計	N_1	N_0	N

* 観察される死亡数とは、 $S_1=x_{01}+x_{11}$, $S_0=n_{10}-x_{10}+n_{11}-x_{11}$ という関係があるだけで、すべての人数を識別することはできない

や $S_0=n_{10}-x_{10}+n_{11}-x_{11}$ が確率的に変動する。ランダム化に基づく推測では、これらの確率的変動が、因果リスク差を推定するときの誤差の源になる。

3.3 一致性, 非干渉性, 効果の均一性

死亡割合の差によって因果リスク差を推定できると述べたが、そこにはいくつかの仮定が隠れている。そこで、対象者 i と j の 2 人がいて、それぞれ ECMO と従来療法により治療されたという単純化した状況で、ここまでの議論を振り返ってみよう。両者の比較 Y_i-Y_j が因果効果としてなんらかの意味を持つための条件はなんだろうか。結論から言うと、(後述する割付けメカニズムに関する仮定に加えて)、一致性 (consistency), 非干渉性 (no interference between unit), 比較可能性 (comparability), 効果の均一性 (between-unit homogeneity of effects) が必要になる (Rubin 1978)。

一致性とは、観察されるデータと潜在結果変数が、 $Y_i=(1-A_i)Y_i^0+A_iY_i^1$ という関係でリンクしていることを意味する。わかりやすくいうと、治療 $A=0$ を受けると Y_i^0 が観察され、 $A=1$ だと Y_i^1 が観察されるという反事実の世界と現実世界を結ぶ仮定である。このとき、 $Y_i-Y_j=Y_i^1-Y_j^0$ のように、対象者 i と j の 2 人を比べることで、潜在結果変数の比較を行うことができる。それでは、どのようなときに一致性が成り立たないのだろうか。それはたとえば、 K 人の術者がいて、ECMO または従来療法の成績が異なるときである。このとき、仮に対象者 i が術者 k ($k=1,\dots,K$) により治療 $A_i=a$ を受けたときの潜在結果変数を定義するには、 Y_i^{ak} のように術者と治療の組み合わせを表現する必要が生じる (VanderWeele and Hernán 2013)。ここから明らかのように、同じ治療 $A_i=a$ であっても別のバージョン ($k=1,\dots,K$) が存在するとき (これを multiple versions of treatment という)、 $Y_i=(1-A_i)Y_i^0+A_iY_i^1$ という単純な関係にはならない。

次に非干渉性とは、他の解析単位の治療が、別の解析単位に影響しないという仮定である。逆に影響があることを干渉 (interference) と呼んでいる (Cox 1958)。仮に対象者 i が治療 $A_i=a$ を、対象者 j が治療 $A_j=b$ を受けたときの、対象者 i のアウトカムを Y_i^{ab} と定義する。これが、対象者 j の治療が、対象者 i に影響を干渉するときのモデルである。このモデルが意味を持つのは、たとえば対象者 i と j の ECMO を行う術者が同じで、同じ治療を続けて行くと技術が向上する状況である。なぜなら、別の治療を行ったとき ($A_i=a \neq A_j=b$) のアウトカム Y_i^{ab} よりも、

Y_i^{aa} の方がよいかもしれないからである。別の例として、同じ地域内に住む対象者間の感染や、1人の対象者の治療部位（たとえば右目と左目）が挙げられる。このように、干渉は一定の範囲内に起こることが多い。そのとき、治療 A_i だけによる効果と、干渉のある範囲内に共通に起こる効果を、区別する必要がある。非干渉性はこの問題を避けるための仮定といえる。

一致性と非干渉性は、合わせて Stable Unit Treatment Value Assumption (SUTVA) と呼ばれることもあるし、別の仮定が論じられることもある（黒木・小林 2011, VanderWeele and Hernán 2013）。いずれにしても、これらの前提を置けば、潜在結果変数を、 $\{Y_i^0, Y_i^1, Y_j^0, Y_j^1\}$ と通常の記法で表すことができる。そして、一致性により Y_i^0 と Y_i^1 のうちどちらかは観察されることになり、もう一方は反事実的結果であり欠測データとみなされる。

一致性と非干渉性が成り立っていたとしても、 $Y_i - Y_j$ という比較は、必ずしも因果効果 ($\tau_i = Y_i^1 - Y_i^0$ または $\tau_j = Y_j^1 - Y_j^0$) とは一致しない。逆に一致するための条件を考えると、 $Y_i^0 = Y_j^0$ と $\tau_i = \tau_j$ が成り立っていればよいことがわかる。前者は、後で詳しく述べる比較可能性に対応しており、対象者 i と j の疾患の状態に差がないことを意味する。後者は、解析単位間で因果効果が等しい状況であり、これを効果の均一性と呼ぶ。効果の均一性は、 Y_i^0 と Y_i^1 の相関が 1 ということを意味するから、強い仮定である。

3.4 Fisher 帰無仮説と Neyman 帰無仮説

有限集団因果リスク差 $\bar{\tau}$ について仮説検定を行うとき、2種類の帰無仮説を考えることができる。有限集団因果リスク差がゼロであれば、ECMO と従来療法に差はないと考えてよい。ただし、効果の均一性は必ずしも仮定しなければならないわけではない。これを表したものを

Neyman 帰無仮説: $\bar{\tau} = 0$

と呼ぶ (Neyman 1923)。これは、 $n_{10} = n_{01}$ と同じ意味である。Neyman 帰無仮説に対して Fisher 正確検定は、治療に効果はなく、ECMO と従来療法で、アウトカム Y は常に変わらないという仮説を調べるものである (Fisher 1935)。これを

Fisher 帰無仮説: $Y_i^0 = Y_i^1$

または強い帰無仮説 (sharp null hypothesis) という。これは、 $n_{10} = n_{01} = 0$ を意味しており、Neyman 帰無仮説よりも強い仮定である。

3.5 共変量

ランダム化臨床試験と観察研究のいずれの場合も、事前知識を参考にして、疾患との関係が深い因子を、共変量として測定するのが普通である。このような因子には2種類が考えられる。1つ目は Y_i と相関のあるもので、これを予後因子 (prognostic factor) という。2つ目は $\tau_i = Y_i^1 - Y_i^0$ と相関するもので、こちらは効果修飾因子 (effect modifier) という。効果の修飾 (effect modification) とは、共変量 $L_i = 0$ と $L_i = 1$ の2つのサブグループがあるときを例にすると、各サブグループの有限集団因果リスク差の差

$$\left(\frac{\sum_{i:L=1} Y_i^1}{\sum_{i:L=1} L_i} - \frac{\sum_{i:L=1} Y_i^0}{\sum_{i:L=1} L_i} \right) - \left(\frac{\sum_{i:L=0} Y_i^1}{\sum_{i:L=0} (1-L_i)} - \frac{\sum_{i:L=0} Y_i^0}{\sum_{i:L=0} (1-L_i)} \right)$$

がゼロかどうかによって定義される。臨床試験では、効果修飾因子は臨床検査や画像などのバイオマーカーであることが多いため、予測マーカー (predictive marker) という用語を同じ意味で用いることも多い。

共変量の用途として、まず考えられるのは、割付因子 (allocation factor) である。英国試験の割付けメカニズムは、原疾患、重症度、施設という3つの割付因子により規定された。もう1つの用途は、統計解析における調整因子である。観察研究では、前述のプロペンシティスコアなどを用いて、共変量の調整が行われる。ランダム化臨床試験であっても、共変量が予後因子であれば、利用した方が無視して解析したときに比べて、検出力が向上する。更に、統計解析では、共変量によりサブグループ解析を行い、効果修飾因子を探索することも多い。効果の修飾があるなら、それを予測マーカーとして用いることで治療がよく効くサブグループを特定できるため、臨床的に有用である。

4. 割付けメカニズム

4.1 割付けメカニズムの分類

対象者全体が、比較するどの治療を受けるかを決める規則を、割付けメカニズムという。臨床試験で用いられる割付けメカニズムでもっとも単純なのは、個人ごとに一定の確率でランダムに割付けるものである。このとき、割付け確率は潜在結果変数 $\{Y_i^0, Y_i^1\}$ と共変量 L_i に依存しない。この単純ランダム化は

$$\Pr(A_1, A_2, \dots, A_N | Y_1^0, Y_1^1, L_1, Y_2^0, Y_2^1, L_2, \dots, Y_N^0, Y_N^1, L_N) = \prod_{i=1}^N \Pr(A_i)$$

と表すことができる。この場合には、後述するように共変量の偏りなどの問題は、期待値の上で生じない。

Imbens and Rubin (2015) によるテキストは、いくつかの割付けメカニズムを取り上げているが、それを無視可能 (ignorable)・非交絡 (unconfounded) という概念を用いて分類している。無視可能とは、割付け確率が観察データだけに依存することと定義される (これは欠測メカニズムにおける無視可能と同義である)。

$$\Pr(A_1, A_2, \dots, A_N | Y_1^0, Y_1^1, L_1, Y_2^0, Y_2^1, L_2, \dots, Y_N^0, Y_N^1, L_N) = \Pr(A_1, A_2, \dots, A_N | Y_1, L_1, Y_2, L_2, \dots, Y_N, L_N)$$

非交絡とは、(Rosenbaum and Rubin 1983 でいう) 強い無視可能 (strongly ignorable) とほぼ同じ意味であり、割付け確率が共変量だけに依存することを指す。強い無視可能割付けメカニズムは、割付けが個人ごとに独立になされるとき、

$$\Pr(A_1, A_2, \dots, A_N | Y_1^0, Y_1^1, L_1, Y_2^0, Y_2^1, L_2, \dots, Y_N^0, Y_N^1, L_N) = \prod_{i=1}^N \Pr(A_i | L_i)$$

と表すことができる。 $e(l) = \Pr(A_i = 1 | L_i = l)$ をプロペンシティスコアという (Rosenbaum and Rubin 1983)。

最近注目されているアウトカム適応的ランダム化は無視可能に分類されるが、アウトカムに依存しているため、因果効果の統計的推測について、単純ランダム化との違いが生じる。多くの観

観察研究は、非交絡を仮定している。また、割付けメカニズムの具体例は、6～7節で取り上げている。

4.2 プロペンシティスコア

観察研究では、単純ランダム化のように治療と共変量が独立ではないため、治療群間で共変量の分布が異なるのが普通である。この現象を疫学では交絡といい、その一例が後で詳しく扱うハーバード試験である。プロペンシティスコアは、共変量の偏りを調整して群間を比較可能にするため、観察研究で広く用いられている。個々の対象者のプロペンシティスコアの値は、ランダム化臨床試験では既知であるが、観察研究ではわからない。そのため、ロジスティック回帰などを用いて推定される。また、プロペンシティスコアは既知であっても、推定された値を用いた方が、因果効果の推定効率がよくなるという性質もある (Imbens 2004, Henmi and Eguchi (2004)。

プロペンシティスコアを用いることで、交絡を完全に調整できるための条件は2つある (Rosenbaum and Rubin 1983)。1つ目は、強い無視可能性 (非交絡性) であり、潜在結果変数 $\{Y_i^0, Y_i^1\}$ と治療 A_i は、共変量 L_i により条件付けた下で、独立と定義される。ここで、 $A \perp\!\!\!\perp B | C$ は、確率変数 A と B は C で条件付けた下で独立であることを意味する記号である。

$$\{Y_i^0, Y_i^1\} \perp\!\!\!\perp A_i | L_i$$

2つ目は、確率的に割付けが行われるという条件である。

$$0 < e(l) < 1$$

このことをオーバーラップ (overlap) や正值性 (positivity) ともいうこともある。

プロペンシティスコアの有用な性質は、この2つの仮定の下で

$$\{Y_i^0, Y_i^1\} \perp\!\!\!\perp A_i | e(l)$$

が成り立つということである (Appendix A に証明を示す)。ここで、 $e(l)$ の部分は、共変量 L_i から計算されたプロペンシティスコアが、同じ値の集団において独立性が成り立つという意味である。また、プロペンシティスコアには、Rosenbaum and Rubin (1983) の意味で、もっとも粗い (coarsest) バランススコアという性質もある。

英国試験では、(最小化法なので厳密には異なるが) プロペンシティスコアは対象者を通じて $e=0.5$ だったと考えられるため、ECMO 群と従来治療群の全体を比較すれば、因果効果をバイアスなく推定できる。逆にいうと、ランダム化臨床試験であったとしても、割付け確率の異なる層別ランダム化を行ったとしたら、それを無視して層を併合して解析してはならない。プロペンシティスコアの値が等しい集団を比べることにならず、人為的な交絡が生じ得るからである。ハーバード試験では、第1ステージと第2ステージでプロペンシティスコアは異なると考えられ、ステージ別に解析が行われた (7節)。プロペンシティスコアに基づく統計手法 (プロペンシティスコアによる重みづけ、層別、回帰、マッチング) は、観察研究で広く用いられている。これらの手法や理論的性質のレビューが、Imbens (2004)、佐藤・松山 (2011) によって与えられている。

5. 統計的推測

統計学には、頻度論と Bayesian の 2 つの流派がある。また、ランダム化臨床試験の解析は、ランダム化に基づく推測とモデルに基づく推測に大別される。しかしながら、 2×2 表のような単純な場合、似たような推定量やその誤差分布を導くことが多い。それは潜在結果変数モデルを考えるとときも同じである。

表 4 に 1996 年の英国試験のデータを示す。この試験では、一般的なりスク差の推定量

$$\hat{\tau} = \hat{\pi}^1 - \hat{\pi}^0 = \frac{S_1}{N_1} - \frac{S_0}{N_0}$$

とその 95% 信頼区間

$$95\%CI = \hat{\tau} \pm 1.96 \sqrt{\frac{\hat{\pi}^1(1-\hat{\pi}^1)}{N_1} + \frac{\hat{\pi}^0(1-\hat{\pi}^0)}{N_0}}$$

そして χ^2 検定と Fisher の正確検定が用いられた (UK Collaborative ECMO Trial Group 1996)。 χ^2 検定、Fisher の正確検定の結果はともに $p < 0.01$ であった。

これらの手法は、観察されるデータの確率分布 $\Pr(Y_i | A_i)$ に、積 2 項分布を仮定することで導かれる。潜在結果変数の枠組みであっても、モデルを仮定するか (5.1~5.2 節)、割付けが並べ替え分布により行われたとき (5.4~5.5 節)、これらの手法は正当化できる (実際の割付けは最小化法である)。さらに、有限集団因果リスク差と母集団因果リスク差のどちらの場合も、同じように計算された推定量が、一致推定量になる。しかし、以下で論じるように、 $\hat{\tau}$ の誤差分布は、無限母集団と有限集団で異なる。これは、前者はモデル (ランダムサンプリング) に基づいて推測を行ったときの誤差を表しているのに対して、後者はランダム化に基づくバラツキを反映しているためである。有限集団を考えたとき、Fisher 帰無仮説以外の状況では、分散

$$\text{Var}(\hat{\tau}) = \frac{\pi^1(1-\pi^1)}{N_1} + \frac{\pi^0(1-\pi^0)}{N_0}$$

には大標本であっても過大評価の方向にバイアスがあり、そのため上の信頼区間は保守的になる。Neyman (1923) は、連続データの場合も類似した結果が導かれることを示した (Appendix B)。

5.1 モデルに基づく頻度論の推測

まず、潜在結果変数にパラメトリックモデルを仮定すると、 2×2 表の積 2 項分布モデルに帰着することを示そう。 N 人の対象者は、無限母集団からのランダムサンプルと仮定する。そして、 Y_i^0 と Y_i^1 の母集団における確率分布に、以下のような独立な Bernoulli 分布を仮定する。

表 4. 英国試験のデータ

	ECMO ($A=1$)	従来療法 ($A=0$)	リスク差 (95%信頼区間)
生存 ($Y=0$)	65	38	
死亡 ($Y=1$)	28 (30%)	54 (59%)	-29% (-42~-15%)
合計	93	92	

$$\Pr(Y_i^0=y, Y_i^1=v; \pi^0, \pi^1) = \binom{1}{y} (\pi^0)^y (1-\pi^0)^{1-y} \binom{1}{v} (\pi^1)^v (1-\pi^1)^{1-v}$$

Y_i^0 と Y_i^1 は独立と仮定するのは、 Y_i^0 と Y_i^1 の相関はデータから識別不能だからである。このとき尤度関数は単純ランダム化の下で

$$\begin{aligned} \text{Likelihood}(\pi^0, \pi^1) &= \prod_{i=1}^N \Pr(Y_i | A_i; \pi^0, \pi^1) \Pr(A_i) \\ &\propto \prod_{i: A_i=0} \Pr(Y_i | A_i=0; \pi^0) \prod_{i: A_i=1} \Pr(Y_i | A_i=1; \pi^1) \\ &= \prod_{i: A_i=0} \Pr(Y_i^0; \pi^0) \prod_{i: A_i=1} \Pr(Y_i^1; \pi^1) \\ &\propto (\pi^0)^{S_0} (1-\pi^0)^{N_0-S_0} (\pi^1)^{S_1} (1-\pi^1)^{N_1-S_1} \end{aligned}$$

ここで第2式から第3式は、潜在結果変数と治療の独立性を利用して導かれる。これは積2項分布モデルと同じ尤度関数であるから、リスク差 $\tau = \pi^1 - \pi^0$ に最尤法を適用することで、 2×2 表のリスク差と95%信頼区間の公式が導かれる。

Y_i^0 と Y_i^1 の独立性について補足する。データは Y_i^0 と Y_i^1 の相関に関する情報を持たないため、尤度関数は真の分布によらず上の式になる。両者の相関関係について、単調性 $Y_i^0 < Y_i^1$ という仮定を置くことで、 Y_i^0 と Y_i^1 の両方に基づく層別 (principal stratification) を行うことがある (Frangakis and Rubin 2002)。また、Imbens and Rubin (2015) の8.6章には、Bayes 流の推測において、事前分布を通じて相関に関する情報を与えると、事後分布の分散が大きくなる数値例が示されている。

5.2 Bayes 流の推測

Y_i^0 と Y_i^1 にパラメトリックモデルを仮定して導かれた π^0 と π^1 の尤度関数は、Bayes 推測でも用いられる。単純ランダム化の下で、 π^0 と π^1 について無情報事前分布を用いるとき、上の95%信頼区間の公式は、 τ の事後分布の95%確率区間 (credible interval) と一致する。このときの95%という確率は、頻度確率 (たとえば割付けを繰り返したときの実現頻度の相対値) とは異なり、主観確率として解釈される。

Bayes 推測の特徴は、潜在結果変数のうち観察できない部分 $Y_i^{\text{mis}} = A_i Y_i^0 + (1-A_i) Y_i^1$ を欠測データとみなし、その事後予測分布を経由して、因果効果の事後分布を求める点である (Rubin 1978)。Imbens and Rubin (2015) に従って、無視可能割付けメカニズムの下で Y_i^{mis} の事後予測分布を導いてみよう。割付けメカニズムを含む確率分布を考えるため、 $\mathbf{Y}^0$ と $\mathbf{Y}^1$ に加え $\mathbf{A}$ の同時分布を考える。 Y_i^{mis} の事後予測分布は、 $\Pr(\mathbf{Y}^0, \mathbf{Y}^1, \mathbf{A} | \mathbf{L})$ のうち、観察できる部分で条件付けたものである。したがって、無視可能性の仮定の下で以下の結果を得ることができる。

$$\begin{aligned} \Pr(\mathbf{Y}^{\text{mis}} | \mathbf{Y}, \mathbf{A}, \mathbf{L}) &= \frac{\Pr(\mathbf{A} | \mathbf{Y}^0, \mathbf{Y}^1, \mathbf{L}) \Pr(\mathbf{Y}^0, \mathbf{Y}^1 | \mathbf{L})}{\mathbb{E}^{\mathbf{Y}^{\text{mis}}} \{\Pr(\mathbf{A} | \mathbf{Y}^0, \mathbf{Y}^1, \mathbf{L}) \Pr(\mathbf{Y}^0, \mathbf{Y}^1 | \mathbf{L})\}} \\ &= \frac{\Pr(\mathbf{A} | \mathbf{Y}, \mathbf{L}) \Pr(\mathbf{Y}^0, \mathbf{Y}^1 | \mathbf{L})}{\mathbb{E}^{\mathbf{Y}^{\text{mis}}} \{\Pr(\mathbf{A} | \mathbf{Y}, \mathbf{L}) \Pr(\mathbf{Y}^0, \mathbf{Y}^1 | \mathbf{L})\}} \\ &= \frac{\Pr(\mathbf{Y}^0, \mathbf{Y}^1 | \mathbf{L})}{\mathbb{E}^{\mathbf{Y}^{\text{mis}}} \{\Pr(\mathbf{Y}^0, \mathbf{Y}^1 | \mathbf{L})\}} \end{aligned}$$

ここで、 $\mathbb{E}^{\mathbf{Y}^{\text{mis}}}$ は、 $\mathbf{Y}^{\text{mis}}$ についての期待値を表す。第1式から第2式は無視可能性から、第2式か

ら第3式は $\Pr(A|Y, L)$ に Y^{mis} が含まれないことから導かれる. これらの式にはパラメータ π が明示的に含まれていないが, $\Pr(Y^0, Y^1|L, \pi)$ を π の事前分布 $f(\pi|L)$ にそって積分することで計算される. すなわち, $\Pr(Y^0, Y^1|L, \pi)$ と $f(\pi|L)$ を特定し, 観察データから事後予測分布を求めれば, 因果効果に欠測データ Y^{mis} が含まれていたとしても, その確率分布を計算することができる.

第3式の分母に含まれている Y^0 と Y^1 の条件付き同時分布は, 一般に

$$\Pr(Y^0, Y^1|L) = \int \prod_{i=1}^N \Pr(Y_i^0|L_i, \pi) \prod_{i=1}^N \Pr(Y_i^1|L_i, \pi) f(\pi|L) d\pi$$

と書ける. これに独立な Bernoulli 分布を仮定し, $f(\pi|L)$ に無情報事前分布を仮定すれば, 5.1 節と同じ τ の事後分布の 95% 確率区間が得られる.

この議論で重要なのは割付けメカニズムが無視可能かどうかである (Rubin 1978). 一般に, 因果効果の事後分布は, 割付けメカニズムが無視可能なとき, 観察データ $\{Y, A, L\}$, Y^0 と Y^1 の条件付き同時分布 $\Pr(Y^0, Y^1|L, \pi)$ とすべてのパラメータ π の事前分布 $f(\pi|L)$ だけで決まる. それに対して, 割付けメカニズムが無視不可能だと, 因果効果の事後分布に割付けメカニズムの部分が含まれるため, 具体的な確率分布を特定する必要が生じる. そしてそのとき, 観察された共変量の分布に群間で違いがなかったとしても, 割付けメカニズムを誤特定すると誤った推測を導きやすい.

5.3 g 推定

5.1~5.2 節では, 潜在結果変数のパラメトリックモデルを特定することで, 母集団因果リスク差 $\tau = E(Y_i^1 - Y_i^0)$ の推測を行った. より仮定の緩い別のアプローチとして, 平均構造 $\tau = E(Y_i^1 - Y_i^0)$ だけを特定する手法もある. これが g 推定である (Vansteelandt and Joffe 2014). 単純な場合には, パラメトリックモデルと g 推定で推定量は同じになるが, 仮定が違うことにより, 異なる 95% 信頼区間が導かれる.

理想的にランダム化されているなら, 潜在結果変数 $\{Y_i^0, Y_i^1\}$ と治療 A_i は独立であるはずである. そのとき, 潜在結果変数の期待値は, 両群で等しい.

$$E(Y_i^a|A_i=0) = E(Y_i^a|A_i=1)$$

g 推定とは, これを利用して因果パラメータの推定方程式を構成する手法である. 導出は Appendix C に示すが, 母集団因果リスク差の推定方程式は,

$$H_i(\tau) = Y_i - \tau A_i$$

と

$$U_i(\tau) = \{A_i - E(A_i)\} [H_i(\tau) - E\{H_i(\tau)\}]$$

を用いて

$$\sum_{i=1}^N U_i(\tau) = 0.$$

と書ける．ここで， $E(A_i)$ と $E\{H(\tau)\}$ は未知なので，実際にはそれぞれに対応する標本平均に置き換えたものを解くことになる．この推定方程式の解は，他の推測アプローチと同様に， $\hat{\tau} = S_1/N_1 - S_0/N_0$ になる． τ の真値 τ^* の下で， $E\{U_i(\tau^*)\} = 0$ であることを用いれば，M 推定の理論から，95% 信頼区間は

$$95\%CI = \hat{\tau} \pm 1.96 \sqrt{\frac{\text{Var}\{U_i(\hat{\tau})\}}{E\{\partial U_i(\tau)/\partial \tau | \tau = \tau^*\}^2}}$$

となることが示される．この結果は，推定関数 $U_i(\tau)$ が独立でなければ成り立たない．つまり，アウトカムや割付けは個人間で独立でなければならない．

英国試験において，g 推定による母集団因果リスク差は -29% (95% 信頼区間 -50~-7%) であった．この 95% 信頼区間は，モデルに基づく推測 (-42~-15%) よりも広い．これは，g 推定は，平均構造だけを特定しており，(確率分布全体を特定する) パラメトリックモデルよりも仮定が緩いためである．一般に，セミパラメトリックモデルの分散下限 (セミパラメトリック効率限界) は，仮定を満たす最も不利なパラメトリックモデルの Cramér-Rao の下限によって決まる．

5.4 Fisher 帰無仮説の推測

これまでは無限母集団を表すパラメトリックモデルの下での議論だった．次に，有限集団因果リスク差 $\hat{\tau}$ の推測について考えよう． N 人の対象者からなるランダム化臨床試験を行い，試験治療群 (N_1 人) とコントロール治療群 (N_0 人) を比較する状況を考える．このとき，割付けメカニズムは既知なのだから，ランダムサンプリングモデルではなく，割付けの確率分布に基づく推測とする方が自然である． N 人のうち， N_1 人が試験治療群に割付けられる組み合わせの数は， $\binom{N}{N_1}$ である．Fisher 正確検定 (Fisher 1935) は，これが等しく起きるような割付けの確率分布を用いる (これを並び替え分布という)．ここでは，表 2・表 3 に示されたデータのうち， $n_{00}, n_{01}, n_{10}, n_{11}$ は未知パラメータであり，割付けパターンが確率変数である．

このとき，試験治療群で観察される死亡数は，割付けパターンにより決まる確率変数である．そこで，この確率変数を S_1 ，その実現値を s_1 と書いて区別する．Fisher 帰無仮説 $Y_i^0 = Y_i^1$ の下で，両群合わせた死亡数は $S = S_0 + S_1$ という固定値とみなすことができる．また， N 人のうち， N_1 人が試験治療群に割付けられる割付けパターンだけを考えていた．したがって，表 2 に示した 2×2 表の確率分布は，周辺度数 (S と N_1) を条件付けた超幾何分布に従う．具体的には， S_1 の確率分布は，

$$\Pr(S_1 = s_1) = \frac{\binom{S}{s_1} \binom{N-S}{N_1-s_1}}{\binom{N}{N_1}}$$

と書くことができる．Fisher の正確検定の片側 p 値は，ここから裾側確率を求めて

$$p = \sum_{i=s_1}^{\min(N_1, S)} \frac{\binom{S}{i} \binom{N-S}{N_1-i}}{\binom{N}{N_1}}$$

である。Fisher の正確検定と χ^2 検定は、漸近的に一致することが知られている。モデルに基づく 95 % 信頼区間の公式

$$95\%CI = \hat{\tau} \pm 1.96 \sqrt{\frac{\hat{\pi}^1(1-\hat{\pi}^1)}{N_1} + \frac{\hat{\pi}^0(1-\hat{\pi}^0)}{N_0}}$$

は、 χ^2 検定と対応しているから、有限集団の場合にも、Fisher 帰無仮説の下に限り被覆確率が大標本で 95 % に近づくことを予想させる。

5.5 Neyman 帰無仮説や対立仮説の下での推測

対立仮説の下での $\hat{\tau}$ の誤差分布も、同じ考え方で導かれる (Copas 1973)。表 3 に示したように、対立仮説の下では潜在結果変数には 4 パターンがある。各パターンの試験治療群の人数 x_{00} , x_{10} , x_{01} , x_{11} の同時分布は、割付け分布に並べ替え分布を考えたとき、多変量超幾何分布

$$\Pr(x_{00}, x_{01}, x_{10}, x_{11}) = \frac{\prod_{a=0}^1 \prod_{b=0}^1 \binom{n_{ab}}{x_{ab}}}{\binom{N}{N_1}}$$

に従う。ただし、周辺度数が観察されていることにより $x_{00} + x_{01} + x_{10} + x_{11} = N_1$, $x_{01} + x_{11} = S_1$, $n_{11} - x_{11} + n_{10} - x_{10} = S_0$ という制約がある。

この確率分布の下で、因果リスク差 $\hat{\tau} = S_1/N_1 - S_0/N_0$ の分散は

$$\text{Var}(\hat{\tau}; \beta) = \frac{N}{N-1} \left\{ \frac{\pi^1(1-\pi^1)}{N_1} + \frac{\pi^0(1-\pi^0)}{N_0} - \frac{\beta}{N} \right\}$$

となる (Appendix D)。この式に含まれる

$$\beta = \frac{n_{01} + n_{10}}{N}$$

は、いずれかの治療によりアウトカムが変わる対象者の割合である。 $\beta=0$ のとき、Fisher 帰無仮説と Neyman 帰無仮説は等しくなり、 N が大きくなるにつれ、 $\text{Var}(\hat{\tau}; \beta)$ は一般的な分散の公式に近づく。ただし、両者の差は、 $\beta \neq 0$ のときには漸近的にゼロにならない。さらに重要なことは、このパラメータは表 2 のデータから識別できず、値をとる範囲

$$|\pi^1 - \pi^0| \leq \beta \leq \min(\pi^0 + \pi^1, 2 - \pi^0 - \pi^1)$$

だけがわかるということである。つまり、 β について制約を置かなければ、因果リスク差 $\hat{\tau}$ の漸近分散を求めることはできない。ただし、幸いなことに β は常に正である。そのため、95 % 信頼区間の公式は、Neyman 帰無仮説の下で被覆確率が 95 % より大きいという意味で、保守的である。リスク差以外の指標でも同様の結果が導かれる (Robins 1988, Ding and Dasgupta 2016)。ま

た、上記の議論は漸近的な結果であるが、並べ替え分布に基づいて正確な信頼区間、p 値、サンプルサイズを計算することもできる。具体的な手順は Chiba (2017) に与えられている。

英国試験の漸近分散の下限は、 $\text{Var}(\hat{\tau}; \beta = \pi^0 + \pi^1) = 0.000098$ であり、一般的な $\text{Var}(\hat{\tau}) = 0.0049$ よりもずっと小さかった。これは極端な状況を想定した分散の計算である。しかし、この事例からは、一般的な 95% 信頼区間は、Neyman 帰無仮説や対立仮説の下で、過度に保守的になることがあるということも示唆される。

6. ミシガン試験

6.1 割付けメカニズムの分類

ミシガン試験のランダム化プレイザウィナルールにおいて、 $i > 1$ 番目以降の割付け確率は

$$\Pr(A_i = 1 | A_1, Y_1, \dots, A_{i-1}, Y_{i-1}) = \frac{1 + \sum_{j=1}^{i-1} \{(1 - A_j) Y_j + A_j (1 - Y_j)\}}{i + 1}$$

となる（死亡すなわち $Y_i = 1$ が悪い結果のため）。これは、割付け確率が観察データだけで決まるため、無視可能割付けメカニズムに分類される。しかし、強い無視可能（非交絡）ではない。割付け確率が共変量だけで決まらず、アウトカム Y に依存しているからである。

表 5 にミシガン試験で実際に用いられた割付け確率を示す。既に述べたように、この試験では実現した割付けパターン以外でどのような割付け確率が生じるかはわからない。そのため、Fisher 帰無仮説の下でなければ、確率計算を行い、リスク差の正確な信頼区間や p 値を求めることはできない。Fisher 帰無仮説の下で計算できたとしても、その結果が Neyman 帰無仮説や対立仮説の下で正しいという保証はない。

一般に、強い無視可能割付けメカニズムでなければ、並び替え検定を構成することは難しい。なぜなら、割付け確率が、観察されない潜在結果変数に依存すると、割付け確率が一部しかわからないからである。ランダム化に基づく推測に対し、Bayes 流の推測では、5.2 節で述べたように、割付けメカニズムが無視可能であることを利用して、 $\Pr(\mathbf{Y}^{\text{mis}} | \mathbf{Y}, \mathbf{A}, \mathbf{L})$ から因果リスク比に関する事後分布を導くことができる。

6.2 モデルに基づく推測

ランダム化プレイザウィナルールの下では、頻度論または Bayes 流のパラメトリックモデルを仮定した推測が有力である。しかし、割付けがアウトカムに依存するとき、頻度論の枠組みでは因果リスク差 $\hat{\tau}$ の誤差分布に影響が出ることには注意が必要である。Wei, et al. (1990) によると、 $\Pr(Y_i | A_i)$ に 2 項分布を仮定したとき、ランダム化プレイザウィナルールは $\hat{\tau}$ の分布に影響するが、大標本ではやはり正規分布に従い、分散は $\text{Var}(\hat{\tau}) = \pi^1(1 - \pi^1)/N_1 + \pi^0(1 - \pi^0)/N_0$ で与えられる。ただしこれはサンプルサイズがかなり大きいときの結果である。サンプルサイズが 50 よりも小さいときは、通常の 95% 信頼区間の公式ではなく、正確な信頼区間が推奨されている。Wei, et al. (1990) には、十分統計量 S と N_1 で条件付けた下での正確な信頼区間と、無条件の正確な信頼区間が示されており、こちらを用いるべきである。

7. ハーバード試験

表 5. ミシガン試験のデータ

登録順序 i	$\Pr(A_i=1 Y_j, A_j, j=1, \dots, i-1)$	A_i	Y_i^1	Y_i^0
1	1/2	1	0	欠測
2	2/3	0	欠測	1
3	3/4	1	0	欠測
4	4/5	1	0	欠測
5	5/6	1	0	欠測
6	6/7	1	0	欠測
7	7/8	1	0	欠測
8	8/9	1	0	欠測
9	9/10	1	0	欠測
10	10/11	1	0	欠測
11	11/12	1	0	欠測
12	12/13	1	0	欠測

7.1 共変量の偏り

ハーバード試験の割付けメカニズムは、無視可能であるが、強い無視可能には分類されない。厳密には個人ごとに独立に割付けがなされているわけではないが、実質的に割付け確率は以下のように近似できる。

$$\Pr(A_i=1 | \text{第1ステージ}) \approx 0.5$$

$$\Pr(A_i=1 | \text{第2ステージ}) \approx 1$$

つまり、第1ステージの19人のプロペンシティスコアは0.5とほぼ等しい。プロペンシティスコアのバランススコアという性質から、第1ステージ内でECMO群と従来療法群の比較は妥当と予想される。一方で第2ステージの20人のプロペンシティスコアは1で、第1ステージとは異なるから、共変量が偏って分布しても不思議ではない。

表6にハーバード試験の結果を示す。第1ステージ内でECMO群と従来療法群を比較すると、リスク差は-40% (95%信頼区間 -70~-10%) で、 χ^2 検定では $p=0.03$, Fisher 正確検定では片側 $p=0.05$ (二倍して両側にすると $p=0.10$) と、ECMO群の死亡確率が低かった。

表6には、院外出生児・院内出生児で層別した結果も示されている。まず注目してほしいのは、第1ステージと第2ステージの違いである。院内出生児の割合は、第1ステージ (10/19) より第2ステージ (5/20) の方が低い。すなわち、第1ステージと第2ステージで対象者の背景が異なるため、従来療法群と第2ステージのECMO群を直接比較できるかどうか慎重に考えなければならない。

ランダム化臨床試験には平均的には交絡はなく、観察研究には交絡があるといわれることがあ

表 6. ハーバード試験のデータ

	第 1 ステージ (ランダム化)			第 2 ステージ
	ECMO (A=1)	従来療法 (A=0)	効果指標	
全体	9	10	リスク差 -40%	20
生存 (Y=0)	9	6	リスク比推定不能	19
死亡 (Y=1)	0 (0%)	4 (40%)	オッズ比推定不能	1 (5%)
院外出生児	4	5	リスク差 -80%	15
生存 (Y=0)	4	1	リスク比推定不能	14
死亡 (Y=1)	0 (0%)	4 (80%)	オッズ比推定不能	1 (7%)
院内出生児	5	5	リスク差 0%	5
生存 (Y=0)	5	5	リスク比推定不能	5
死亡 (Y=1)	0 (0%)	0 (0%)	オッズ比推定不能	0 (0%)

る。これは常に正しいわけではない。ハーバード試験はランダム化臨床試験として報告されているが (O'Rourke, et al. 1989), 割付けメカニズムが特殊であるため, 院外出生児という共変量の偏り (交絡の一種) が問題になった珍しい例といえる。

7.2 交 絡

ここで交絡という概念について述べておく。共変量の偏りがなぜ問題かという点, ECMO 群と従来療法群で疾患の状態に違いがあることを意味するからである。仮に, ECMO 群の対象者を従来療法で治療したとする。そのときの死亡割合は, 従来療法群の死亡割合とは異なるかもしれない。何もしなくても差があるのだから, ECMO 群と従来療法群の差をとっただけでは, ECMO の効果を正しく調べられない。これを比較可能性 (comparability) または交換可能性 (exchangeability) がないという。比較可能性の欠如は, 交絡の定義の 1 つである (Greenland, Robins, and Pearl 1999, 佐藤・松山 2011)。

Greenland, Robins, and Pearl (1999) と Imbens and Rubin (2015) は, それぞれ潜在結果変数に基づく数学的に等価な交絡の定義を与えているが, その説明は異なっている。前者は, 比較可能性を定式化したものである。集団 A と B が比較可能とは, 同じ治療を受ければ同じようなアウトカムが生じることだから, これは潜在結果変数の確率分布を用いて

$$\Pr(Y_i^a=1|\text{集団A})=\Pr(Y_i^a=1|\text{集団B})$$

と表すことができる ($a=0, 1$)。コントロールとして集団 B をうまく選べばこれが成り立つかもしれない。

関心のある母集団因果リスク差は

$$\tau=\pi^1-\pi^0$$

$$= \Pr(Y_i^1=1) - \Pr(Y_i^0=1)$$

$$= \sum_{a=0}^1 \Pr(Y_i^1=1 | A_i=a) \Pr(A_i=a) - \sum_{a=0}^1 \Pr(Y_i^0=1 | A_i=a) \Pr(A_i=a)$$

と書くことができる。これは4つの条件付き確率を含んでいる。既に述べた一致性さえ成り立てば、 $\Pr(Y_i^0=1 | A_i=0)$ は、従来療法群の死亡割合 $\Pr(Y_i=1 | A_i=0)$ により推定することができる。 $\Pr(Y_i^1=1 | A_i=1)$ も同様に識別可能である。残りの2つはどうだろうか。

コントロールの選択がまず比較可能性がない状況は

$$\Pr(Y_i^0=1 | A_i=1) \neq \Pr(Y_i^0=1 | A_i=0)$$

と表される。このとき、 $\Pr(Y_i^0=1 | A_i=1)$ は、従来療法群の死亡割合から推定するとバイアスがあるし、ECMO群では Y_i^0 は観察できないため、識別することはできない。同様の議論から、

$$\Pr(Y_i^1=1 | A_i=1) \neq \Pr(Y_i^1=1 | A_i=0)$$

のとき、 $\Pr(Y_i^1=1 | A_i=0)$ は識別できない。Greenland, Robins, and Pearl (1999) による交絡とは、比較可能性が成り立っておらず、因果効果を識別できないという意味である。

Imbens and Rubin (2015) では、既に述べたように、非交絡（強い無視可能）という概念で、割付けメカニズムを分類している。これは、以下の条件付き独立性が成り立つことだった。

$$\{Y_i^0, Y_i^1\} \perp A_i | L_i$$

すなわち非交絡とは、共変量 L_i が等しい層の中では、ECMO群と従来療法群で Y_i^0 と Y_i^1 の分布が等しいことを意味しており、このことから

$$\Pr(Y_i^a=1 | A_i=1, L_i=l) = \Pr(Y_i^a=1 | A_i=0, L_i=l)$$

が成り立つ ($a=0,1$)。つまり、共変量 $L_i=1$ となるような対象者内に限ると、ECMO群と従来療法群は比較可能である。このことから、比較可能性が保たれていることと割付けメカニズムが強い無視可能であることは、本質的に同じ意味であり、共変量が不要などの単純な場合には一致することがわかる。

7.3 交絡因子

観察研究やプロペンシティスコアの文献では、しばしば交絡因子 (confounder) という概念が出てくる。しかし、交絡因子の定義は複数あり、因果推論の数学的な側面に議論を限定したとしても、どれがよいかを論じることは簡単ではない（詳細な議論が VanderWeele and Shpitser 2013 に与えられている）。Rubin (1978) や Rosenbaum and Rubin (1983) は、ランダム化臨床試験では割付けメカニズムに従って解析すべきであり、観察研究では治療が選ばれる前に生じるすべての（意味のある）変数を調整すべきである、という立場をとる。この点についてコンセンサスが得られているわけではなく、全ての治療前変数を調整すると、合流点 (collider) 層別バイアスが生じる可能性があるという批判がなされている (VanderWeele and Shpitser 2011)。また、交絡

因子とは治療とアウトカムの両方と相関するものを指すことがあるが、これは正しくなく、治療とアウトカムの途中にある因子（中間因子, intermediate）で調整すると、逆にバイアスが生じる。観察研究では交絡因子の選択が問題になるのが普通だが、そのための手法は Greenland (2008) や黒木・小林 (2011) が簡潔にまとめている。

7.4 併合可能性、交絡、効果の修飾の区別

ハーバード試験では、ECMO 群全体と従来療法群の間で、共変量（院内出生児・院外出生児）に偏りが生じた。量的な見方に立ったとき、これは因果リスク差の推定にどれくらいの影響があったのだろうか。

表 7 に、ECMO 群全体と従来療法群を比較した結果を示す。表 6 と比べると、層別前のリスク差（ -40% と -37% ）、層別後のリスク差（ -80% と -75% , 0% と 0% ）に違いはないようにみえる。更に、表 7 の全体の結果と院外出生児の結果を比べると、リスク差は大きく異なる（ -37% と -75% ）が、リスク比・オッズ比はすべて 0.1 より小さく、大きな差はない。疫学では、昔から原因と結果の両方に関係する第三の因子があると、層別する場合としない場合で、結果が異なることがあると教えられてきた。これが古典的な意味での交絡である。最近では、共変量で調整した前と後で、効果の指標の値が変化することを、併合可能性 (collapsibility) がない、と呼んでいる (Greenland, Robins, and Pearl 1999)。

表 6 に戻ると、第 1 ステージ内での比較は妥当なはずが、院外出生児・院内出生児で分けたサブグループ解析には併合可能性がない。つまり、院内出生児の予後は良好であり、死亡はみられなかったため、全体・院外出生児・院内出生児の結果はそれぞれ傾向が異なっていた。第 1 ステージ全体 19 人におけるリスク差は -40% であった。一方で、院内出生児 10 人からなるサブグループにおけるリスク差は -80% 、院内出生児 10 人からなるサブグループにおけるリスク差は 0% であった。このように、サブグループ間で因果効果が異なることを、効果の（指標の）修飾という。効果の修飾の程度は、効果の指標によって異なる。つまり、リスク差でみると ECMO の効果の大きさがサブグループ間で違うが、リスク比やオッズ比ではあまり変わらない（またはその逆）ということが生じ得る（表 7）。そのため、単に効果の修飾と呼ぶこともあるが、効果の指標の修飾 (effect measure modification) といった方がより正確である。

表 6 と 7 からわかるのは、複数の 2×2 表が併合可能かどうかを決める要素は 3 つあるということである。1 つ目は、層別する共変量が治療およびアウトカムと相関するかどうかである。このような変数を、疫学では交絡因子と呼ぶことがある（この定義は必ずしも正しくない）。2 つ目は、共変量によって効果の修飾があるかどうかである。そして 3 つ目は、効果の指標の特性である。リスク差・リスク比（の真値）は、共変量が治療またはアウトカムと独立で、効果の修飾がないとき、併合可能である。この性質はオッズ比では成り立たない。同じような議論は、 2×2 表以外のデータでも成立する。アウトカムが指数型分布族に従うとき、一般化線型モデルに基づいて考えると、上の 3 要素は、共変量間の相関、交互作用項の有無、リンク関数の種類に対応している (Appendix E 参照)。 2×2 表の層別解析では、交絡を調整しているのか効果の修飾を調べているのかあいまいになることがあるが、統計モデルを導入することで、両者は異なる概念であることが明らかになる。

表 7. ハーバード試験のデータ (ECMO 群を併合し観察研究として解析)

	ECMO 全体 ($A=1$)	従来療法 ($A=0$)	効果の指標
全体	29	10	リスク差-37%
生存 ($Y=0$)	29	6	リスク比 0.09
死亡 ($Y=1$)	1 (3%)	4 (40%)	オッズ比 0.05
院外出生児	19	5	リスク差-75%
生存 ($Y=0$)	18	1	リスク比 0.07
死亡 ($Y=1$)	1 (5%)	4 (80%)	オッズ比 0.01
院内出生児	10	5	リスク差 0%
生存 ($Y=0$)	10	5	リスク比推定不能
死亡 ($Y=1$)	0 (0%)	0 (0%)	オッズ比推定不能

7.5 Bayes 流の推測

もう 1 つの有力なアプローチは Bayes 推測である。Ware (1989) による総説では、ハーバード試験 (O'Rourke, et al. 1989) の結果は、Bayesian の立場でどのような解釈になるかを考察している。中心的な役割を果たすのは、第 1 ステージで得られた尤度 $(\pi^0)^4(1-\pi^0)^6(1-\pi^1)^9$ である。ECMO が従来療法に比べて、劣るのか、同等か、優るのかに関する事前確率を

$$\Pr(\pi^0 < \pi^1) = \Pr(\pi^0 = \pi^1) = \Pr(\pi^0 > \pi^1) = \frac{1}{3}$$

と設定した。尤度を用いて更新すると、事後確率はどうなるだろうか。

Bayes の定理を適用するためには、これらの事前確率だけではなく、 π^0 と π^1 の事前同時分布が必要になる。事前分布全体を特定するため、パラメータ a と b を持つベータ分布 $f(\pi^0 | a, b)$ 、一様分布 $f(\pi^1 | \pi^0 = c, \pi^0 < \pi^1) = 1/c$, $f(\pi^1 | \pi^0 = c, \pi^0 > \pi^1) = 1/(1-c)$ が用いられた。

まず関心があるのは、 π^0 について無情報事前分布に対応するベータ分布のパラメータ $a=b=1$ を用いたときである。このとき、事後確率は

$$\Pr(\pi^0 < \pi^1 | \mathbf{Y}, \mathbf{A}) = 0.01$$

$$\Pr(\pi^0 = \pi^1 | \mathbf{Y}, \mathbf{A}) = 0.10$$

$$\Pr(\pi^0 > \pi^1 | \mathbf{Y}, \mathbf{A}) = 0.89$$

と計算された。つまり、最初の 19 人のデータだけで ECMO が従来療法より同等以上であるという強い証拠が得られていたことがわかる。試験前のチャートレビューでは、従来療法の死亡割合は 13 人中 11 人であった。これは $a=12$, $b=3$ に対応する。このとき、 $\Pr(\pi^0 > \pi^1 | \mathbf{Y}, \mathbf{A}) = 0.9957$ というさらに強い証拠が得られた。Ware (1989) は、これらの結果に基づいて、第 2 ステージで

ランダム化を継続することを正当化することは難しいと述べている。

8. 考 察

因果効果に関する推測方式は、モデルに基づく頻度論の手法、Bayes 流、推定方程式を用いる手法（g 推定など）、ランダム化に基づく手法の4つに大別される。ECMO の事例で興味深いのは、ミシガン試験・ハーバード試験でのみ、推測方式による解析結果やその解釈の違いが顕著になったことである。

英国試験は典型的なランダム化臨床試験の一例といってよい。このとき、割付けメカニズムは、p 値などの確率計算の基礎を与えるとされる。しかし、議論が簡単なのは Fisher 帰無仮説のときに限られる。潜在結果変数を導入すると、 2×2 表のような単純なケースですら識別不能性が生じる。ただし、積 2 項分布の下でのリスク差の分散は保守的な信頼区間を導くだけなので、実用的に大きな問題ではない。また、解析が単純な 2 群比較であれば g 推定を用いるメリットはない。g 推定は、パラメトリックモデルを前提にしないという意味で仮定が緩い方法であるが、その代償として検出力が低下する。適用事例では g 推定の 95% 信頼区間の幅は、通常よりも 1.6 倍ほど広がった。この手法のよさが出るのは、ノンコンプライアンスや治療切り替えが生じているときである。このような状況では、パラメトリックモデルを仮定した Intention-To-Treat 解析に比べ、検出力が向上することがある（Sato 2000, Matsuyama 2010）。

アウトカム適応的ランダム化では、中程度のサンプルサイズでは漸近理論がうまく働かないことが問題になる。実践的な指針として、潜在結果変数にパラメトリックモデルを仮定し、正確な p 値や信頼区間を求めるか（Wei, et al. 1990）、Bayes 流の推測を行うことが薦められる。最近のがん第 II 相臨床試験では、I-SPY2 試験（Park, et al. 2017）などのアウトカム適応的ランダム化が採用される例が増えてきた。有望な薬を迅速にスクリーニングするという第 II 相試験の目的を考えると、柔軟性の高い Bayes 推測は、アウトカム適応的ランダム化臨床試験にマッチしている。しかしながら、頻度論・Bayesian 共通の問題として、コントロール群との比較可能性に注意を払う必要がある。割付けがランダムだったとしても、ハーバード試験のように途中で割付け確率（プロペンシティスコア）が大きく変わると、比較可能性が保たれていないことがあり得る。

9. 結 語

ECMO 臨床試験は、医療技術をどのように評価すればよいか、という文脈で行われた研究事例である。しかし医学は、医療技術以外に様々な対象を扱う。たとえば、環境問題を取り上げると、喫煙、大気汚染、ハウスダストのように、様々な因子や条件が、喘息などの疾患に影響すると考えられる。そのとき、ある環境因子の因果効果は、他の効果と分離することが難しい。しかし、だからといって因果関係を否定したり、結論を先送りしたりすることが、正しいとは限らない。潜在結果変数モデルは、特定の原因に注目して、事実と反事実を考えるモデルであるが、これにとらわれすぎると現実にはまずいこともある。このように、モデルは有効に働くケースもそうでないケースもある。しかし、潜在結果変数モデルを共通言語の 1 つとすることで、医療技術評価と環境問題のどこが違うのかを論じることが可能になる。ここで論じられた諸概念は、因果

問題を考えたり，解決したりするための，いわば実用上の基礎というべきものである．

謝 辞

貴重なコメントを頂いた佐藤俊哉先生，篠崎智大先生，査読者に感謝する．本論文は，国立研究開発法人日本医療研究開発機構（AMED）の臨床研究・治験推進研究事業/生物統計家育成支援事業（18lk0201061t0003）の資金によるものである．

参考文献

- Bartlett, R.H., Roloff, D.W., Cornell, R.G., Andrews, A.F., Dillon, P.W., Zwischenberger, J.B. (1985). Extracorporeal circulation in neonatal respiratory failure: a prospective randomized study, *Pediatrics* **76** (4), 479-487.
- Chiba, Y. (2016). Sample size calculation of exact tests for the weak causal null hypothesis in randomized trials with a binary outcome, *Open Journal of Statistics* **6**, 766-776.
- Copas, J.B. (1973). Randomization models for the matched and unmatched 2×2 tables, *Biometrika* **60**(3), 467-476.
- Cox, D.R. (1958). *The Planning of Experiments*, John Wiley, New York
- Ding, P., Dasgupta, T. (2016). A potential tale of two-by-two tables from completely randomized experiments, *Journal of the American Statistical Association* **111**, 157-168.
- Fisher, R. A. (1935). *The Design of Experiments*, Oliver & Boyd, Edinburgh (遠藤健児，鍋谷清治 訳 (1971). 実験計画法，森北出版，東京)
- Frangakis, C.E., Rubin, D.B. (2002). Principal stratification in causal inference, *Biometrics* **58**(1), 21-29.
- Greenland, S. (2008). Invited commentary: variable selection versus shrinkage in the control of multiple confounders, *American Journal of Epidemiology* **167**(5), 523-529.
- Greenland, S., Brumback, B. (2002). An overview of relations among causal modelling methods, *International Journal of Epidemiology* **31**, 1030-1037.
- Greenland, S., Robins, J.M., Pearl, J. (1999). Confounding and collapsibility in causal inference, *Statistical Science* **14**(1), 29-46.
- Henmi, M., Eguchi, S. (2004). A paradox concerning nuisance parameters and projected estimating functions, *Biometrika* **91**, 929-941.
- Imbens, G.W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review, *The Review of Economics and Statistics* **86**(1):4-29.
- Imbens, G.W., Rubin, D.B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press, Cambridge

- Kempthorne, O. (1952). *The Design and Analysis of Experiments*, John Wiley, New York
- 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** (2), 119-144.
- Matsuyama, Y. (2010). A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial, *Statistics in Medicine* **29**(20), 2107-2116.
- Neyman, J. (1990). On the application of probability theory to agricultural experiments. Essay on principles. Section 9, translated from the 1923 Polish original and edited by Dabrowska, D.M. and Speed, T.P., *Statistical Science* **5**, 465-472.
- O'Rourke, P.P., Crone, R.K., Vacanti, J.P., Ware, J.H., Lillehei, C.W., Parad, R.B., Epstein, M.F. (1989). Extracorporeal membrane oxygenation and conventional medical therapy in neonates with persistent pulmonary hypertension of the newborn: a prospective randomized study, *Pediatrics* **84**(6), 957-963.
- Park, J.W., Liu, M.C., Yee, D., Yau, C., van 't Veer, L.J., Symmans, W.F., et al; I-SPY 2 Investigators. (2016). Adaptive randomization of neratinib in early breast cancer, *New England Journal of Medicine* **375**(1), 11-22.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference 2nd edition*, Cambridge University Press, Cambridge (黒木学 訳 (2009). 統計的因果推論—モデル・推論・推測—, 共立出版, 東京)
- Robins, J.M. (1988). Confidence intervals for causal parameters, *Statistics in Medicine* **7**, 773-785.
- 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, *Annals of Statistics* **7**, 34-58.
- Rubin, D.B. (1990). Formal mode of statistical inference for causal effects, *Journal of Statistical Planning and Inference* **25**(3), 279-292.
- Sato, T. (2000). Sample size calculations with compliance information, *Statistics in Medicine* **19**(19), 2689-2697.
- 佐藤俊哉, 松山裕. (2011). 交絡という不思議な現象と交絡を取りのぞく解析—標準化と周辺構造モデル—, 計量生物学 **32**, S35-S49.
- UK Collaborative ECMO Trial Group. (1996). UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation, *Lancet* **348**(9020), 75-82.
- VanderWeele, T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*, Oxford University Press, Oxford
- VanderWeele, T.J., Hernán, M.A. (2013). Causal inference under multiple versions of treatment, *Journal of Causal Inference* **1**(1), 1-20.

- VanderWeele, T.J., Shpitser, I. (2011). A new criterion for confounder selection, *Biometrics* **67**, 1406–1413.
- VanderWeele, T.J., Shpitser, I. (2013). On the definition of a confounder, *Annals of Statistics* **41**, 196–220.
- Vansteelandt, S., Joffe, M. (2014). Structural nested models and g-estimation: the partially realized promise, *Statistical Science* **29**(4), 707–731.
- Ware, J.H. (1989). Investigating therapies of potentially great benefit: ECMO, *Statistical Science* **4**(4), 298–306.
- Wei, L.I., Durham, S. (1978). The randomized play-the-winner rule in medical trials, *Journal of the American Statistical Association* **73**:840–843.
- Wei, L.J., Smythe, R.T., Lin, D.Y., Park, T.S. (1990). Statistical inference with data-dependent treatment allocation rules, *Journal of the American Statistical Association* **85**, 156–162.
- Woodcock, J., LaVange, L.M. (2017). Master protocols to study multiple therapies, multiple diseases, or both, *New England Journal of Medicine* **377**(1), 62–70.

付録

A. プロペンシティスコアの性質

証明は Rosenbaum and Rubin (1983) による. プロペンシティスコア $e(L_i=1)$ がバランススコアであるとは,

$$\Pr \{A_i=1 | L_i, e(L_i=l)\} = \Pr \{A_i=1 | e(L_i=l)\}$$

を意味する. $e(L_i=1)$ は L_i の関数であるから, この式の左辺は

$$\begin{aligned} \Pr \{A_i=1 | L_i, e(L_i=l)\} &= \Pr \{A_i=1 | L_i=l\} \\ &= e(L_i=l) \end{aligned}$$

また, 右辺は期待値の反復から

$$\begin{aligned} \Pr \{A_i=1 | e(L_i=l)\} &= E\{A_i=1 | e(L_i=l)\} \\ &= E[E\{A_i | L_i, e(L_i=l) | e(L_i=l)\}] \\ &= E\{e(L_i=l) | e(L_i=l)\} \\ &= e(L_i=l) \end{aligned}$$

従って, 左右が一致するため, $e(L_i=l)$ はバランススコアである. バランススコアの性質から,

$$\{Y_i^0, Y_i^1\} \perp\!\!\!\perp A_i | e(L_i=l)$$

が示される.

また, プロペンシティスコア以外の任意のバランススコアを $b(L_i=l)$ と表す. このとき, 2つの異なる値 l と l' について, $b(l)=b(l')$ かつ $e(L_i=l) \neq e(L_i=l')$ となるものがなければならない.

$$\Pr (A_i=1 | L_i=l) = e(L_i=l) \neq e(L_i=l') = \Pr (A_i=1 | L_i=l')$$

これは, $b(l)$ がバランススコアという定義に矛盾する. よって, $e(L_i=l)$ はあらゆるバランススコアの関数であり, これが最も粗い (coarsest) の意味である.

B. 連続データにおける平均の差の漸近分散

わかりやすい解説が Imbens and Rubin (2015) のテキストに与えられている。 Y_i^0 と Y_i^1 が連続データで、固定されており、割付けの並び替え分布を考える。そのとき、平均の差

$$\bar{\tau} = \frac{\sum_{i: A_i=1} Y_i}{N_1} - \frac{\sum_{i: A_i=0} Y_i}{N_0}$$

の分散は

$$\text{Var}(\bar{\tau}) = \frac{V_1}{N_1} + \frac{V_0}{N_0} - \frac{V_\tau}{N}$$

で与えられる。ここで、 $V_a = \sum_{i=1}^N (Y_i^a - \bar{Y}^a)^2 / (N-1)$ (ただし $a=0, 1$) は、有限集団における潜在結果変数の分散、 $V_\tau = \sum_{i=1}^N (\tau_i - \bar{\tau})^2 / (N-1)$ は、個人の因果効果 $\tau_i = Y_i^1 - Y_i^0$ の分散である。Fisher 帰無仮説 $Y_i^0 = Y_i^1$ の下では、 $V_\tau = 0$ だから、 $\text{Var}(\bar{\tau})$ は一般的な分散の公式に帰着する。

C. g 推定量の漸近分散

g 推定は、強い無視可能性（非交絡性）

$$\{Y_i^0, Y_i^1\} \perp\!\!\!\perp A_i \mid L_i$$

の下で、母集団因果リスク差 $\tau = \pi^1 - \pi^0$ や母集団因果リスク比 $\tau = \pi^1 / \pi^0$ の推定関数を構成する方法である。

母集団因果リスク差の場合は

$$H_i(\tau) = Y_i - \tau A_i$$

母集団因果リスク比 $\tau = \pi^1 / \pi^0$ の場合は

$$H_i(\tau) = Y_i \exp(-\tau A_i)$$

と定義する。どちらの場合も、 $H_i(\tau)$ には、真値 τ^* の下で

$$E\{H_i(\tau^*) \mid A_i, L_i\} = E(Y_i^0 \mid A_i, L_i)$$

という性質がある。従って、強い無視可能性が成り立てば、 Y_i^0 と同様に

$$E\{H_i(\tau^*) \mid A_i=0, L_i\} = E\{H_i(\tau^*) \mid A_i=1, L_i\}$$

という関係が成り立つ。このとき、 $H_i(\tau)$ と A_i の条件付き共分散が、母集団においてゼロになる。言い換えれば、標本における条件付き共分散（これは τ の関数である）をゼロと置くことで、以下の推定方程式が導かれる。

$$\sum_{i=1}^N U_i(\tau) = \sum_{i=1}^N \{A_i - E(A_i \mid L_i)\} [H_i(\tau) - E\{H_i(\tau) \mid L_i\}] = 0.$$

ここで、 $E(A_i \mid L_i) = \Pr(A_i=1 \mid L_i)$ はプロペンシティスコアである。ただし、英国試験では共変量

L_i で条件付ける必要はない. M 推定の理論から, この推定方程式の解は漸近的に以下の正規分布に従うことが示される.

$$N^{1/2}(\hat{\tau} - \tau^*) \rightarrow N(0, ABA)$$

ここで, $B = \text{Var} \{U_i(\tau^*)\}$, $A = -E\{\partial U_i(\tau)/\partial \tau | \tau = \tau^*\}^{-1}$ である.

D. 多変量超幾何分布の下でのリスク差の漸近分布

証明は Copas (1973) による. $x_{00}, x_{01}, x_{10}, x_{11}$ の同時分布には, $x_{00} + x_{01} + x_{10} + x_{11} = N_1$, $x_{01} + x_{11} = S_1$, $n_{11} - x_{11} + n_{10} - x_{10} = S_0$ という制約がある. 多変量超幾何分布の確率関数の分子は, これらを代入すると

$$\binom{n_{00}}{N_1 - n_{11} + S_0 - n_{10} - S_1 + x_{11}} \binom{n_{01}}{S_1 - x_{11}} \binom{n_{10}}{n_{11} - x_{11} + n_{10} - S_0} \binom{n_{11}}{x_{11}}$$

に帰着する. ただし, これは $x_{00}, x_{01}, x_{10}, x_{11}$ の理論的な確率分布であり, 実際に観察されるわけではない. 観察されるのは表 2 のデータである.

ここで知りたいのは, 因果リスク差の推定値がどのように分布するかである. 因果リスク差は, 両群の死亡数 S_0 と S_1 によって決まる. N_0 と N_1 が十分に大きいとき, この 2 つの確率変数は中心極限定理により 2 変量正規分布に従う. 2 変量正規分布のパラメータは

$$E\left(\frac{S_a}{N_a}\right) = \pi^a$$

$$\text{Var}\left(\frac{S_a}{N_a}\right) = \frac{(N - N_a)\pi^a(1 - \pi^a)}{N_a(N - 1)}$$

$$\text{Cov}\left(\frac{S_0}{N_0}, \frac{S_1}{N_1}\right) = -\frac{\pi^1(1 - \pi^1) + \pi^0(1 - \pi^0) - \beta}{2(N - 1)}$$

$$\beta = \frac{n_{01} + n_{10}}{N}$$

よって, 因果リスク差 $\hat{\tau} = S_1/N_1 - S_0/N_0$ の分散は

$$\begin{aligned} \text{Var}(\hat{\tau}; \beta) &= \frac{1}{N - 1} \left\{ \frac{N_0 \pi^1(1 - \pi^1)}{N_1} + \frac{N_1 \pi^0(1 - \pi^0)}{N_0} + \pi^1(1 - \pi^1) + \pi^0(1 - \pi^0) - \beta \right\} \\ &= \frac{N}{N - 1} \left\{ \frac{\pi^1(1 - \pi^1)}{N_1} + \frac{\pi^0(1 - \pi^0)}{N_0} - \frac{\beta}{N} \right\} \end{aligned}$$

となる.

E. 一般化線型モデルにおける調整前後の回帰係数の差

アウトカム Y_i が 2 値変数で, リンク関数 g を持つ一般化線型モデルに従うとする. 治療変数を A_i , 共変量 L_i とすると (すべて 2 値変数), Y_i は以下の確率を持つ Bernoulli 分布に従う確率変数になる.

$$\Pr(Y_i = 1 | A_i, L_i) = g(\beta_0 + \beta_1 A_i + \beta_2 L_i + \beta_3 A_i L_i)$$

このモデルによって, 層別した 2×2 表の確率分布を表すことができる. 一方で, 層別前の 2×2

表のセル確率は,

$$\Pr(Y_i=1|A_i)=E\{\Pr(Y_i=1|A_i,L_i)|A_i\}$$

のように, L_i について期待値をとることで計算できる.

まずはリスク差の性質を調べよう. リンク関数が恒等関数で, 交互作用がないとき ($\beta_3=0$), β_1 は共変量で調整したリスク差と呼ばれる. 一方で, 層別前の 2×2 表で, リスクの差を計算したときの真値 (期待値) は

$$\tau = \Pr(Y_i=1|A_i=1) - \Pr(Y_i=1|A_i=0)$$

である. $\Pr(L=l|A=a)=P_a$ と置く. 上の式の各項に, 一般化線型モデルの平均構造を代入すると

$$\begin{aligned}\Pr(Y_i=1|A_i) &= E\{\Pr(Y_i=1|A_i,L_i)|A_i\} \\ &= E(\beta_0 + \beta_1 A_i + \beta_2 L_i + \beta_3 A_i L_i | A_i) \\ &= (\beta_0 + \beta_2 P_0) + \{\beta_1 + \beta_2(P_1 - P_0) + \beta_3 P_1\} A_i\end{aligned}$$

のように, 回帰係数による表現が得られる. この差が無調整リスク差 τ だった. つまり調整リスク差と無調整リスク差の関係は

$$\beta_1 + \{\beta_1 + \beta_2(P_1 - P_0) + \beta_3 P_1\} = \tau$$

であり, $P_0=P_1$ または $\beta_2=0$ で, $\beta_3=0$ のとき, 両者は等しくなる.

同じ計算を, リスク比とオッズ比について行うとどうなるだろうか. 調整リスク比は, リンク関数が対数のモデルにおける $\exp(\beta_1)$ のことである. 簡単のため $\beta_3=0$ のときを考えると

$$\begin{aligned}\Pr(Y_i=1|A_i) &= E\{\exp(\beta_0 + \beta_1 A_i + \beta_2 L_i) | A_i\} \\ &= \exp(\beta_0 + \beta_1 A_i) \{\exp(\beta_2) P_A + (1 - P_A)\} \\ &= \exp\left\{\text{constant} + \left(\beta_1 + \log \frac{\exp(\beta_2) P_1 + (1 - P_1)}{\exp(\beta_2) P_0 + (1 - P_0)}\right) A_i\right\}\end{aligned}$$

ここで, constant は A_i に依存しない定数部分である. ここから, 調整回帰係数と無調整回帰係数の差 (リスク比に変換すると比) は, P_0, P_1, β_2 によって決まっており, $P_0=P_1$ または $\beta_2=0$ で, $\beta_3=0$ のとき, 両者が等しいことがわかる.

調整オッズ比は, 共変量を調整したロジスティック回帰における $\exp(\beta_1)$ を指す. ロジスティック回帰に同じ計算を適用すると

$$\frac{\Pr(Y_i=1|A_i)}{1-\Pr(Y_i=1|A_i)} = \beta_0 + \beta_1 A_i + h(A_i)$$

$$h(A_i) = \log \frac{\exp(\beta_2) \{1 + \exp(\beta_0 + \beta_1 A_i)\} P_A + \{1 + \exp(\beta_0 + \beta_1 A_i + \beta_2)\} (1 - P_A)}{\{1 + \exp(\beta_0 + \beta_1 A_i)\} P_A + \{1 + \exp(\beta_0 + \beta_1 A_i + \beta_2)\} (1 - P_A)}$$

という結果が得られる。一般に、ロジスティック回帰で共変量について期待値をとると、 A_i の線型の項と非線型の項が出てくる。非線型の項は、 $P_0 = P_1$ または $\beta_2 = 0$ だったとしても、必ずしもゼロにならない。言い換えると、オッズ比は調整前後で併合可能な指標ではない。