

# なぜ ベイズを使わない のか!?

臨床試験デザインのために

手良向 聡 著



## まえがき

医学・薬学・看護学の分野に身を置くと、統計学は難しくて分からないという人に多く出会う。それは、私たち専門家の教え方にも問題があると思うが、それだけが原因ではなく、統計学は確率を扱う学問であり、確率とは何かについて多くの考え方があり、適用場面によってそれらを使い分ける必要があるから難しいのである。また、技術的な面は学習によってある程度は克服可能であるが、統計学を適用する場面は社会のほぼすべての領域にわたっていて、その分野の知識と経験がなければ、結果の正しい解釈ができないのである。さらに、統計学は数学の一分野であると思っている人がいるが、それは誤りである。もう30年以上前になるが、私が通っていた大学の数学科に統計学の講義はなく、おそらく今もないと思う。統計学は積分や行列など数学で用いる道具を用いるが、数学の延長線上にはない学問なのである。それは一言でいうと、演繹の学問と帰納の学問という違いである。ただし、数学の一分野である確率論は1500年代にジロラモ・カルダーノ、1600年代にブレーズ・パスカルとピエール・ド・フェルマーが議論した賭け金の分配問題<sup>1)</sup>に端を発しており、そこからヤーコブ・ベルヌーイによる大数の法則、アブラム・ド・モアブルによる正規分布などが生まれた。従って、確率論を土台にして数理統計学という学問が生まれたと言うことはできる。ちなみに、ベイズ流統計学の始祖トーマス・ベイズ(1702-1761)はド・モアブル(1667-1754)や偉大な数学者レオンハルト・オイラー(1707-1783)とほぼ同じ時代に生きた。

統計学を本格的に勉強し始めると思想・哲学の重要性に気付く。従って、統計学を本当に理解したければ、まずは科学哲学、特に確率の哲学やそれにまつわる思想を学んだ方がよい(たとえば『偶然を飼いなす』<sup>2)</sup>『確率の哲学理論』<sup>3)</sup>『偶然とは何か』<sup>4)</sup>『科学と証拠』<sup>5)</sup>『確率の出現』<sup>6)</sup>など)。特に、プラグマティズムという思想は統計学と関係の深い重要な思想である

(『パースのプラグマティズム』<sup>7)</sup>など)。本書で主に扱うベイズ流統計学は、古典的と呼ばれている頻度流統計学に対して異端の統計学として扱われてきた歴史をもつ<sup>8)</sup>。しかしながら、近年はベイズ流統計学ブームと言ってもよいくらいの人気ので、ここ数年間に膨大な数の参考書が刊行されている。これらの多くはデータサイエンス、ビッグデータのブームと連動し、主にデータ解析に主眼が置かれている。本書では、それらとは異なる視点で臨床試験のデザイン（いわゆる実験計画）にこれまで主として用いられてきた頻度流の方法の利点と欠点を述べながら、それを補うような形でベイズ流の方法をどう利用するかについて分かりやすく解説することを意図している。また本書では、標本サイズの計算やデータ解析を行うための詳細な技術についてはほとんど述べず（多くの良書を参照し）、その考え方の基本を伝えることを目標としている。最初に断っておくが、臨床試験のデザインと解析に頻度流の方法が有用な場面は多く、すべての臨床試験をベイズ流の方法で行うべきだと主張するつもりはない。

ベイズ流臨床試験を行う際には、臨床研究者と統計家の真の協同が不可欠である。臨床研究者は、事前情報の確かさ、目標値の見積もり、臨床試験を継続するかどうかを決定する確率閾値などについて、専門家としての見解を持っている。統計家は、事前情報や目標値の妥当性、実施可能性などについて臨床研究者と議論し、デザインを決定しなければならない。近年、臨床試験にベイズ流接近法が有用であるという報告が増えつつある。米国の食品医薬品局（Food and Drug Administration: FDA）医療機器・放射線保健センターは、2010年に企業とFDAスタッフ向けに「医療機器の臨床試験におけるベイズ流統計学の利用に関するガイダンス」を公表した（第3部「ベイズ流臨床試験のガイダンス」参照）。今後、新しい臨床試験デザインの開発は、医薬品や医療機器の承認申請にインパクトを与える可能性がある。特に、資源を有効に活用するという観点から効率よく臨床試験を行うことが今後ますます重要になるであろう。

本書は、ベイズ流臨床試験デザインを多くの方に知っていただくことを意図しているが、それらは第2部（第11章から第15章）に書かれている。

第1部（第1章から第10章）は、その準備のための章であり、臨床試験デザインの基本および古典的な頻度流接近法の長所と短所について解説している。特に、第10章には頻度流統計学の問題点をまとめている。それらの部分を読む必要のない方は第2部から読み進み、必要に応じて第1部、および第3部を参照していただければ幸いである。

本書の内容は、多くの先生方からの指導・助言に基づいている。まずは、臨床試験の方法論についてご指導いただいた福島雅典先生、生物統計学について長年ご指導いただいた大橋靖雄先生、折笠秀樹先生、佐藤俊哉先生に深く感謝したい。また、共同研究者または同僚として多くの助言をしてくれた松山裕先生、松井茂之先生、森田智視先生、山中竹春先生、大門貴志先生、大庭幸治先生、吉村健一先生、田中司朗先生、横田勲先生に感謝したい。さらに、本書の原稿を読んで有益な助言をしてくれた大門貴志先生、木下文恵氏、金芳堂の浅井健一郎氏には改めて感謝したい。

## 文 献

---

- 1) デブリン K (原 啓介訳). 世界を変えた手紙: パスカル, フェルマーと〈確率〉の誕生. 岩波書店. 2010.
- 2) ハッキング I (石原英樹, 重田園江訳). 偶然を飼いなす: 統計学と第二次科学革命. 木鐸社. 1999.
- 3) ギリース D (中山智香子訳). 確率の哲学理論. 日本経済評論社. 2004.
- 4) 竹内啓. 偶然とは何か: その積極的意味. 岩波書店. 2010.
- 5) ソーバー E (松王政浩訳). 科学と証拠 - 統計の哲学 入門. 名古屋大学出版会. 2012.
- 6) ハッキング I (広田すみれ, 森元良太訳). 確率の出現. 慶應義塾大学出版会. 2013.
- 7) 伊東邦武. パースのプラグマティズム: 可謬主義的知識論の展開. 勁草書房. 1985.
- 8) マグレイン SB (富永 星訳). 異端の統計学ベイズ. 草思社. 2013.

## 目 次

まえがき

<b>第1部 臨床試験デザイン入門</b>	<b>1</b>
<b>1 医学研究と統計学</b> .....	<b>3</b>
1 医学・医療と臨床試験－基礎研究と臨床研究の違い	4
<b>2 臨床試験の計画</b> .....	<b>9</b>
1 臨床試験実施計画書	9
2 探索的臨床試験と検証的臨床試験	10
3 臨床試験デザイン	10
4 評価項目	12
<b>3 統計的仮説</b> .....	<b>16</b>
<b>4 解析対象集団とサブグループ解析</b> .....	<b>18</b>
<b>5 中間モニタリング</b> .....	<b>21</b>
1 探索的試験における中間モニタリング	22
2 検証的試験における中間モニタリング	22
3 適応的デザイン	23
<b>6 頻度流統計学－仮説検定</b> .....	<b>25</b>
<b>7 頻度流統計学－推定と信頼区間</b> .....	<b>30</b>
<b>8 標本サイズ設定と検出力解析</b> .....	<b>32</b>
<b>9 頻度流の標本サイズ設定</b> .....	<b>33</b>
1 2項検定に基づく標本サイズ設定	33
2 2段階デザインに基づく標本サイズ設定	36
<b>10 頻度流統計学の問題点</b> .....	<b>39</b>
1 統計的有意性とP値の問題	39
2 検定の多重性	40

## 第2部 ベイズ流臨床試験デザイン 45

- 11 ベイズ流統計学 ..... 47
- 12 事前分布, 尤度, 事後分布, 予測分布 ..... 51
- 13 単群臨床試験デザイン ..... 55
  - 1 標本サイズ設定 55
  - 2 事前分布の設定方法 67
  - 3 中間モニタリング 69
  - 4 方法の選択と実施手順 75
  - 5 標本サイズの再設定 76
  - 6 効果と安全性を組み合わせた中間モニタリング 79
- 14 2群臨床試験デザイン ..... 81
  - 1 標本サイズ設定 81
  - 2 中間モニタリング 83
- 15 メタアナリシス ..... 85
  - 1 結果の異質性 85
  - 2 結果の統合 86

## 第3部 ベイズ流臨床試験のガイダンス 91

- 1 はじめに..... 94
- 2 序章..... 94
  - 2.1 ベイズ流統計学とは? 94
  - 2.2 なぜ医療機器に関してベイズ流統計学を用いるのか? 95
  - 2.3 なぜ現在ベイズ流手法がより一般的に用いられているのか? 97
  - 2.4 いつFDAはベイズ流試験の計画に参加すべきか? 97
  - 2.5 ベイズ流接近法は健全な科学の代わりにはならない 97
  - 2.6 ベイズ流手法を用いる潜在的な利益は? 98
  - 2.7 ベイズ流接近法を用いる潜在的な難しさは? 100
  - 2.8 ベイズ流解析を行うために利用できるソフトウェアプログラムは? 103
  - 2.9 ベイズ流統計学についてもっと学ぶために利用できる情報源は? 103
- 3 ベイズ流統計学 ..... 105
  - 3.1 アウトカムおよびパラメータ 105
  - 3.2 ベイズ流の理論的枠組み 106

3.3	事前分布とは？	107
3.4	観察データの尤度とは？	108
3.5	事後分布とは？	108
3.6	予測分布とは？	110
3.7	交換可能性とは？	111
3.8	尤度原理とは？	114
<b>4</b>	<b>ベイズ流臨床試験の計画</b>	<b>115</b>
4.1	ベイズ流試験は健全な臨床試験デザインから始まる	115
4.2	適切な評価項目の選択	116
4.3	他の重要な情報の収集：共変量	117
4.4	比較するものを選ぶ：対照	117
4.5	評価項目についての初期情報：事前分布	118
4.6	他の試験から力を借りてくる：階層モデル	121
4.7	標本サイズの決定	125
4.8	ベイズ流デザインの動作特性の評価	127
4.9	ベイズ流の多重性調整	128
4.10	ベイズ流の適応的デザイン	129
<b>5</b>	<b>ベイズ流臨床試験の解析</b>	<b>132</b>
5.1	事後分布の要約	132
5.2	仮説検定	132
5.3	区間推定	132
5.4	予測確率	133
5.5	中間解析	136
5.6	モデルの点検	136
5.7	感度分析	137
5.8	決定分析	138
<b>6</b>	<b>市販後調査</b>	<b>139</b>
<b>7</b>	<b>技術的詳細</b>	<b>140</b>
7.1	プロトコルに含めるのが望ましい情報	140
7.2	動作特性を得るためのシミュレーション	144
7.3	モデル選択	147
7.4	事後予測分布を用いた交換可能性の点検	148
7.5	計算	149
	参考文献	152

第1部

# 臨床試験デザイン 入門

---



# 1 医学研究と統計学

最近、データサイエンス、ビッグデータ、データマイニングなどの統計学に関連したカタカナ用語をよくみかける。どのような用語を使うとしても、データを役立つ情報に変えるためには統計学の論理が必要であり、正しく計算することはソフトウェアにまかせるとしても、結果を正しく解釈した上で、正しい判断を行う能力が必要である。統計学とは、

- ・ 賢明な意思決定を行うための科学
- ・ 不確実性を数量化し表現する方法
- ・ すべての学問のための学問
- ・ データから情報へのはしごを一段昇るための論理

であると言われている<sup>1)</sup>。従って、多くの学問分野で統計学は利用されているが、理解不足による誤用も後を絶たない。

統計学は、大きく二つに分類される。一つは「記述統計学」であり、そこでは位置の尺度（算術平均、中央値）、バラツキの尺度（分散、標準偏差、範囲）などを算出し、集団としての特徴を記述するためにデータを整理・要約する。もう一つは、「推測統計学」であり、そこではデータを分析し、母集団についての推測を行う。推測統計学には、大まかに言うと、頻度流とベイズ流と呼ばれる二つの接近法（approach）が存在する。頻度流統計学の代表的な手法が仮説検定と推定である。詳細は後述するが、簡単に言うと、頻度流では確率を「仮想的繰り返しに基づく相対頻度」と解釈し、パラメータ（関心のある母集団の特性）を「未知の定数」と考える。一方、ベイズ流統計学ではパラメータ

の不確実さをそのまま確率分布として定式化する。

## 1 医学・医療と臨床試験－基礎研究と臨床研究の違い

医学は普遍的な真実を追求する科学の一分野である。一方で、医療には個人に対して最適な技術を選択して適用することが要求される。医学が生み出した技術の評価（技術評価）は、主に統計学に基づく科学的方法を用いて技術を相対的に評価し、医学から医療に橋を架ける役目を担う（図1）。

統計的方法を医学や医療の世界に導入する際には、二つの大きなギャップを理解しておく必要がある。医学と技術評価の間のギャップは、「決定論」対「非決定論（確率論）」である。コラム「確率の哲学」に示すように、医学の世界では長い間にわたって決定論的な思想が支配的であった。技術評価と医療の間のギャップは、意思決定の対象に関連しており、「対集団の確率（統計学）」対「対個人の確率」である。例えば、ある医薬品の製造販売を許可すべきかどうかという判断は、国民という集団に対するベネ

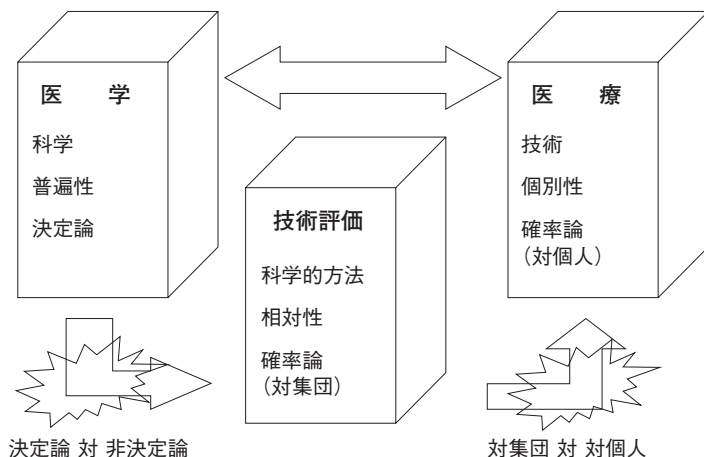


図1 医学・医療と技術評価<sup>2)</sup>

フィットとリスクのバランスで決定される。従って、その決定は「対集団の確率」に基づくが、医療現場では個人に対するベネフィットとリスクを評価しなければならない。がんの手術後、もし医師が「術後5年以内に再発する可能性は30%です」と言ったとき、その30%はある集団から推定された割合に基づいたものである。しかしながら、それを聞いた患者は「私は再発するか(100%)、しないか(0%)のどちらか」とこの確率を解釈するしかない。降水確率や地震発生確率なども同様の例であるが、これは「確率とは何か」という問題から生じている。

臨床試験の方法論は、英国の統計学者ロナルド・フィッシャー(1890-1962)が考案した統計的実験(技術的実験)の方法論を基礎として発展してきた。人工的に作った純粋な条件下での因果関係を立証しようとする科学的実験に対して、統計的実験は以下の特徴を有する<sup>3)</sup>。

- ・ 実験の場は、現実の応用の場に近い状況に設定される。
- ・ 結果の分析には誤差の存在を前提にしなければならない。
- ・ いくつかの因子を同時に変化させて結果を見る必要があることがある。
- ・ 目的は、何らかの基準によって現実の場において最も良い結果が得られるような条件を求めることである。

実験の場で生じる様々な種類の誤差を伴うデータを扱うためには統計的方法が不可欠である。結果を再現性によって保証することが比較的可能な基礎医学・生物学研究(科学的実験)と異なり、臨床試験(統計的実験)は、同じ条件で繰り返すことが困難で、結果の再現性による評価が難しいため、デザインと手続きの正しさでしか結果を保証できない

ロナルド・フィッシャーは次のように述べている<sup>4)</sup>

「同じだけの時間と労力をかけたとしてもデータ収集の過程、または実験計画を厳密に検討しているか否かによって、得られる収穫は10倍から12倍にもなる。実験終了後に統計学者に相談を持ちかけるのは、統計学者に、単に死後診察を行って下さいと頼むようなものである。統計学者はおそらく何が原因で実験が失敗したかという実験の死因について意見を述べてくれるだけであろう」。

すなわち、計画段階から統計学の専門家（統計家）が参画していれば、質の高い臨床試験を実施できる可能性は高くなる。ちなみに、臨床試験の統計的側面に責任をもつ統計家は、試験統計家（trial statisticians）と呼ばれる。

#### ◎文 献

- 1) ラオ CR (藤越康祝ほか訳). 統計学とは何か: 偶然を生かす. 筑摩書房. 2010.
- 2) 手良向聡. 臨床試験の統計的デザイン. In: 川上浩司編. 遺伝子医学 MOOK 別冊 はじめての臨床応用研究. メディカルドゥ社. 2010: 111-117.
- 3) 竹内 啓. 統計学的な考え方ーデザイン・推測・意思決定ー. 統計学の基礎 II. 岩波書店. 2003: pp1-53.
- 4) Fisher RA. Presidential address. Indian Journal of Statistics 1938; 4: 14-18.

### ☪ ☪ ☪ 「確率の哲学」

決定論的な思想が支配していた時代に、ピエール・シモン・ラプラス（1749–1827）は「確率の理論は、実は、計算された常識にほかならない」と、ラプラスの悪魔と呼ばれる考えを示した。クロード・ベルナール（1813–1878）は「実験医学序説」の中でデテルミニズム（決定論のフランス語）を提唱し、「統計学に立脚している限り、医学は永久に推測科学に止まるであろう」と述べた。決定論とは、「ある系の未来と過去の状態は、しかるべき方程式と、その系の現在の状態に関する知識から予測できるという、科学の世界にある哲学的な原理」である。

確率の解釈には、以下に示す大きく 2 通りの解釈が存在する。

1. 認識論的解釈 — 1) 論理説  
— 2) 主観説
2. 客観的解釈 — 1) 頻度説  
— 2) 傾向説
- (3. 多元的解釈 — 間主観説)

一つは認識論的解釈であり、確率が人間の知識や信念に関係するものと考えられる。その中の一つが論理説であり、「確率は合理的信念の度合い」と解釈する。パートランド・ラッセル (1872–1970)、ジョン・メイナード・ケインズ (1883–1946)、ルドルフ・カルナップ (1891–1970) らがこれを支持した。もう一つが主観説であり、「確率は特定の個人が持つ信念の度合い」と解釈する。フランク・ラムジー (1903–1930)、ブルーノ・デ・フィネッティ (1906–1985)、レオナルド・サベジ (1917–1971) らがこれを支持した。なお、サベジがベイズ流統計学を再評価したことなどから、主観説はベイズ流統計学と大きく関連しているが、ベイズ流統計学を利用している統計家 (いわゆるベイズアン) のすべてが主観説を支持しているわけではない。

もう一つの解釈が客観的解釈であり、確率が客観的な物質世界の性質であると考えられる。その中の一つが頻度説であり、「ある集合において観察された相対的頻度の極限值」を確率とする。ジョン・ヴェン (1834–1923)、リヒャルト・フォン・ミーゼス (1883–1953) が支持した。この頻度説は頻度流統計学と密接に関係している。もう一つが傾向説であり、「ある条件がほぼ確率に等しい頻度を生む傾向を備えている」と考える。カール・ポパー (1902–1994) が提唱した。そのほかに多元主義的解釈として間主観説「確率はある社会集団が持つ信念の度合い」というものもある。

意思決定者の立場、または意思決定の場面によって確率の解釈は異なってくると考えた方がよいかもしれない。ここでは、医薬品等を承認する立場 (社会の立場) と治療を行う立場 (医師の立場) を対比して考える。社会の立場では、その集団に対する医薬品等の効果を客観的確率と解釈する頻度説と相性がよく、頻度流統計学の手法、特に長期的なリスクを一定の確率で制御しなければならないという第 I 種の過誤の制御という考え方を受け入れることが自然である。一方、医

師の立場は、その個人に対する医薬品などの効果を客観的確率と自らの経験に基づく主観的確率の混合と考える傾向があり、主観説と相性がよい。極端に言うと、患者は個々に異なることから常に単一事象に対する確率を扱っており、同一試行の長期的繰り返しを前提とした頻度流の考え方に違和感を持つのではないかと思う。しかも、医師は利用可能な方法から最善と考えるものを選択しなければならないことから、ある仮説の正しさを確かめる仮説検定の問題ではなく、多くの仮説から一つを選択しなければならないという選択の問題に直面しているのではないかと考える。

#### ◎参考文献

---

- ギリース D (中山智香子訳). 確率の哲学理論. 日本経済評論社. 2004.  
タバク J (松浦俊輔訳). はじめからの数学 (4) 確率と統計. 青土社. 2005.

## 2 臨床試験の計画

### 1 臨床試験実施計画書

臨床試験実施計画書（clinical study protocol）は一般に略して「プロトコル」と呼ばれ、臨床試験を行う上で最も重要な文書である。その作成の目的は、

- ・ 計画を明文化し、科学的・倫理的審査を受けるため
- ・ 実施関係者に内容・手順などを伝達し、試験の質を一定に保つため
- ・ 計画と実施結果の差異を明らかにすることにより、試験の質を評価するため

である。表 1 にプロトコルの項目立てを示す。「目的」と「背景と根拠」は論文の Introduction に相当する重要な部分であり、論文の Method に相当する試験の骨格は、誰に対して、何を行い、何を評価するか、すなわち項目としては、

- ・ 対象集団（または適格規準）
- ・ 治療計画（または介入計画）
- ・ 評価項目

の三つである。また、「登録・割付」、「統計的事項」の二つの項目は臨床試験デザインに最も関連しており、統計家の関与が必要な部分である。

表1 プロトコルの項目立て

0. 概要	11. 評価項目
1. 目的	12. 統計学的事項
2. 背景と根拠	13. 倫理
3. 本試験で用いる基準・定義	14. 個人情報保護と患者識別
4. 治験薬／治験機器情報	15. 治験実施計画書内容の変更
5. 対象集団、適格基準	16. 補償
6. 登録、割付	17. 利益相反と資金源
7. 治療計画	18. データの品質管理、品質保証
8. 有害事象の評価・報告	19. 研究内容の発表
9. 臨床検査、観察、調査項目・スケ ジュール	20. 文献
10. データの収集	21. 付録

## 2 探索的臨床試験と検証的臨床試験

臨床試験の性格は探索的試験と検証的試験に大きく分けられる。通常は、薬物動態、安全性、用法・用量などを探索・評価する探索的試験を経て、有効性・安全性の確固たる証拠を提示するために検証的試験が行われる。ただし、個々のいかなる試験も検証的側面と探索的側面を持ち、プロトコルには、各試験について検証的な証明として用いられる側面と、探索的解析のためにデータを提供する側面とを明確に区別しておくべきである<sup>1)</sup>。探索的解析から得られた結果は仮説に過ぎず、その仮説は検証的試験によって確認しなければならない。

## 3 臨床試験デザイン

臨床試験はデザイン・計画の段階から始まり、試験実施・データ管理・モニタリングを経て、データ解析・報告書作成に至る。この各ステップが統計的方法を必要としている。臨床試験は、良いデザインおよび正しい遂行がなければ、台無し（例えば、結論の出ないまたは誤った知見を導く）になり、さらには大惨事（例



えば、不必要に多数の患者が毒性または死を被ることを引き起こす)になる可能性さえある。臨床試験は効率的かつ倫理的であるべきで、資源を節約し、より多くの患者に恩恵を与え、より迅速に正しい結論を引き出し、結果として不必要な毒性をより少なくすべきである<sup>2)</sup>。

試験デザインとは、

- ・ 統計的仮説
- ・ 対照の選択
- ・ ランダム化
- ・ 盲検化
- ・ 並行群間比較, クロスオーバー, 用量漸増など (狭義の試験デザイン)
- ・ 中間モニタリング計画
- ・ 標本サイズ設定

を含む。

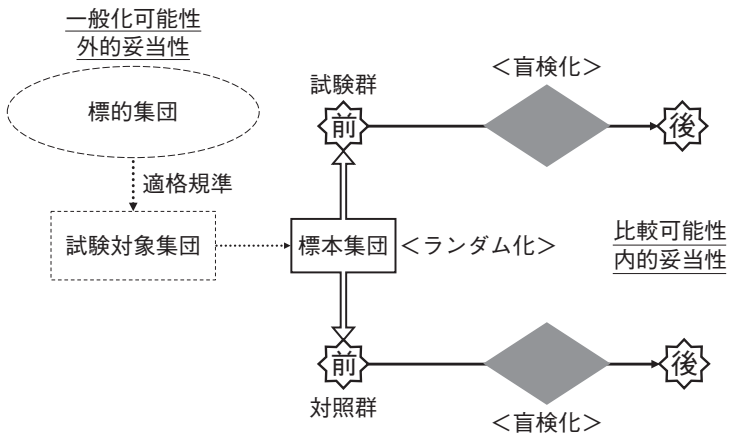


図 2 ランダム化対照試験

検証的試験の標準はランダム化対照試験（RCT: randomized controlled trials）である（図2）。ランダム化という操作を行うことにより、試験の比較可能性（内的妥当性）を保証する。一方、試験対象集団が適格規準によって厳密に規定されることから、標的集団（結果を適用したい集団）への一般化可能性（外的妥当性）は保証されていない。

## 4 評価項目

評価項目とは、「試験の目的に関連する仮説を検証するうえで臨床的に意味があり、客観的に評価できる観察・検査項目またはそれらの合成指標」である。結果尺度、反応変数と呼ばれることもある。なお、評価項目は、「生存期間」や「腫瘍反応の有無」など各被験者について測定される項目として定義すべきである。「5年生存割合」や「反応割合」など集団から計算される尺度は、推定の対象となる「関心のあるパラメータ」であり、評価項目と呼ぶと混乱を招く。試験目的に合致した評価項目を設定する必要があり、その試験で最も重要なものを主要評価項目、それ以外を副次評価項目と呼ぶ。

評価項目を選択する際には、それが臨床評価項目であるのか、代替評価項目であるのかをよく吟味しなければならない。臨床評価項目（真の評価項目とも呼ばれる）とは、「患者がどのように感じ、あるいは機能し、どのくらい生存しているかを反映する特性あるいは変数」である。バイオマーカーは「正常な生物学的プロセス、病態形成プロセスあるいは治療的介入に対する薬理学的反応の指標として客観的に測定および評価されるある特性」であり、代替評価項目は「バイオマーカーのうち、臨床評価項目の代わりになることが意図されたもので、疫学、治療学、病態生理学または他の科学的根拠に基づき、臨床上的便益・害の有無を予測することが期待されるもの」と定義される。すべての疾患と医療

表 2 代替評価項目の候補

代替評価項目	臨床（真の）評価項目
血中コレステロール値	心疾患の発生
血圧値	心疾患，脳血管障害の発生
血糖値	糖尿病合併症の発生
腫瘍縮小効果	癌死
不整脈	心臓死
心機能指標	心不全の発生
血管狭窄度	狭心症，心筋梗塞
眼圧	緑内障の視野狭窄

技術の組み合わせについて臨床評価項目を用いた試験を行うことが理想ではあるが，コストや時間の制約によりそれは現実的に不可能である。従って，代替評価項目を用いなければならない状況はまれではない。代替評価項目の候補と臨床評価項目との対応の一覧を表 2 に示す。

疾患，代替評価項目，臨床評価項目の理想的な関連は図 3 のように，代替評価項目が疾患過程の唯一の因果経路上に存在し，臨床評価項目への治療効果が，完全に代替評価項目への効果を介

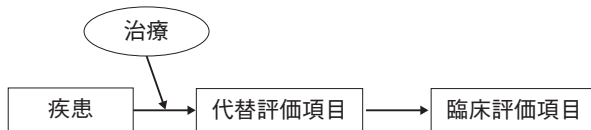


図 3 疾患，評価項目と治療との関係（理想）

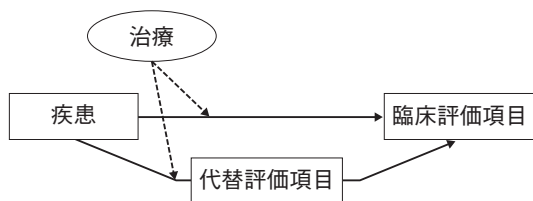


図4 疾患，評価項目と治療との関係（現実）

して伝達されるような場合であるが，現実には図4のように，治療が代替評価項目を介す経路と介さない経路の両方に作用する可能性があるため，代替評価項目での結果は慎重に解釈する必要がある。

臨床試験で用いられる評価項目のほとんどは以下の三つの変数の型に分類される。

#### 連続変数

臨床検査値，身体的特徴など。関心のあるパラメータは平均値の差など。

#### 分類変数（2値）

腫瘍反応の有無，感染の有無など。関心のあるパラメータは反応割合の差，リスク比，オッズ比など。

#### 時間イベント変数

生存時間，無イベント生存時間など。関心のあるパラメータは生存割合の差，ハザード比など。

それぞれの変数の型に対応する統計解析手法を表3に示す。ランダム化対照試験では，比較可能性が保証されているので単純な検定を用いればよいが，観察研究では交絡の調整が必要であり，複雑な手法（層別検定や回帰モデルなど）を用いる必要がある。

表 3 変数の型別の標準的な統計解析手法

目的	連続変数	分類変数	時間-イベント変数
分布の記述	ヒストグラム, 箱ヒゲ図, 散布図	ヒストグラム, 分割表	生存曲線 (カプラン・マイヤー法)
要約統計量	平均, 分散, 中央値, パーセント点, 相関係数	頻度, 一致度, 相関係数	x 年生存率, 中央生存期間
検定	t 検定, 分散分析, ウィルコクソン検定	$\chi^2$ 検定, フィッシャー正確検定, ウィルコクソン検定	ログランク検定, 一般化ウィルコクソン検定
検定 (層別)	共分散分析	マンテル・ヘンツェル検定	層別ログランク検定
回帰モデル	一般線形モデル	ロジスティックモデル	コックス比例ハザードモデル

## 文 献

- 1) 厚生省医薬安全局審査管理課長:「臨床試験のための統計的原則」について, 臨床評価 27: 161-206, 1999.
- 2) Yin G (手良向聡, 大門貴志訳). 臨床試験デザイン-ベイズ流・頻度流の適応的方法. メディカル・パブリケーションズ, 2014.

## 【著者略歴】

手良向 聡 (てらむかい さとし)

京都府立医科大学大学院医学研究科生物統計学 教授

京都府立医科大学研究開発・質管理向上統合センター 副センター長 (生物統計・データ  
マネジメント部門長)

## 【著書】

実践医学統計学 (統計ライブラリー) (共訳, 朝倉書店, 2005)

Rによる統計解析ハンドブック (共訳, メディカル・パブリケーションズ, 2010)

臨床試験デザインーベイズ流・頻度流の適応的方法 (共訳, メディカル・パブリケーションズ, 2014)

## なぜベイズを使わないのか!? 臨床試験デザインのために

---

2017年8月15日 第1版第1刷©

著 者 手良向聡 TERAMUKAI, Satoshi

発 行 者 宇山閑文

発 行 所 株式会社金芳堂

〒606-8425 京都市左京区鹿ヶ谷西寺ノ前町34番地

振替 01030-1-15605 電話 075-751-1111 (代表)

<http://www.kinpodo-pub.co.jp/>

組版印刷 西濃印刷株式会社

製 本 藤原製本株式会社

---

落丁・乱丁本は直接小社へお送りください。お取替え致します。

Printed in Japan

ISBN978-4-7653-1723-8

**JCOPY** <(社) 出版者著作権管理機構 委託出版物>

本書の無断複写は著作権法上での例外を除き禁じられています。複写される場合は、その都度事前に、(社) 出版者著作権管理機構 (電話 03-3513-6969, FAX 03-3513-6979, e-mail: info@jcopy.or.jp) の許諾を得てください。

●本書のコピー、スキャン、デジタル化等の無断複製は著作権法上での例外を除き禁じられています。本書を代行業者等の第三者に依頼してスキャンやデジタル化することは、たとえ個人や家庭内の利用でも著作権法違反です。