

EBNのための研究計画

介入研究の原理と方法

林 邦彦・高木廣文

EBN のための研究計画

介入研究の原理と方法

林 邦彦 HAYASHI Kunihiro

群馬大学医学部保健学科

高木廣文 TAKAGI Hirofumi

新潟大学医学部保健学科

Key Point

- 介入研究は比較研究と非比較研究に大別できる。
- 非比較研究とは、比較対照（コントロール）を研究内にもたず、調べたい治療法についてのみ検討する研究デザインで、「症例報告」「瀕死研究」「オープン試験」「追跡研究」ともよばれる。
- 比較研究とは、研究のうちに比較対照をもつ研究デザインで、群間で比較する「並列群間比較研究」と群内（個人の中）で比較する「クロス・オーバー比較研究」などがある。
- 比較試験における「比較」とは、検討する要因以外の要因を同じにすることである。
- 検討する要因以外の予後要因を同じにするために、ランダム化を行う。
- 研究にかかわる対象者・評価者・データ分析者の心理的なバイアスを除く方法として「盲検」が行われる。
- 対象者だけに盲検をかけることを「単盲検」、評価者にもかけることを「二重盲検」、分析者にもかけることを「三重盲検」という。
- 介入研究の計画に必要な統計学的事項として、①研究の目的・仮説・対象集団の明確な定義、②エンドポイントの選択、③研究デザインの選択、④比較対照の選択、⑤必要最小例数の設定、⑥CRFの設定、⑦統計解析法の決定、がある。

臨床試験は、ヒトを対象にしたある種の実験である。よくも悪くも、誰かが新しい治療法に曝露する最初の例にならざるを得ない^{1, 2)}

オースティン・ブラッドフォード・ヒル卿

介入とは

前号まで観察研究の計画の特徴をみてきたが、本号からは介入研究の計画に話を進める。

観察研究と介入研究を区別する「介入（intervention）」とは、いったい何を指すのだろう。おそらく、臨床現場では手術や投薬といった医療行為のことを、介入とよんでいるだろう（interventionの語を介入

とは訳さずに、単に、治療法とか医療行為と訳すことも多い）。ところが、疫学でいう介入は意味が異なるので注意を要する。

疫学での介入とは、「研究者による予防法や治療法などの意図的な変更（intentional change）」³⁾のことを指す。観察研究では、研究対象になっても、日常生活、治療法、予防法などが、研究のために意図的に変えられることはない。たとえ、手術や投薬といった医療行為に関する研究であっても、研究のために治療法に変更が生ずること（もしくはその可能性）がなければ、その研究は観察研究に分類される。

一方、介入研究では、何らかの研究者の意図的な働きかけが必ず存在する。そのため、「実験的研究」

ともよばれる。特に、治療法の評価のための実験的研究を、「臨床試験 (clinical trial)」という。ヒトを対象に行われる研究に、実験という名前を冠するのに躊躇があるかもしれない。しかし、科学一般で、研究者が要因に操作を行い、その後の経過を観察する研究のことを「実験」とよぶ。臨床試験は、研究者が患者の予後に影響を与えると考える要因に意図的に介入を行い、その後の変化を観察するという、まさに実験の形式をとる研究デザインである。

最初のランダム化比較臨床試験といわれるストレプトマイシンの臨床試験の実施など、臨床試験の研究デザイン確立に大きく貢献した医学統

計学者、ヒル卿⁴⁾は、「治療法の評価は、実験という形で行わざるを得ない。実験とは、ある特定の患者にその治療法を割り付け、その後に起きる事象を観察記録することである。本来は解釈が困難な事象について、事前と事後とを区別することによって成り立っている」^{1, 2)}と述べている。植物の種の発芽実験において、要因と考える水を与えること（介入）により、その後（事後）の種の変化を観察し、水分と発芽の関連を検討するのと、原理は同じである。

通常、研究者の介入は、対象者個人単位に行われる。しかし、ワクチンやマス・スクリーニングの効果の評価などでは、地域、学校での学年といった集団単位で介入を行う場合もある。このような研究は「field trial」とか「cluster trial」とよばれ、介入研究に分類される。

比較と非比較

①に、いくつかの介入研究の例を示す。介入研究は、

1) 非比較研究 (症例報告, 漸踏み研究, オープン試験, 追跡研究)



2) 比較研究

・並列群間比較研究



・クロス・オーバー比較研究



① 介入研究の例

比較研究と非比較研究に大別できる。①の上段は非比較研究の例である。比較対照（コントロール）を研究内にもたず、調べたい治療法についてのみ検討する研究デザインである。新しい治療法の開発などでは、まず1例ずつ安全性を確かめながら有効である症例経験を増やしていくことが行われる。こういった比較対照をもたない非比較研究は、「症例報告」「漸踏み研究」「オープン試験」「追跡研究」などともよばれる。しかし、数多くの非比較研究を実施しても、その新規治療法の効果を実証的に示すことはできない。たとえ実験的研究であっても比較対照がないと、事後に起きた事象が確かに要因への介入のせいだと結論することは難しい。水を与えた種のいくつかが発芽したとしても、水がないときには種は発芽しないという情報がないと、確かに水分が発芽に必要な要因だと結論できないだろう。比較対照がないと、「僕は患者にこう言うんだよ。風邪をひいたときはきちんと治療をすれば1週間で治る。治療しなくても、やはり1週間で治るって」^{5, 6)}といった現象があっても

認識できない。

研究のうちに比較対照となるものが存在すれば、「比較研究 (controlled study)」とよぶ。現在では、多くの治療法評価において、最終的には比較介入研究の研究デザインによって検証を行う。比較研究の必要性をみるため、過去の実験的研究の事例をみてる。

介入研究における比較の歴史的事例

明治初期には、わが国において脚気が大きな問題であった。脚気はビタミンB₁欠乏により起こり、重症の場合には死に至る疾患であるが、当時はビタミンの存在すら知られておらず、脚気の原因について脚気菌説など諸説があった。

そのような状況のなか、海軍の高木兼寛は脚気の原因は食事にあると考えて、ある実験を行った⁷⁾。遠洋航海に出かける海軍練習船「筑波」において、麦と肉を積み込み、白米中心であった食事を西洋食に変えた。すると、その航海ではほとんど脚気患者は発生しなかった。白米食であった前年の練習船「龍驤」の航海では333人中169人に脚気が発生したことを考え合わせると、この実験によって、食事を白米からパンや麦飯に変えることが有効であることを示したものと考えた。

ところが、陸軍の森林太郎（鷗外）らは、この高木の実験に猛烈な反論を行った。それは、前年の「龍驤」の航海での脚気患者の発生成績は比較対照にはならず、高木の実験には比較対照とするものがないという主張であった。比較するなら、「一大兵団ヲ中分シテ、一半ニハ麦ヲ給シ、一半ニハ米ヲ給シ、両者ヲ同一ノ地ニ住マワシメ、ソノ他ノ生活ノ状態ヲ齊一ニシテ、食米者ハ脚気ニ罹リ、食麦者ハ罹ラザルトキハ、マサニ、ワズカニソノ原因ヲ説クベキノミ」と、食事以外のほかの要因を揃え、兵隊を2つ

に分けて同時に発生する脚気患者を、麦群と米群とで比較すべき（同時比較：①における並列群間比較研究のデザイン）というものである。彼らは、高木の実験では食事と脚気との因果関係が実証できないと考えた。

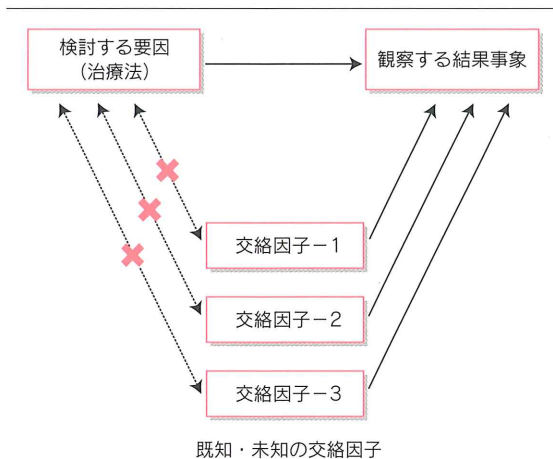
森らが主張した同時比較の実験の事例は、140年ほど遡り、1747年のイギリスにある。イギリス海軍に当時、流行したビタミンC欠乏による（もちろん当時はその原因はわからない）壊血病の例である。イギリス海軍の医師ジェームス・リンドが、実験的研究を行った⁷⁾。できるだけ似た12人の壊血病の水夫を選び、オートミール、マトンの肉汁など共通の食事を全員に与えたうえで、2人ずつ6群に分け、①リンゴ酒、②硫酸エリキシル、③食酢、④海水、⑤オレンジ・レモン、⑥ナツメグを各群に割り付け、その後の変化を観察した。他群に比較して、最も劇的な変化があったのはオレンジ・レモン群で、6日後には作業可能になり他群の患者の看護にあたった。つまり、6群の並列群間比較の研究デザインであった。

「ランダム」に割り付ける

「比較」における要点は、検討する要因以外の要因は同じにすることである。これを「比較可能性」という。

水分を与える種と与えない種の間で、温度・日光・養分といった発芽に影響しそうな他要因は同じ条件にする。前述の森林太郎の主張した研究デザインやジェームス・リンドの研究でも、同じような対象者について、注目している要因（脚気では米と麦、壊血病だと6つの処置法）以外の要因を同じにするため、比較対照を過去の成績に求めるのではなく、研究内に同時に対照群を設定している。

しかし、この並列群間比較研究のデザインにおいても、対象者個人間の予後要因をすべて完全に同じにすることは困難である。そこで、検討したい要因



✕：ランダム化による関連の排除

② ランダム化の特徴：交絡因子の影響の排除

以外の要因を揃えるために、いくつかの方策が利用されている。最も多く利用されている方法が、「ランダム化（ランダム割付，無作為割付，randomization, random allocation）」とよばれる方法である。この方法は，対象者に治療群をランダムに割り付けていく。対象者がある程度の人数になれば，今，割り付けている要因以外のいずれの予後要因も各群で均等になっていく。この方法を使えば，本誌前号・前々号で説明した交絡因子^{8, 9)}の影響を排除することができ，既知の予後要因はもちろん未知の予後要因があったとしても，比較への影響を排除できる（②）。

前号の例では，70歳以上の患者では新看護法Nが多く，70歳未満の患者では標準看護法Sが多くなり，観察された看護法N群とS群の結果事象の比較では，看護法の効果と年齢の効果が交絡していた。そこで，ランダム化を行って研究すれば，②に示すように，今，調べたい看護法と，年齢など他要因との間に関連が生じない。結果事象に影響を与える要因がいくつかあっても，また，そのような要因の存在がわからなくても，ランダム化を行えば，今，調べたい要因を独立させて比較することができる。

通常は，各治療法群で同じ対象者数になるように

割り付ける。各対象者からみると，いずれの治療法に割り付けられる確率も同じとなる。治療法群数が2であれば各治療に割り付けられる確率は1/2，3であれば1/3となる。実際のランダム割付は，乱数表，乱数サイコロなどを使っても行えるが，現在はコンピュータによって乱数を発生させることが多い。医薬品の評価のための臨床試験では，後述する盲検の手法を併用して，あらかじめランダムに割り付けられた薬剤を入れた箱をその番号順に用いることで，ランダム化が行われる。また，多施設共同で実施される比較研究では，割付センターを設けて，全施設における対象者を必ず中央登録することを行う。該当する対象者の組み入れに先立ち，各施設は，電話，FAX，電子メールなどを利用して，割付センターへ連絡する。そして，割付センターが一括してランダム割付を行い，各対象者への治療群を決めていく。中央登録を行うことは，各施設の対象者組み入れ基準を一括して確認できるほか，予後要因によって患者を層に分け，層内でランダム化したり，各治療群の間で複数の予後要因の違いを最も小さくなるようにランダム化できるなどの利点をもつ。

かつてはランダム化の方法として，患者の組み入れ順に（治療法A，治療法B，治療A，治療Bといったように）交互に割り付ける，患者の名前のインisialによって割り付ける，封筒に割付指示票を入れ対象者組み入れごとに開封していく封筒法とよばれる方法なども使われていたようである。しかし，後述する盲検が併用しにくく，正確なランダム化が実施できないため，現在はほとんど用いられていない。

群間比較と群内比較

対象個人間の予後要因の差をなくす別の方法として，比較を群と群の比較ではなく，個人のなかでの比較にすることも考えられている。①の，「クロス・

オーバー比較研究（チェンジオーバー比較研究）」とよばれる比較デザインである。この方法は、比較的安定した病状の疾患で、治療法への反応が治療を止めるとともに戻るような場合に適しているが、治療の時期や順番に効果があるような疾患や治療法では適さない。

クロス・オーバー比較研究によって個人内比較をするには、時期や治療法の順番に効果がないことを確認できるデザインが必要である。治療法が2つで治療時期が2つ、治療法が3つで治療時期が3つなどでは、「ラテン方格」という割付方法を用いる^③。このような割付でなく、対象者すべてに、治療法A→治療法Bの順番で治療すると、治療法の効果と、時期や順番の効果が区別できなくなる。かつて、シェーバーの宣伝で、朝、自宅では国産シェーバーAでひげを剃ったと思われるサラリーマンをつかまえて、シェーバーBでひげを剃ってもらい、その効果をみ

る様子が流れていた。あの場合も、2治療2時期（正確には2ひげ剃り2時期）のクロスオーバー比較研究であれば、朝、自宅ではシェーバーBでひげを剃ったサラリーマンをつかまえ、国産シェーバーAでひげを剃ってもらい、その効果もみなくてはいけない。本来、ひげは時間に伴い伸びる性質をもつので、もともとクロスオーバー比較研究には向かないかもしれない。

異なる時期での個人内比較のほかにも、クロスオーバー比較研究と同様のデザインで、片眼ずつ治療法を変え両眼で比較する研究や、皮膚反応を左右上下で比較するパッチテストなど、いくつかの個人内比較の形式をとる研究がある。

盲検をかける

医学研究において、研究対象者も、観察評価者も、またデータの分析者も、ヒトである以上は何らかの心理的バイアスを受ける。このようなバイアスは、「プラセボ効果」「ホーソン効果」「ラベリング効果」などとよばれる。治療している・治療されている、観察している・観察されている、と意識するだけで反応が変わることはよく経験する。特に、今、検討している治療法に思い入れがあればあるほど、また評価する項目が主観的であればあるほど、この心理的なバイアスは大きくなるのかもしれない。適切な治療を受けていると感じると症状がより改善する患者の経験を、多くの読者はもっていると思う。評価者においても、症状改善など同じ事象を観察していても、効果があると信じている治療法では過大に、そうでない治療法では過小に評価するかもしれない。また、データ解析など分析者においても、2つの治療法には差があるはずといった信念があると、無意識のうちに統計学的有意差が出やすい統計解析手法の採用やデータ処理を行うかもしれない。

自分が研究の対象になっている、自分が評価観察

2治療法2時期：

第1期	第2期
治療法A →	治療法B
治療法B →	治療法A

3治療法3時期：

第1期	第2期	第3期
治療法A →	治療法B →	治療法C
治療法B →	治療法C →	治療法A
治療法C →	治療法A →	治療法B
治療法A →	治療法C →	治療法B
治療法B →	治療法A →	治療法C
治療法C →	治療法B →	治療法A

③ クロス・オーバー比較研究における要因順序の割付の例（ラテン方格の利用）

をしている、自分が研究のデータ解析をしていると意識することによる反応は避けることはできないことかもしれない。問題は、治療法を比較する研究において、いずれかの治療法に偏ってその影響が出てしまうことである。このような心理的なバイアスによる反応の違いは、研究対象者、観察評価者、データ分析者が、個々の対象者がどちらの治療法を受けているのかを研究終了まで知らなければ防ぐことができる。これを「盲検 (blind, mask)」という。対象者だけに盲検をかければ「単盲検」、評価者にもかければ「二重盲検」、分析者にもかければ「三重盲検」という (4)。

医薬品評価の臨床試験においては、この盲検はプラセボ (placebo) を用いて行われる。たとえば、錠剤である薬剤Aの効果を無治療と比較して検討したいとしよう。薬剤A群にはその錠剤を服用してもらい、無治療群には、概観、におい、味、重さ、色、つやなどがそっくりだが成分が入っていないプラセボ錠剤を服用してもらう。そうすると、薬剤A群と同様に無治療群でも、治療を受けている、薬剤を飲んでいると意識するための反応 (プラセボ効果) が出るので、両群の効果の差が真の薬剤成分の効果の大きさとなる。では、薬剤Aと薬剤Bの比較ではどのように盲検をかけられるのか。薬剤A群では「薬剤A」と「薬剤Bのプラセボ」、薬剤B群では「薬剤Aのプラセボ」と「薬剤B」を服用してもらえば、見た目ではいずれの群も薬剤Aと薬剤Bの両方を服用しているようになり、実際にどちらの薬剤成分が含まれているかはわからない。この方法はプラセボを二重に使うことから「ダブル・ダミー法」とよばれる。

手術などの治療法評価研究では、医薬品の臨床試験のようにプラセボを用いるわけにはいかない。しかし、このような場合でも術者と評価者を別の医師にして、治療の評価者には研究終了時までどちらの術式を受けた対象者かを知らせないといった形で、盲検がかけられている。意識的であれ無意識であれ

	対象者	評価者	分析者
非盲検 non-blind			
単盲検 single blind	○		
二重盲検 double blind	○	○	
三重盲検 triple blind	○	○	○

○：盲検をかける箇所

④ 盲検の種類

比較に混入しそうな心理的バイアスを防ぐには、どのような比較研究においても盲検が最も有効な手段といえる。治療法を知っていると、「あれ？ この治療法ならこの検査値はもっと高くなると思うけど、もう一度検査してみましょう」と、評価者が予想する値になるように検査が何度か行われることもありうるだろう。たとえ、観察測定する項目 (エンドポイントとよぶ) が客観的な数値として機械的に測定されるものであっても、このようにバイアスの影響を受ける可能性があるので盲検がかけられている。

介入研究の計画における統計学的事項

最後に、介入比較研究の計画で必要となる統計学的事項を以下に列記する。

- 1) **研究の目的・仮説・対象集団の明確な定義**：特に仮説は、「治療法Aは治療法Bより効果がある」といった医学的な仮説のほか、明確な統計学的仮説の設定が計画時に必要である。
- 2) **エンドポイントの選択**：統計学的な観点から、どんな尺度を用いて何を測るかの決定。また、代替エンドポイントを用いるには、ほかの知見から真のエンドポイントとの関連性などの統計学的評価も必要となる。
- 3) **研究デザインの選択**：群間比較にするのか個人内

(群内) 比較にするのかの決定や群の数の決定など。

4) **比較対照の選択**：研究デザインの決定にもかかわるが、何を比較対照とするかの決定も統計学に重要な事項である。それによって、研究の統計学的仮説が変わる場合がある。

5) **必要最小例数の設定**：仮説検証に必要となるサンプル・サイズの算出。対象者数があまりに小規模で仮説検証など期待できないような研究や、逆に、対象者数が大規模になりすぎ不必要に患者を研究対象にしてしまう研究にならないように、事前に、仮説

検証に必要と予想される対象例数を算出する。

6) **CRF の設計**：調査票、測定値記録、経過記録など「ケース・レポート・フォーム (CRF)」とよばれるデータ報告用紙を誤記入のないように設計することも、データ品質管理、データ統計処理において重要となる。

7) **統計解析法の決定**：データ収集前の計画の段階で、主要な統計解析法を決定して、研究計画書や解析計画書に明記しておく。

文献

- 1) Hill AB : Medical ethics and controlled trials. British Med J 1963 April ; 20 : 1043-1049.
- 2) Piantadosi S : Clinical trials-A methodologic perspective. New York : John Wiley & Sons, Inc ; 1997. p.11.
- 3) Last JM, editor : A dictionary of epidemiology. New York : Oxford University Press ; 1995. p.90.
- 4) Johnson NL, Kotz S : Leading personalities in statistical sciences-From the seventeenth century to the present. New York : John Wiley & Sons ; 1997. p.334-335.
- 5) アーサー・ヘイリー著、永井淳訳：ストロング・メディスン（上）。新潮文庫；1998。
- 6) 椿広計，藤田利治，佐藤俊哉編：これからの臨床試験－医薬品の科学的評価－原理と方法。朝倉書店；1999。
- 7) コールダー，R 著，佐久間昭訳：物語 人間の医学史。平凡社；1996。
- 8) 高木廣文，林邦彦：EBN のための統計の読み方5 交絡因子の影響を除く方法。EB NURSING 2002；2 (1)：94-100。
- 9) 林邦彦，高木廣文：EBN のための統計の読み方6 観察研究におけるバイアスを考える。EB NURSING 2002；2 (2)：96-100。

