ランダム化臨床試験の実践的方法論

Practical Methodology for a Randomized Clinical Trial

松 山 裕

Yutaka Matsuyama, Ph.D.

Key words :
clinical trial
randomization
biostatistics

Spinal Surgery 32 (1) 29-33, 2018

はじめに

主要医学雑誌に掲載された日本発の臨床研究論文数は世界第 25 位で、その順位は年々低下傾向にある¹⁾。確かに、インパクト・ファクターが高い雑誌での掲載数だけが研究の活発さや優秀さを表す指標ではないが、研究成果が得られると世界に向けて論文発表することは事実であり、研究者にとっては常識でもある。このような現状、あるいはディオバン事件をはじめとする臨床研究の不祥事の影響もあって、2018 年 4 月施行予定の臨床研究法の制定など、昨今のわが国における臨床研究を取り巻く環境が大きく変化しつつある。本総説では、ランダム化比較試験 (randomized controlled trial : RCT) をデザインする際のいくつかの重要な概念について概説する。

1. 臨床研究実施のために必要なもの

臨床研究を実施する際には、さまざまなもの (ヒト・モノ・カネ・ノウハウ) が必要である (Fig. 1)。研究目的が不明確で、進むべき針路が共有されておらず、出口を見据えないで実施している臨床研究の実施は絶対に避けるべきで、当該分野の unmet medical needs を満たす研究テーマ (clinical question : CQ) であることが大前提である。

研究デザインで重要な位置を占めるのが、研究の具体的な目的、方法などを文書化した研究計画書 (プロトコル) の作成である。基本的にはプロトコルと研究実施マニュアル (プロトコルには記載できない詳細な実施手順) をみれば、行われた研究を再現できるようにすべきであり、プロトコルには最低限、Table 1 に示す内容が記載

Fig. 1 臨床研究を実施するにあたって必要なもの

されていなければならない²⁾。プロトコル、実施マニュアル、調査票の作成などは、臨床研究者、疫学者、生物統計家、医療倫理の専門家、データマネージャーなどの複数の専門家がチームとして行わなければならない作業であり、その内容は、簡潔で、その研究に携わるすべての者にとってわかりやすいものでなくてはならない。プロトコルにおける「研究計画のまとめ」などは、対象者への説明の手助けとなるように、専門的な用語はできるだけ避け、平易な表現で書く必要がある。また、作成されたプロトコルは、研究開始前 (および、研究終了まで定期的に)、研究 (あるいは、倫理) 審査委員会 (institutional review board : IRB) により審査され、承認されなければならない。

プロトコルを作成したとしても、当然研究はこれからであり、プロトコルに則った研究の遂行が一番重要である。わが国では、臨床研究を実施・サポートしていくシステムが少なくともアカデミアにおいてはまだ十分整備

東京大学大学院医学系研究科公共健康医学専攻生物統計学分野 / Department of Biostatistics, School of Public Health, The University of Tokyo

連絡先 : 〒113-0033 文京区本郷 7-3-1 東京大学大学院医学系研究科公共健康医学専攻生物統計学分野 松山 裕 [Address reprint requests to : Yutaka Matsuyama, Ph.D., Department of Biostatistics, School of Public Health, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan]

Table 1 プロトコルの章構成と内容

項 目	内 容
研究計画のまとめ	研究全体の簡潔なまとめ
研究の目的	背景・理論的根拠・研究仮説・エンドポイント
研究の実施期間・場所	いつ、どこで、いつまで
研究対象集団の設定	適格条件・除外条件
インフォームド・コンセント	説明と同意の具体的方法
研究の実施方法	研究の型・割り付け方法・用量変更・併用治療・症例の取り扱い・測定項目・測定時期
統計解析を行う項目・方法	エンドポイントの設定と具体的な解析方法
研究に必要なサンプルサイズ	必要な対象者数算出の根拠
研究の組織・責任の所在・連絡先	研究組織一覧・連絡先

Fig. 2 データの正確度 (accuracy) と精密度 (precision)

がなされていないことが問題である。各大学病院・医療施設において、臨床研究支援センターに相当するものがつくられているが、生物統計家・データマネジャー・clinical research coordinator (CRC) などの人材不足、雇用体系の問題などもあり、臨床研究を有機的に動かしていくシステムづくりが喫緊の課題となっている。

2. 統計学的視点からみた臨床研究デザインに必要なこと

統計学で問題にするのはデータに内在するバラツキ (variation) であるが、そのバラツキを2種類に明確に分けて捉えることが重要である。バラツキの理由が同定できない (あるいは、あえてしない) ものを誤差的なバラツキ (ランダム誤差、純粋誤差) と呼び、一方、性別や年齢、疾患の重症度の違いなどによって結果の食い違いを説明できる意味のあるバラツキを系統的誤差 (バイアス) と呼ぶ。平均値を計算すれば真値に一致する結果は

正確 (accurate) であり、その真値の周りのランダムなバラツキが小さい結果は精密 (precise) と呼ばれる (Fig. 2)。Fig. 2a, b の場合は、平均値を求めれば真値に一致しているので、得られる結果にバイアスはない。この2つの違いは、真値の周りのランダムなバラツキの程度であり、できるかぎりサンプル数が多い研究ほど結果の精密度が高まる。一方、Fig. 2c, d の場合は、平均値を求めたとしても真値から系統的にずれているので、いくらサンプル数を多く集めたとしても得られる結果にはバイアスが生じる。

純粋な統計学では、バイアスがないことを前提に方法論の構築が行われるが (その意味では、サンプル数が多い研究ほど信頼度は高い)、ほとんどの臨床研究では得られたデータになんらかの理由によるバイアスが混在していることが多い。たとえば、日常診療においては、患者背景に依存してなんらかの意図をもって治療されるのが通常であり、得られる結果のバラツキはランダムなバ

Table 2 医学研究で頻用される統計解析手法（エンドポイントの型から分類）

解析の目的	エンドポイントの型		
	二値データ	連続データ	生存時間データ
分布の記述	頻度集計 分割表	ヒストグラム 平均・標準偏差 散布図・相関係数	Kaplan-Meier 法
単純な群比較	χ^2 検定・Fisher 検定 リスク差・リスク比（オッズ比）推定	t-検定 Wilcoxon 検定	log-rank 検定 ハザード比推定
層別解析	Mantel-Haenszel 法 調整リスクの推定	分散分析	層別 log-rank 検定 調整ハザード比の推定
回帰モデル	ロジスティック回帰 調整リスク比（オッズ比）の推定	重回帰分析 分散分析	Cox 回帰 調整ハザード比の推定

層別解析と回帰モデル：バイアス調整のための方法

ラッキーだけでは説明できないのが通常である。したがって、統計学的視点からみた臨床研究デザインに必要なことは、①誤差的バラツキを小さくすること（精度を高くすること）に加えて、②偏り（バイアス）を小さくすることが重要である。①のためには、サンプル数を増やすことが最も効果的であり、データの測定精度を上げることも有効である。一方、②のためには、サンプル数を増やしても効果がないため、デザイン上の工夫が必要となる。最も強力なバイアス減少の手段がランダム化（randomization）であり、盲検化は対象者に対する医師の評価、データ管理にバイアスが入ることを避けるための工夫である。また、バイアス補正のための統計解析手法を利用することも多い。医学研究でよく利用される統計解析手法をエンドポイント（アウトカム）の型の観点から分類したものを **Table 2** に示す。層別解析とは、バイアスを引き起こす要因（たとえば、性、重症度）別に結果を求め、そのサブグループでの結果を統合する解析手法である。一方、回帰モデルによる調整法とは、原因と結果との間に数学的なモデルを想定し、一度に多くのバイアス要因を調整する方法である。

3. なぜランダム化？

新薬や新治療法のヒトに対する効果を調べるための標準的実験方法として、ランダム化臨床試験が広く用いられている。なぜ治療法をランダム化する必要があるのかを、次の仮想的な臨床試験の例を通して考えてみる。

「脂質異常症に対する治療法として食事療法のみよりも、食事療法に抗脂質異常症薬を加えたほうが心筋梗塞の発症を減少させることができるかどうか」という仮説を調べるために、6人の脂質異常症患者を対象とした臨床試験を行うことを考えた。試験治療（抗脂質異常症薬＋食事療法）グループとコントロール治療（食事療法のみ）グループ（以下、対照治療）の2グループ間で、

試験開始から5年後の心筋梗塞発症割合を比較することをこの研究におけるおもな評価項目（エンドポイント）とする。この6人の対象者における、ある特定の個人に関する試験治療の効果を調べるためには、以下のことを「同時に」知る必要がある。

- A）その個人が試験治療を受けた場合、5年以内に心筋梗塞を発症するかどうか
- B）その個人が対照治療を受けた場合、5年以内に心筋梗塞を発症するかどうか

この2つの状況を知ることができれば、その個人にとって、試験治療の効果があるかどうかを証明することができる。すなわち、状況Aでは心筋梗塞を発症せず、かつ状況Bで発症すれば、その個人にとって試験治療はプラスの効果があると判断できる。逆に、状況Aで心筋梗塞を発症し、かつ状況Bでは発症しなければ、その個人にとって試験治療はマイナスの効果があると判断できる。また、どちらの状況でも心筋梗塞を発症しない、あるいは発症する場合には、その個人にとって試験治療は効果なしということになる。

このように考えると、上の仮想的な臨床試験の例における6人の対象者全体での治療効果（平均的な治療効果）を調べるためには、6人全員に試験治療を行った場合の5年間の心筋梗塞発症数と、「同じ」6人に対照治療を行った場合の5年間の心筋梗塞発症数を「同時に」比較しなければならないことになる。しかしながら、これら2つの状況は、一方が観察されれば他方を観察することは不可能である。

つまり、理想的な実験研究を行うためには、「その研究で関心のある結果（心筋梗塞の発症）にかかわるすべての要因のうち、操作可能な1つの要因（抗脂質異常症薬を与えるかどうか）のみがグループ間で異なる状況」をつくり上げる必要がある。この目的を達成するためには、結果に影響を及ぼす可能性のあるすべての要因（リ

Table 3 6人の対象者に対するコントロールのレベル（文献3より引用）

番号	コントロールグループ	設定の仕方	バイアス
①	ランダム化コントロール	試験治療と対照治療をランダムに3人ずつに割り付ける	測定できない特徴も平均的には同一にできるが、偶然のバラツキによるバイアスは起こりうる
②	同時コントロール	試験治療を実施したグループと同時期に対照治療を実施した別のグループ6人	測定できる特徴は同一であるかどうか確認できるが、測定できない特徴はどうしようもない
③	既存コントロール	過去に対照治療を実施したグループの別の既存データ6人	記録に残っていた特徴は同一であるかどうか確認できるが、記録されなかった特徴はどうしようもない
④	コントロールなし	「対照治療で30%が改善する」などの研究者の常識（暗黙のコントロール）	バイアスがないとはまったく保証できない

スク因子)を制御する必要があるが、残念ながら、本質的にバラツキを伴うヒトを対象とした研究においては、すべての要因が既知であることはまれであり、人為的に操作した要因のみが異なる理想的なコントロールグループをつくることは不可能である。

理想的なコントロールグループをつくることは実際にはできないが、試験治療の効果を比較するための現実的なコントロールグループの設定方法には、4つの方法がある (Table 3)³⁾。いずれの方法も、研究者自身による人為的な操作により試験治療を行うかどうかを決定することができるが、どの方法が科学的な実験研究（研究者自身が行う人為的な操作によってどのような結果が研究対象に生じるかを、適切なコントロールグループが存在する状況で調べるための一連の観察）に一番近いであろうか？

前述の「同じ6人に対照治療を行った場合」という理想的なコントロールグループは実際には存在しないが、このような研究が実施できたと仮定し、そのような研究を「比較可能性 (comparability) がある」と呼ぶことにすると、①から④へ番号が大きくなるに従って、この比較可能性を保証することが難しくなる。

①のランダム化臨床試験の場合でさえも、上の仮想例のように対象者数が6人と少なければ、偶然の要素により一方のグループに総コレステロール値の高い対象者が集まってしまうこともあり、比較の妥当性が損なわれる可能性がある。このような2グループ間の治療以外のリスク因子が比較可能でない現象は交絡バイアス (confounding) と呼ばれ、その研究での内部妥当性が損なわれる (バイアスが存在する) 1つの原因となる。

しかしながら、このランダム化コントロールは、ほかの②～④のコントロールとは違い、(たとえば、コインを投げたりサイコロを振ったりという) 純粋に偶然の要素に基づいて対象者を2つのグループに分けているために、対象者数が増えれば、性、年齢、総コレステロール値などの心筋梗塞発症にかかわる既知の要因のみなら

ず、生活習慣などの (われわれの現在の知識ではわからない) 未知の要因をも含めて平均的には心筋梗塞発症リスクが均一な2つの (理想的に近い) グループをつくり上げることができるという利点をもっている。いいかえれば、④のコントロールが存在しない状況、③の過去に記録された情報のみに基づいた既存コントロール (historical control)、あるいは、②の未知の要因を考慮できない同時コントロール (concurrent control) と違い、ランダム化コントロールを採用し、対象者を「平均的には心筋梗塞発症リスクが均一」という比較可能な2グループに分けることにより、理想的な実験研究の実施に近づくことができる。

ランダム化コントロールを採用したとすると、試験治療、対照治療グループそれぞれの対象者に割り付けられた内容を5年間守ってもらい、心筋梗塞の発症割合を比較することで、「対象者全員に試験治療を行った場合の5年間の心筋梗塞発症数と、同一の対象者に対照治療を行った場合の5年間の心筋梗塞発症数の比較」を平均的な比較可能性という観点からバイアスなく行うことができる。さらに、この治療法のランダム化という「強い」人為的な操作を行うことにより、治療効果の検定や推定といった統計計算の基礎も与えられる。

4. 臨床試験結果の一般化可能性

ランダム化比較臨床試験、特に標準治療の構築と普及を目指す第Ⅲ相臨床試験計画の目標は、精確性 (clarity)、比較可能性 (comparability)、一般化可能性 (generalizability) の3つにまとめることができる⁴⁾。精確性の要求 (2で述べた①) とは、主たる評価項目の計測について、そのランダムなバラツキを小さくし、研究の精度を向上させることである。比較可能性の要求 (2で述べた②) とは、交絡バイアスを減らし、研究の内部妥当性 (internal validity) を向上させることである。研究の第一義的な目標は、研究対象集団からこの2つの要求を満たした結果を得ることであるが、得られた結果の一般化可能性の検

討も重要である。

結論を適用する有限母集団が明確に定義され、その母集団からランダムに対象者がサンプリング(無作為抽出)されているのであれば、内部妥当性を満たす結果の外挿(外的妥当性の確保)は容易である。しかしながら、通常の臨床試験においては、研究成果の適用を考えている目標集団(target population)と実際に研究が行われる標本集団の間には数のうえでも質のうえでも大きな隔たりが生じる。患者を対象とし、インフォームドコンセントを前提とした臨床試験においては、患者集団からのランダムサンプリングを行うことは不可能であり、そもそも、試験参加に関する適格・除外基準の設定には幅がある。このような場面において、ランダム化によって実現されるのは、試験に参加した患者集団に対する内部妥当性の確保のみであり、結果がより広い対象集団に適用できるかの検討は、別の角度からなされなければならない。

一般化可能性の検討は、論文の考察部分(特に研究の限界)として、対象集団の特徴の議論・ほかの研究結果との比較を踏まえて述べられるのが通常である。対象者を性、年齢、疾患重症度などの背景要因によって分け、これによって得られるサブグループごとに解析を行い(フォレストプロット)、結果がサブグループごとに大きく異ならないのであれば、全体での結果の一般化可能性が高いと推察される(サブグループ解析は、有意なサブグループをみつけるための手段ではない)。また、多施設臨床試験で、施設内での治療効果が施設ごとに大きく異ならないのであれば、一般化可能性は高いであろう。研究の繰り返しによる結果の確認により一般化可能性は高まり、メタアナリシスを一般化可能性の検討手段と捉えることもできる。

しかしながら、ある研究からの結論の一般化を確保するためには、「得られた結果を抽象化し、特定の研究結果からその研究を実施した時期や地域を越えた普遍的な仮説を産み出す」というプロセスの積み重ねが重要であり⁵⁾、これは決して統計的な問題だけではない。臨床試

験における結論の一般化のためには、統計的な議論だけでなくデザイン段階において、患者選択を意図的に行うのではなく、「試験の適格基準を満たす患者にはすべて参加をお願いする」という原則を守ることも重要である。

まとめ

臨床研究における生物統計学の貢献は、本質的にバラツキを伴うデータ、およびその曖昧さ・不確実さの中で得られる結果に対する信憑性を確保・評価することである。これらの信憑性を支える柱が、研究計画・データ管理・統計解析の3つである。いくら素晴らしい統計解析手法を用いたとしても、研究計画とデータ管理がよい加減であれば、そこから得られる結論は弱いことが予想される。したがって、妥当で効率のよい研究デザインはどのようなものであり、プロトコルをどのように開発するか、調査票をどのように設計するかなどの研究計画の問題・モニタリング・データ入力・標準化とデータベース管理をどのように行うか、そしてそれらの品質保証をどうするかなどのデータ管理の問題(組織、システム、人材面)をまずクリアすることが大事である。

利益相反開示

本総説に関し、著者は開示すべき利益相反を有していない。

文 献

- 1) 辰巳邦彦：主要基礎・臨床医学論文掲載数の国際比較。政策研ニュース 35：43-45, 2012
- 2) 佐藤俊哉：調査計画書の作成—調査を始める前の自己チェック。保健の科学 37：72-76, 1995
- 3) 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 原理の方法。東京, 朝倉書店, 1999, pp21-33
- 4) 大橋靖雄：Biostatistician からの本邦臨床試験への提言。癌と治療 20：2483-2495, 1993
- 5) Rothman KJ, Greenland S, Lash TL：Modern epidemiology. (3rd ed). Philadelphia, PA：Lippincott Williams & Wilkins, 2008