

## 第4回 研究のリフォーム, 耐震補強

横浜薬科大学医療統計学

奥田千恵子 Chieko OKUDA

### 投稿論文がリジェクトされる理由

投稿論文がすんなりと通ることは稀である。リジェクトされないまでも、回答不可能な質問が並んでいたり手間のかかる追加実験やデータ収集を要求されて、途方にくれた経験をもつ研究者は少なくない。

British Medical Journal (以下, BMJ) 誌に「投稿論文がリジェクトされる理由 (Why were papers rejected for publication)」<sup>1)</sup>として挙げられた項目は以下のようなものである (番号は原文にはない)。

- (1) 科学的に重要な問題が述べられていない。
- (2) オリジナルな研究ではない (誰かがすでに同じあるいは類似の研究をしている)。
- (3) 実際には著者の仮説を検証していない。
- (4) ほかのタイプの研究をすべきであった。
- (5) 現実的な困難 (例えば被験者を募集する場合など) のため、著者らが妥協して元の研究計画通りに実施しなかった。
- (6) データ数が小さすぎる。
- (7) 対照がない, あるいは不十分である。
- (8) 統計解析が不正確, あるいは不適當である。
- (9) データに合わないような結論が導かれている。
- (10) 重大な利害関係の対立がある (著者の1人, あるいは資金提供者がその論文の公表により経済的に利益を得ているかもしれず, バイアスを防ぐための保護が不十分になっている)。
- (11) 論文の書き方があまりにひどくて理解不能である。

(1), (2), および(11)は研究者としての資質が問われる事項であり本稿の守備範囲を超えている。(10)利益相反 (conflict of interest: COI) に関しては, 現在, 各誌において, この論説文が書かれた当時 (1997年) 以上に厳しい規範が設けられている。

本シリーズ最終回では, 残る(3)~(9)の各項について, 論文が瑕疵物件とならないようにするためには事前にどんな注意が必要か, はからずも欠陥建築ができてしまったり地震で亀裂が生じた時にはどのように改修, リフォームして, 新たな物件として売りに出すことができ

るかを考える。

### 具体的な命題を提示する

#### (3)研究仮説の検証に関して

論理が破綻してしまっている場合は(11)に含めることにして, まず研究仮説の検証とは何かを考えてみよう。

統計学的な検定を行うには, 必ず具体的な命題を研究仮説として提示しなければならない。例えば, 「治療法Aの方が治療法Bより優っている」という研究仮説を立てて, 生存率など直接的な指標の差を統計学的に検定する研究であれば, (3)のような指摘を受けることは少ない。しかし, 治療法の優劣を決めるのは延命だけではない。様々な臨床的指標が検定の対象となり得る。研究目的のセクションで, 研究仮説と指標の関係およびその妥当性を論理的で説得力のある形で述べる必要がある。

明確な研究仮説を立てずに研究をスタートしてしまい, 論文にまとめる段階になって初めて統計学の教科書やソフトのマニュアルを開くという研究者が多いのが現状である。研究仮説の検証のためにはどのように対象を選び, どのように群分けし, どんなデータの集め方をすればよいのか……。リジェクト理由の多くは研究デザインについての知識の欠如が原因となっている。統計リテラシーなくしては研究をデザインすることはできない。

### 計画段階で十分な検討を加える

#### (4)研究のタイプに関して

推測統計を用いる分析的研究 (analytical study) では, できるだけ標準的な研究デザイン (第1回表1) を用いたほうが先行研究との比較がしやすい。可能なかぎりバイアスの入り込みにくいエビデンスレベルの高いデザインにするのが望ましいが, (5)で指摘されているように, 常に時間や費用, 倫理的配慮などの現実的な困難が付きまとう。計画段階で十分な検討を加えて何度も設計図を描き直してから研究を始めなければならない。

研究者はややもすれば群間に有意差があるとか, 変数間で相関がみられるとかにこだわりがちであるが, 統計

学的な有意性や相関の有無はアクセプトの必要条件ではない。貴重な症例や珍しい現象、独創的な方法が用いられている研究、あるいは、探索段階の研究であっても検証的な研究の設計に貢献できる重要な知見を含んでいれば、それをセールスポイントとして記述統計のみを用いた記述的研究 (descriptive study) としてまとめてみてはどうだろう。

研究内容によっては質的研究 (qualitative research) が適している場合もある。本シリーズでは、データ収集や分析において「数」を重視する量的研究 (quantitative research) のみを扱っている。例えば、質問票を用いた調査研究で得られるデータを「数」として扱うには回答選択肢を設ける必要がある。しかし、この方法では被験者が心にもつ多様な回答を十分とらえきれない。

質的研究では、「数」ではなくて「言葉」を重視する。統計は用いない。自由な語りによって研究ターゲットとする研究対象の経験を描きだし、その意味や意義を研究する。語られた経験の意味は、主観的であるがゆえに時間と共に変化し社会的文脈によって異なる可能性があるため、以前は非科学的と批判されることが多かったが、現在では医療分野でも質的研究の価値が認められるようになった。具体的な研究方法については成書を参考にされたい<sup>2)</sup>。

## 医療研究は公表の義務がある

### (6) データ数の不足に関して

事前に必要なデータ数を算出しておくというのが原則である (第2回参照)。研究の計画段階で、実現可能なデータ数では有意差が出ないことがわかった場合はデザインを見直す必要がある。

対象の性別や年齢、基礎疾患や薬剤使用の有無などの背景因子は、研究目的とする変数と多かれ少なかれ関係をもつ。観察的研究では、通常、背景因子の影響 (交絡) を除くために多変量解析を用いる。組み込まれる独立変数が多いほどデータ数を増やさなければならない。1つの独立変数あたり必要なデータ数は、多変量線形回帰では約20例、2項ロジスティック回帰やコックス比例ハザード回帰では、アウトカムの少ないほう (通常は、疾患発症者や死亡者) が10例程度とされている。データ数が足りない場合は何らかの方法で独立変数を減らさなければならない<sup>3)</sup>。

十分と思われるデータ数でスタートしても予期せぬ不足が起こり、研究仮説の検証に至らないことがあり得る。そのような場合には信頼区間 (以下、CI) を求めたうえ

で適切な考察を行えばよい (第2回参照)。

人間を対象とする研究を行う医療研究者は、研究の結果を一般に公表する倫理的な義務を有している。有意差が出なかった研究を引き出しに仕舞いこむべきではない。加えて、出版バイアスを引き起こさないために、医療分野の雑誌の多くは掲載を考慮する条件として、介入研究 (実験的研究) の場合、公的な試験登録システムへの登録を義務づけている。日本でも、厚生労働省が指定している大学病院医療情報ネットワーク (<http://www.umin.ac.jp>) などへの事前登録が必要である。

## 対照の設定の仕方

### (7) 対照に関して

分析的研究には必ず対照 (control) が必要である。対照とは、通常は、同一の研究者によって同時に研究に組み込まれ、観察される内部対照 (internal control) を指す。何らかの特徴をもっている、あるいは、何らかの因子に曝された群と、そのような特徴や曝露がない対照群との間で比較を行うというのが群間比較デザインの基本型である。3群以上の間での比較や、研究目的とする因子の量や程度により用量反応関係を求めるデザインにすることもできる。

例えば、喫煙者 vs 非喫煙者の2群に分けて何らかのデータを比較する観察的研究の場合、過去の喫煙者をどう扱うかにより分け方が変わり、導かれる結論も変わってしまう。食品や薬物の摂取量に関しても同様の問題が起こる。そのような場合には、摂取量や摂取年数などによって数個のカテゴリに分けて、最小 (最短) 摂取群を基準カテゴリ (対照) とし用量反応関係を示すとよい。

## 間違えやすいポイント

(8) 解析手法の選択基準に関しては第3回を参照された。また、(9) データに合った結論を導くには、解析結果であるp値やCI (第2回参照) を研究仮説や研究デザインと関連づけてわかりやすくまとめる必要がある。

ここでは研究者が間違えやすいポイントを幾つか挙げる。

### 1. 検定の繰り返し

カルテから得た検査値を手当たり次第に検定して有意差の出たものを報告する; 経時的な測定データで、時点を変えて検定を繰り返す; 群全体で有意差が出ない場合、性別や年齢で層化して検定し直す; 順序カテゴリで表されたデータを、例えば、「非常によい」と「よい」を合

併して1つのカテゴリにして検定し直し、有意差が出た区切り方を採用する；長期間にわたる研究の途中で何度も検定して、有意差が出た時点で研究を終える、など、恣意的な群分けによる検定を繰り返してはいけない。

探索段階ではこのような試行錯誤はやむを得ないが、検証的研究では、偶然有意差が出る確率が増えるという多重性の問題（第3回参照）を避けるため、研究を始める前に解析計画を示す必要がある。

## 2. 比率データの扱い

臨床検査データのなかには、ヘマトクリットやヘモグロビンA<sub>1c</sub>など、パーセントで表示されるものがあり、通常の測定単位（例：mgやmM）で表されたデータと同様、連続量として扱われる。そのため、生存/死亡や再発/非再発などの2値データを比率で表した生存率や再発率なども、誤って連続量として扱ってしまうことがある。このようなデータには離散量の解析法（ $\chi^2$ 検定や、リスク比（以下、RR）やオッズ比の形で区間推定）を適用しなければならない。

## 3. 時間データの扱い

時間データの扱いは研究内容によって様々である。分や時間単位の場合は、連続量データとしてt検定や分散分析などのパラメトリック検定法を用いることもあるが、正規分布には従わないことが多いのでノンパラメトリック検定のほうが適切である。

一定の期間中のアウトカム（死亡など）の発生頻度で表されたデータは、介入の時点やリスクファクターに曝された時点からアウトカムが発生するまでの期間（月、年）がデータとして得られる場合は、カプラン・マイヤー法や、コックス比例ハザード回帰による生存時間分析を行う。

## 4. 用量反応関係の扱い

多数の用量カテゴリがあるデザインでは、各用量群間の対比較ではなく、系全体としての傾向性を調べるほうが望ましい。例えば、食品や薬物の摂取量と様々なアウトカムとの関係を調査した観察的研究では、多変量解析により交絡因子の影響を除いたうえで、用量反応関係があるかどうか調べなければならない。傾向性の検定には高度な手法が必要となるため、探索段階の研究では記述統計にとどめてグラフなどにより傾向を視覚的に示すとよい。

## 5. 経時的变化の扱い

個々の時点での群間比較や、群内で時点間の比較を繰り返してどこかに有意差をみつけるのではなく、用量反応関係と同様、系全体としての傾向をとらえるべきである。最大の反応が起こるまでの時間や、曲線下面積（ベ-

表 研究デザインに基づいたエビデンスの序列

- ・患者データに基づかない、専門委員会や専門家個人の意見
- ・記述的研究（症例報告、症例集積など）
- ・分析疫学的研究（ケース・コントロール研究、コホート研究など）
- ・非ランダム化比較試験
- ・1つ以上のランダム化比較試験
- ・系統的レビュー/メタアナリシス

（表の下のほうほど信頼性が高い）

スライン値からの時間的変動パターン（面積）を求めれば、高度な統計手法を使わなくてもt検定などによる群間比較が可能である。

## エビデンスの序列

統計学的にみた研究デザインの序列は、第1回表1「医療研究における研究デザインの分類」としてすでに示したが、医療研究は統計学に裏付けられた科学的妥当性だけでなく、結果の臨床的重要性という見地からも評価される。従って、動物から得られた情報より人から得られた情報のほうがエビデンスのレベルが高い。

複数の研究結果が一致していればさらに信頼性が増す。過去に行われた複数の独立した研究結果を統合し分析するメタアナリシス（meta-analysis）は、科学的根拠に基づく医療（evidence-based medicine：以下、EBM）のための最も強力な情報を提供できると考えられている（表）。

## メタアナリシスとは

過去に行われた特定のテーマに関する研究を、計画的、包括的、かつ、再現可能な方法で収集し、研究結果を要約したものを系統的レビュー（systematic review）という。著者の自由裁量に委ねられている従来の総説（narrative review）とは異なり、文献の選択バイアスを避けるために、MEDLINEやEMBASEなどのデータベースの系統的な検索が必要である。

メタアナリシスは系統的レビューで特定した研究の数値的な結果を統計学的に併合したものである。収集した研究資料から、適格基準（研究デザイン、出版・研究の時期、言語、被験者の選択基準、標本数、観察期間など）を決めてデータを抽出する。研究間の許容できない異質性や出版バイアスの有無を検討したうえでデータを統合する。通常、フォレストプロット（forest plot）と呼ばれる表示方法を用いる。

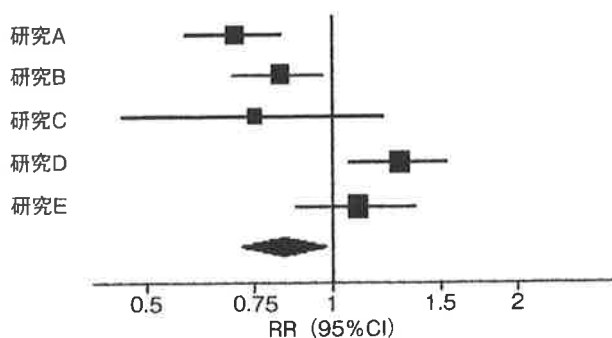


図 フォレストプロットの例

例えば、試験薬Xの治療効果（RRが小さいほど効果がある）を調べた5つのランダム化比較試験のメタアナリシス（図）では、統合RRの95% CI（ひし形の横幅）が1を含まないことから、試験薬Xの治療効果があると判定される。

## おわりに

医療従事者が自ら研究を行うために、また、医療情報を適切に判断し患者との良好な協力関係を築くためにも統計リテラシーが必要である。

最終回では研究の最終段階である論文投稿を取り上げた。本稿が対象とする統計初学者にはややハードルが高く感じられたかもしれないが、できるだけ著者の独断を避けるために、敢えて、The Lancet, Journal of American Medical Association (JAMA), The New England Journal of Medicine (N Engl J Med) などと並んで臨床分野の雑誌を代表するBMJの論説文を引用した。加えて、

EBMのための最も信頼性の高い情報を提供できると考えられているメタアナリシスを紹介した。

上述の雑誌に掲載される論文やメタアナリシスは、多くの研究者が共同で何年もかけて行った大規模な研究が中心である。これから研究を目指す医療従事者は、ビルや橋の建造を目指す前に、まず、日曜大工道具を使って物置小屋作りや、壊れたままで放置してある庇の修理から始めればよい。建造物の大小にかかわらず、地震や風雨に晒されても倒壊しないものをつくるための知識や技術の基本は同じである。

大学などの研究機関に属する研究者と比較して、病院薬剤師が日常業務の傍ら研究に割ける時間は限られている。しかし、医療現場という環境は研究室に籠ってはい手に入らない研究テーマの宝庫である。何か面白そうなテーマがみつかったら恐れず第一歩を踏み出していきたい。本シリーズがそのための小さな手助けとなれば幸いである。

## 引用文献

- 1) T. Greenhalgh : How to read a paper, *BMJ*, 315, 243-246 (1997).
- 2) P. Liamputtong *et al.* : “現代の医学的研究方法、質的・量的方法、ミクストメソッド”, 木原雅子, 木原正博訳, メディカル・サイエンス・インターナショナル, 東京, 2012.
- 3) H.K. Mitchell : “医学研究のための多変量解析”, 木原雅子, 木原正博監訳, メディカル・サイエンス・インターナショナル, 東京, 2008.

## 参考文献

1. 奥田千恵子 : “医薬研究者のための研究デザインに合わせた統計手法の選び方”, 金芳堂, 京都, 2009.