

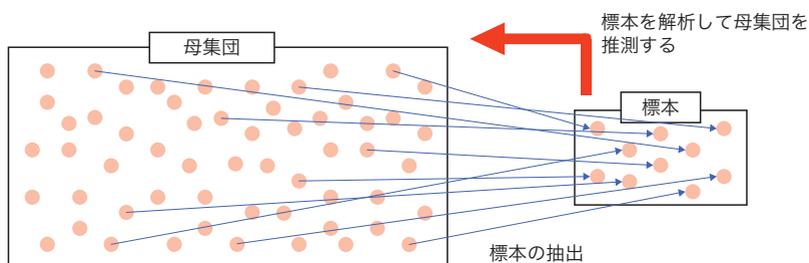
1 なぜ統計解析が必要なのか？

人間は自分自身の経験に基づいて、感覚的にものごとを判断しがちである。例えばある疾患に対する標準治療薬の有効率が50%であったとする。そこに新薬が登場し、ある医師がその新薬を5人の患者に使ったところ、4人が有効と判定されたとしたら、多くの医師はこれまでの標準治療薬よりも新薬のほうが有効性が高そうだと感じることだろう。しかし、たまたま有効性が出やすい5人に治療が行われたにすぎないかもしれない。同じ疾患を有する患者であったとしても、疾患の細かな分類や進行度、患者の年齢、性別、臓器の状態などによって有効率は左右される。さらに背景の条件が全く同じであったとしても有効率にばらつきは生じる。治療に対して思い入れが強ければ強いほど、治療結果に大きく一喜一憂し、客観的な評価が困難となる。印象に残る結果は感覚的な判断を偏らせてしまう。

統計解析の目的の1つは、前提としてこのような様々なばらつきが存在する状況の中で、限られた**標本 (sample)** から**母集団 (population)** を**推測**し、より一般的な結論を導き出そうとすることである。**母集団**の定義は状況によって異なるが、例えばある疾患に対する新薬の有効性を評価するのであれば、その疾患を有する全ての患者が母集団となる。統計解析をしていると、目の前にあるデータだけを対象としているような錯覚にとらわれることがあるが、実際に行っていることは、その標本を用いて本当の**母集団の全体像を推定**しようとしているのである（選挙の出口調査による全体の投票数や議席数の予測をイメージすればよい）。



統計解析の目的の1つは、母集団から抽出した**標本 (サンプル)** を用いて解析することによって、**母集団を推測**することである



2 変数の種類とその要約

① 変数の種類

統計解析で扱う主な変数は**連続変数 (continuous variable)**、**順序変数 (ordinal variable)**、**名義変数 (nominal variable)** の3つに分けられる。連続変数は身長、体重など、数値で表される**定量的なデータ**を意味する。順序変数、名義変数はいずれも**質的なデータ (カテゴリー変数、categorical variable)**であるが、順序変数は尿蛋白の(-)、(±)、(+)、(2+)、(3+) や、腫瘍の進行度のステージ I、II、III、IVのように順序づけられたものである。一方、名義変数は、性別の男性、女性や、ABO血液型のA、B、O、AB型のように順序の関係がない(男性、女性、あるいは有効、無効のように二値だけを持つ場合は**二値変数**あるいは**二区分変数 (binary variable)**とも呼ばれる)。

特殊な変数として医学統計ではしばしば**生存期間**の解析が行われる。正確にいうと、必ずしも生存期間だけを対象とする解析ではなく、ある時点からあるできごと(イベント)が発生するまでの期間(**time-to-event variable**)の解析であり、死亡がイベントとして定義された場合に生存期間の解析が行われることになる。この解析方法の特徴は、ある時期まで生存していた(あるいはイベントが発生していなかった)ことは知られているが、その後の情報が得られないような場合に**観察打ち切り (censor)**として解析に含めることができる点である。例えば、ある疾患に対して特定の治療を行った後の生存期間を解析する場合に、最終観察時点で生存中の患者の真の生存期間は不明であるが、その時点で打ち切りとして扱うことによって解析に含めることができる。この解析においては、イベントが1回しか発生しないものであることと、打ち切りとなる理由が解析対象のイベントの発生とは無関係であることが必要である。例えば悪性腫瘍に対する化学療法後の生存期間の解析において、打ち切りとなった理由が他院への転院のような場合は、病状が増悪して死期の近づいた患者がしばしばホスピスに転院するという背景が解析上の**偏り (バイアス)**を生じてしまう危険性がある。

② 変数の要約、信頼区間

各変数を**要約**して記述する方法はそれぞれの解析のところで詳しく述べるが、まずは全体像を眺めることが重要である。名義変数なら**頻度分布**を、連続変数であれば**散布図**、**ヒストグラム**、**箱ひげ図**などを描いてみる。生存期間を表すためには**Kaplan-Meier 曲線**が用いられる。各変数を端的に記述するには、そ

似的な母集団として、そこから無作為に重複を許しながらサンプル抽出を繰り返したデータセットの解析を反復して検証とするような方法（**Bootstrap 法**）も用いられている

医学領域でしばしば行われている変数の選択方法は、まずは研究の興味の対象となる変数、理論的に重要であると考えられる変数、過去の研究結果によって重要であると考えられる変数、単変量解析で有意となった変数（抑制因子の影響を考慮して $P < 0.10 \sim 0.20$ のように広めに選択する）を全てモデルに含めるという方法である。これらを全て含んだモデルを最終モデルとすることもあるが、サンプル数が不十分であるために変数を限定する場合には、機械的に $P \geq 0.05$ の変数の中から P 値が最も高い変数を1つずつ除外して再計算することを繰り返し、最終的に有意な ($P < 0.05$) 変数だけを残す方法 (**backward stepwise deletion**) や、モデルの適合度のよさの指標 (F 検定、尤度比検定、AIC、BIC など) をみながら変数を増減することによって、最も当てはまりがよいモデルを決定する方法などが行われる。ただし、ある特定の変数に関心があるような場合は、その変数は常にモデルに残す必要がある。理論的に、あるいは過去の研究結果などによって明らかに重要である因子も同様に除外しないようにする。

8 EBM と臨床研究

1990年代から **EBM (evidence-based medicine)** という用語がもてはやされるようになった。EBMは**臨床疫学 (clinical epidemiology)** を臨床現場の診療により近づけたものであり、個々の患者の治療を決定する際に、その時点での最良の**エビデンス** (臨床研究の実証結果) を解釈したうえで、患者の状態や価値観をふまえて慎重に適用することである。EBMの実践はステップ1「問題点を定式化する」、ステップ2「その質問に答えるためのエビデンスを収集する」、ステップ3「得られたエビデンスの信頼性を批判的に吟味する」、ステップ4「得られた情報が目の前の患者にあてはめられるかどうかを検討する」、ステップ5「自分が行った1から4のステップを自己評価する」に沿って行われる。

ステップ1では臨床現場で直面する疑問を明確な言葉 (**クリニカルクエスション**) に置き換える作業であり、「どのような患者に、どのような治療を行ったら、対照群と比較して、どのような結果の違いが生じるか」という文章で表す。構成要素のそれぞれの頭文字をとって **PICO** あるいは **PECO** と呼ばれる (表)。一般に観察研究では何らかの要因への曝露 (**exposure**) の影響を



クリニカルクエスションの構成要素

P	Patients	対象とする患者
I (E)	Intervention (Exposure)	介入や曝露（治療、検査など）
C	Control	比較する対照群
O	Outcome	結果、転帰（治療効果、生存率など）



EBMの5つのステップ

ステップ1	患者の問題の定式化
ステップ2	問題についてのエビデンスの収集
ステップ3	エビデンスの批判的吟味
ステップ4	エビデンスの患者への適応
ステップ5	1~4のステップの評価

評価するので PECO となり、介入研究では何らかの介入（intervention）の影響を評価するので PICO となる。

ステップ2では PICO で得られたキーワードを利用して文献検索を行う。網羅的な検索を行うにはインターネットで PubMed などの**一次資料**（主として専門誌に掲載された学術論文）を利用することができるが、数多くの文献が抽出された場合にその中から重要なものを見極めていくにはある程度の知識が必要である。そこで、不慣れな分野で検索を行う場合は、まずはエビデンスを集積した**二次資料**から読み始めることを推奨する。例えば専門雑誌の**総説**（review article）や UpToDate、Clinical Evidence、Cochrane library なども有用である。

ステップ3では得られたエビデンスの信頼性、妥当性について客観的な評価を行う。このステップでは基本的に各研究の中で行われた解析、結論が**バイアス**や偶然によって歪められていないかどうかを科学的に検証する（すなわち**内的妥当性**の評価である）。また、結果の臨床的な重要性についても検討しなければならない。有意差が得られたということと臨床的な重要性（効果の大きさ、**エフェクトサイズ**）とは別問題である。有用な二次資料はこのステップ2、3を集約してくれる場合もある。

そして、ステップ1~3で得られたエビデンスが目の前の患者に適用できるかどうか（**外的妥当性**の評価）を検討するのがステップ4である。臨床試験の多くは年齢、疾患の進行度、合併症など、様々な要素で適格条件を絞りこんで

いるため、その試験で得られた結果が一般的な患者に応用できるとは限らないのである。患者や家族の人生観や社会環境なども治療決断に組み込まなければならない。

最後にステップ1~4で行ったことをステップ5で評価し、次の機会に役立てる。

これらのステップ（特にステップ4）をみればわかるとおり、EBMとは画一的な料理帳のような治療を強要するものではない（David L SackettはEvidence-based medicine is not “cook-book” medicine.と記述している）。数多くの診療ガイドラインが出版されているが、これらについてもガイドラインに従っていればよい、あるいは従わなければならないというような金科玉条ではない。ガイドラインはあくまで「一般的な」診療を示すものにすぎず、それが該当する患者はおそらく60~80%程度であろう。また、決してメタアナリシスや大規模無作為割付比較試験によって生み出された、いわゆるLevel Iエビデンスだけが求められるわけでもない。現時点で存在するエビデンスをいかにして個々の患者の診療に役立てるかに医師の力量が問われる。

一方、日常診療においてエビデンスを検索すると、十分なエビデンスが見つからないことが多い。そのため、よりよいエビデンスを生み出すための臨床研究が必要になる。すなわちEBMも臨床研究も、いずれも日常診療において生じる疑問から始まる表裏一体の関係にある。



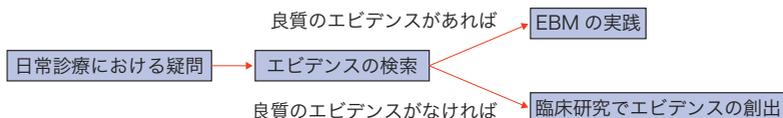
エビデンスレベルの分類の例

（厚生労働科学研究費のガイドラインの作成の手順より）

- | | |
|------|----------------------------|
| I. | システマティックレビュー/メタアナリシス |
| II. | 1つ以上のランダム化比較試験による |
| III. | 非ランダム化比較試験による |
| IV. | 分析疫学的研究（コホート研究や症例対照研究による） |
| V. | 記述研究（症例報告やケース・シリーズ）による |
| VI. | 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 |



EBMと臨床研究の関係



定の p_0 , p_1 , α , β の条件下に、最も少ない症例数が期待されるモデルが計算によって求められる。

無作為割付比較試験（第Ⅲ相試験） の症例数を設定するためには、第Ⅱ相までの試験結果に基づいて各群の主要評価項目の結果を設定する。その際に群間の差がどの程度以上であれば臨床的に意味があるかを考えなくてはならない。次に α , β の値を設定する。この値は、第Ⅰ種、第Ⅱ種それぞれの過誤による損失の大きさを考慮して決定されるが、 $\alpha=0.05$, $\beta=0.10\sim 0.20$ と設定されることが多い。生存期間の比較の場合、死亡数（**イベント発生数**）が重要であり、追跡期間内での死亡率が高いほど必要症例数は少なくなる。実際の割付においては年齢、性別、既知の重要な予後因子などが各群で均等になるように層別化した割付を行う。

EZR ではこれらのサンプルサイズは

【統計解析】

→ 【必要サンプルサイズの計算】

で算出できる。

15 臨床試験のプロトコール

プロトコールとは臨床試験の実行のために詳細に記載された計画書である。臨床試験から有意義な情報を得るためには、何を知りたいかを明確にした上で、実行前に入念に検討されたプロトコールを作成することが必須である。プロトコールは表「プロトコールを構成する要素」に示すような要素から構成される。患者の適格性については疾患の種類、診断の方法、病期、評価病変の有無、臓器機能などを定義しなければならない。

1998年10月から施行されるようになった**新GCP (good clinical practice)**では、モニタリング、監査などの基準が厳密になった。**モニタリング (品質管理、QC, quality control)**とは指名されたモニターが臨床試験の遂行状態を調査し、試験が実施計画書、標準業務手順書、薬事法の規定基準、新GCPに準拠して実施・記録・報告されているかを確認する活動である。モニタリングは臨床試験実施中の全期間を通して継続的に行われ、問題点の早期発見や改善に役立てられる。**監査 (品質保証、QA, quality assurance)**とは臨床試験の実施ならびに実施計画書の遵守状態を評価することにあり、モニタリングなどの品質管理とは別に分離独立した形で品質を保証する活動である。臨床試験の開始前や終了後に行われ、臨床試験の信頼性が確保されているかどうかを評価

する。モニタリングや監査では必要に応じてカルテや検査結果などの原資料の**直接閲覧（SDV, source data verification）**が行われる。新GCPに準拠する臨床試験においては、モニタリング、監査の方法についてもプロトコール中に明記しなくてはならない。日本国内の臨床試験のサポート体制においては新GCPに準拠した臨床試験の実施は容易ではないが、**治験**、すなわち医薬品もしくは医療機器の製造販売に関して薬事法上の承認を得るために行われる臨床試験は新GCPに準拠して行われなければならない。2016年に公布された「人を対象とする医学系研究に関する**倫理指針**」では治験以外にも、侵襲を伴う介入試験については信頼性確保のためのモニタリングが必須となり、必要に応じて監査も求められるようになった。ただし、必ずしも全症例の厳密なモニタリング、監査を求めるものではなく、試験の内容に応じてモニタリングや監査のレベルを設定してプロトコールに記載し、倫理委員会の審査を受けることになる。なお、この倫理指針にもプロトコールに記載すべき項目が列挙されている。また、様々な大学や研究機関の倫理委員会が臨床研究プロトコールのひな形を公開しているので参考にすることができる。



プロトコールを構成する要素

(0. 概要)

1. 目的と意義
2. 背景
3. 患者の適格条件、除外条件
4. 登録、割付の方法
5. 治療計画とその変更基準
6. 薬物の情報
7. 予測される有害事象とその対応
8. 必要な検査と臨床データ
9. 評価項目、検査予定
10. 有害事象の報告、健康被害への補償の有無
11. 効果の判定基準
12. 統計学的事項（症例数、研究期間、解析方法を含む）
13. 倫理的事項（利益相反、個人情報の取り扱い、同意取得の方法を含む）
14. データ収集
15. モニタリングと監査の方法
16. 研究組織（責任者、共同研究者、連絡先など）
17. 研究結果の発表方法
18. 参考文献
19. 付表
20. データ管理用紙
21. 説明同意文書