

研究計画時に覚えておきたい 統計の考え方

順天堂大学医学部附属順天堂医院
臨床研究・治験センター 臨床研究支援室
内山沙樹



はじめに：今回の話題

■臨床研究には様々なデザインがある

- 記述研究・観察研究などの非介入研究もとても大事な研究ですが、今回は介入研究に焦点を当ててお話しします

■どのような研究でも研究仮説の設定が重要

- Clinical Question(CQ)からResearch Question(RQ)への落とし込み
- この段階で統計が関わるイメージがしにくい

■RQの設定と統計がどうかかわるか？をご紹介します



研究仮説の定式化：PICOとFINER

- PICOは研究計画の初期段階から考え始める
 - FINERに照らし合わせながら計画を立てていく

PICO	
P	Patient：患者
I	Intervention：介入
C	Control：比較対照
O	Outcome：アウトカム

FINER	
F	Feasibility：実現可能性
I	Interesting：興味深さ
N	Novel：新規性
E	Ethical：倫理性
R	Relevant：必要性

PICO/FINERを計画に反映するしようと思った時に
統計の視点を持っていると役に立つ

PICOから明確になったRQ

■PICOを用いて設定されたRQから明確になること

- 「何を」 比べるか
 - 治療群と対照群における介入前後のLDL-Cの値
 - 治療群と対照群における投与開始からイベントが起こるまでの期間
 - …の具体的な中身

■では、「どのように」 比べるか？ = 統計解析

- 解析手法には様々な種類がありますが、その前に…

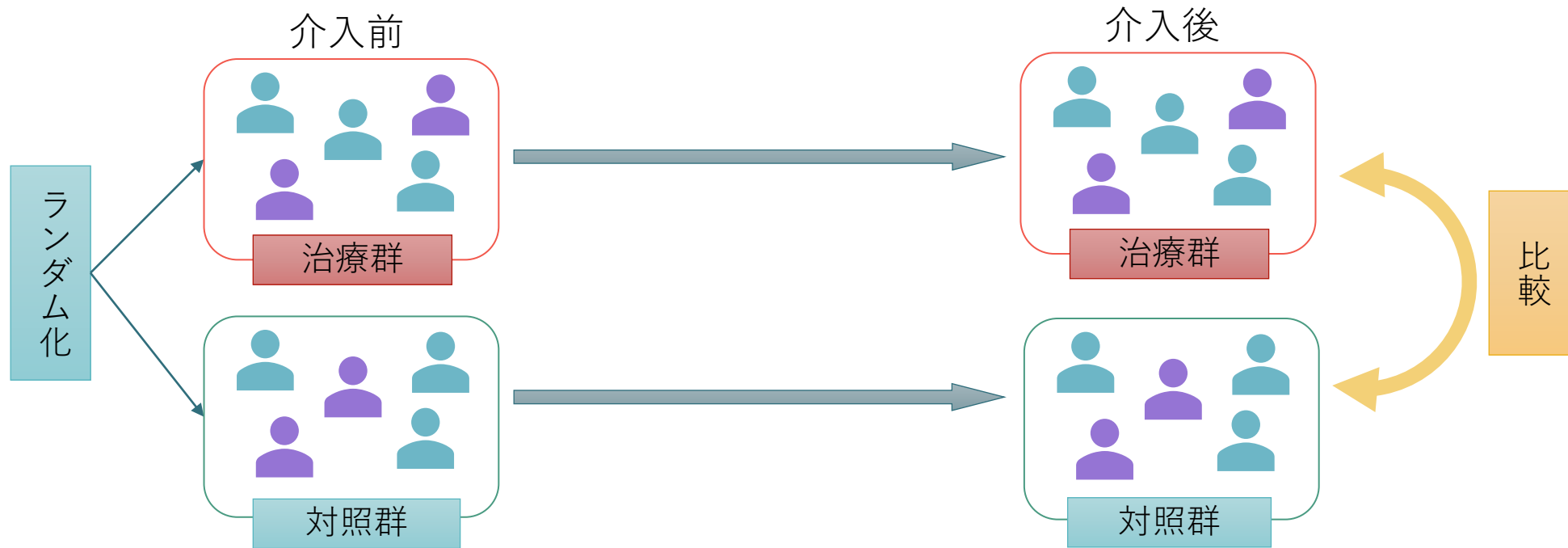
■データはどのように集められるのか？ = 試験デザイン

- 同じ解析手法でも情報の集められ方が異なると結果の解釈は変わる



簡単な例：群間の介入後のLDL-Cの値を解析

- 以下のデザインの試験の解析でt検定を行うとすると…
 - 理想的なランダム化試験

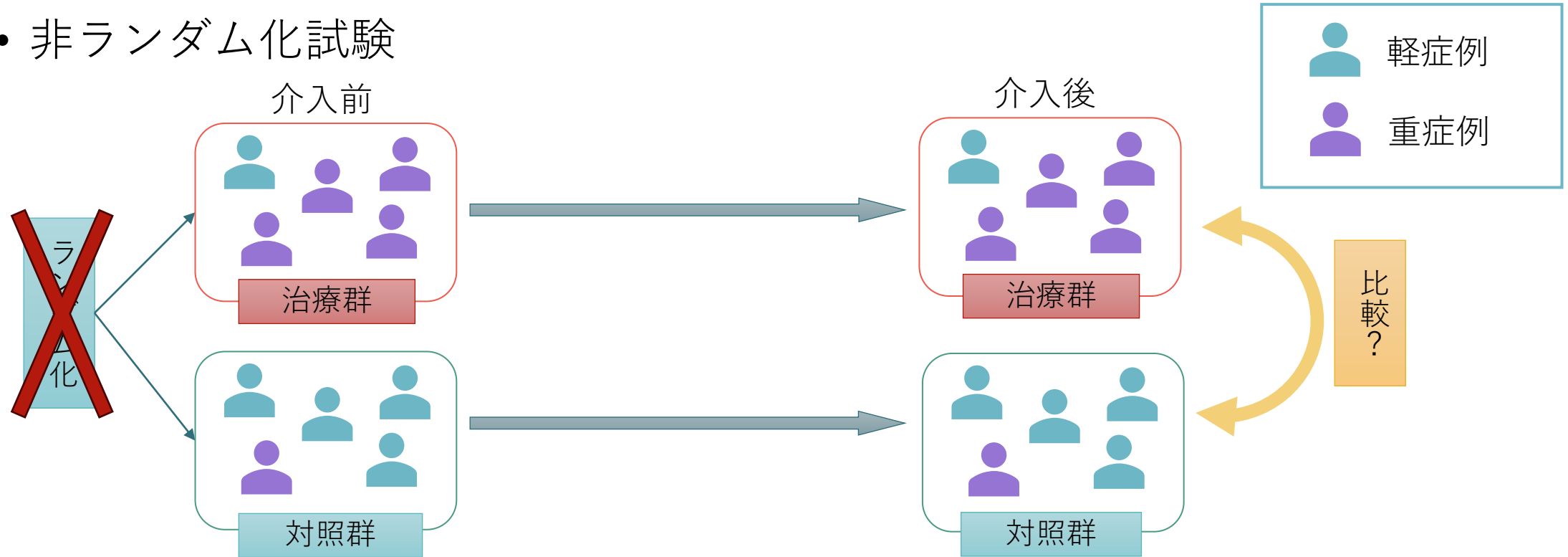


- 対象集団について、「介入後の治療群と対照群の平均の差の推定値は0mg/dLで、有意な差がある/あるとは言えない」などと言える
- (結論としてこのようなことを言いたいのではない場合はPICOや解析方法を検討)

簡単な例：群間の介入後のLDL-Cの値を解析

■以下のデザインの試験の解析でt検定を行うとすると…

- 非ランダム化試験

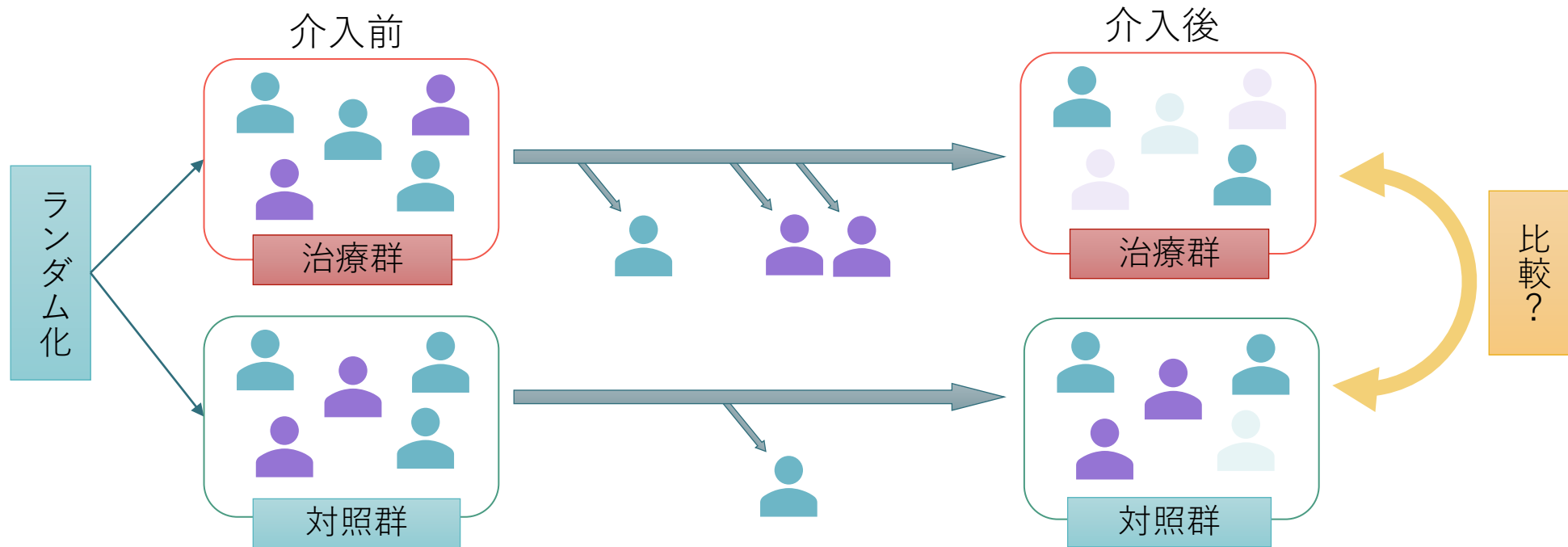


- 例えばより重症な人が治療群となりやすい場合、「重症な治療群と軽症な対照群」でのt検定で有意差があったとして、そこから何が言えるか？
- 適切にデータが収集されていれば解析による調整が可能

簡単な例：群間の介入後のLDL-Cの値を解析

■以下のデザインの試験の解析でt検定を行うとすると…

- ・ランダム化試験だが、疾患が悪化した人が途中で試験から離脱



- 悪化した人の介入後のデータは収集されない
- 「悪化した症例を除いた各群の平均の差」のt検定で有意差があったとして、これは研究の目的に合致した結果か？



試験デザインの重要性

- 統計解析の結果解釈には試験デザインの情報が不可欠
 - 「ある試験デザインである統計解析を実施したときに、研究仮説 (PICOで設定したRQ) に対してどのような答え方ができるか」という視点からの研究計画
 - 試験デザインの設計時点から統計家が関与するのが望ましい理由
- 試験デザインを適切に設計するために
 - PICOの設定をするとともに考えをまとめておくの良い点をお伝えします



結論として何が言いたいのか

■ 介入の効果をどのような形で結論したいのか？：研究目的のさらなる明確化

- 対照より優れている？
- 対照に劣っていない？
- 用量反応関係(どの用法でどの程度の効果があるか)？
- 安全性の確認？ …etc.

→何が言いたいかによって各々のことを言うために適切な試験デザインも解析方法も異なる



臨床的に意味のある差

- 例えば、「ある治療が対照より優れている」ことを明らかにしたいとき
 - 何をもって「優れている」とするか？
 - 例) 脂質異常症の領域
 - 「LDL-Cについて、対照群より治療群が〇mg/dL低くなれば優れていると言えるだろう」
 - これが1mg/dLであったときにそれは臨床的に重要な差なのか？
 - 具体的にどの程度の差(効果)を持って優れている = 臨床的な意味がある
と考えるかが重要
 - ここは統計的に決められるものではなく、研究者がどう考えるか、あるいはその領域でどう考えられているかによる
 - そして、特に検証的な目的の試験ではその治療がそれくらい効果があることを示唆する先行研究があること(期待する効果が現実的である根拠)も大切



なぜ計画時にこのような観点が必要なのか？

- サンプルサイズ計算は「臨床的に意味のある差」を検出するために必要な最小限の人数を見積もるもの
 - 検出すべき差を設定することで初めてサンプルサイズ計算ができる
- もし組み入れ可能であるならば、症例数は多ければ多いほど良いか？
 - 介入＝効果・安全性が明らかでない試験治療(あるいは治療効果がないプラセボ)を不必要に過剰な人数に実施することは、倫理的な観点から問題がある
 - (一方で、症例数が不足すると検出すべき差が検出できなくなる)

適切な仮定に基づくサンプルサイズ計算から得られた必要症例数通りに組み込むことは、倫理的な面、科学的な妥当性の面からも重要(組み入れ可能かどうかの実現可能性の確認のためだけではない)



サンプルサイズの計算に必要な情報

■二群比較試験の場合の典型的な例

- 連続変数：群間のアウトカムの差の平均値と標準偏差
 - あるいは、各群のアウトカムの平均値と標準偏差
- 二値変数：各群のアウトカム(成功可否など)の割合
 - あるいは、どちらか一方の群の割合 + 想定される割合の差
- (生存時間解析が主目的である場合は各群のハザード比やある時点での生存率など)

■全く同じデザイン・対象集団のデータ・要約値があることは稀

- 先行研究と自身の研究の差異を加味した上で、報告された値を自身の研究に適用できるか検討が必要



現実には生じる問題

■計算した結果、症例数が多すぎる

- サンプルサイズ計算は一般に以下の挙動をする
 - 連続変数：設定した効果(群間差)が小さいほど、標準偏差が大きいほど増大
 - 二値変数：設定した群間の割合の差が小さいほど増大
- (ばらつきと比較して)小さな差について二群比較試験で「統計学的に差がある」と言おうとすると、サンプルサイズは大きくなる
 - 組み入れ人数に制約がある場合、ある程度大きな差が見込まれるRQを見つけられるかも重要
 - 無理やりサンプルサイズを小さくする方向へ見積もりを大きく(小さく)すると(真の効果からの乖離が大きくなると)、本当は効果がある治療の場合でも有意差があると言える確率が小さくなっていく(検出力不足)



現実に生じる問題

- 参考にできそうな情報(先行研究)がない
 - もしかしたら、先に効果を見積もるための研究が必要かもしれない
 - そのRQはまだ「対照より優れているかどうか」といったことを評価できる段階にないかもしれない
- 研究の目的を再確認
 - 結論として何が言いたいか？ (p.9)
 - その結論を言うために必要な情報はそろっているか？



検証的試験と探索的試験

■今までの例でモデルにしてきたのは検証的試験

- 1. 「ある疾患はある集団においてxx%の割合で発症するようだ」
- 2. 「ある介入は対照に比べてその疾患の発症をxx%予防するらしい」
- 「1と2の情報が正しいなら、適切な計画をしてその通りに試験を実施し、得られた結果を比較すれば有意な差が出るはずだ」
- …というのが検証的試験の枠組み
- 結果が仮定の通りであればその治療についてより強く結論できる

■検証的試験は集団や疾患および介入(治療法)に対する知見がある程度そろっているから計画できる

- 実際は、「2」が具体的な数値としてわかっていないことも多々ある
 - (「1」に関しては記述的・観察的な研究が得意とする範囲)



検証的試験と探索的試験

■探索的試験

- 目的の一つとして、ある対象集団について当該治療の効果がどの程度かを探索・推定する(前項の「2」の情報を得る)ことが挙げられる
 - もちろん、探索的試験の目的はこれだけに限らず多岐にわたる
- 位置づけは早期に決まっていることが望ましいが、計画を立てているうちにゴールが不明瞭になった場合にはどの枠組みでの研究を計画するのか立ち返ると整理しやすい
- 可能性を探索する研究のため、結論としては「有用な可能性がある」という形になりやすいことには留意
 - 探索的試験の結果が検証的試験の基礎となるため、長期的な目線では検証的試験の設計(対象集団、アウトカムの効果量)を見据えながら計画できるとよい



探索的試験におけるサンプルサイズ計算

■ 検証的試験のサンプルサイズ計算(再掲)

- 「臨床的に意味のある差」を検出するために必要な最小限の人数を見積もるもの
 - その計算には先行研究などからのある程度明確な根拠が必要

■ 探索的試験では

- 「実現可能な例数を考えて〇例とする」ことも多い
 - 症例数が決まっているのにサンプルサイズ計算？
- この場合に行うのは、「この人数で試験を行った時、仮に治療効果がxxであったとしたらそれを検出できる確率は〇%」という計算
 - やはり「臨床的に意味のある効果」を設定するのは重要、ただし検証的な目的の場合に比べると主観の許容幅は大きい
 - 仮定した差を検出できる確率が極端に小さくないか、予め検討することができる

まとめ

- 統計的な視点は解析だけでなく試験デザインの設計時に重要
 - 解析は独立しているものではなく、研究仮説や試験デザインと連動している
- 結論まで見据えた研究仮説の設定
 - RQとともに結論の形や臨床的に意味のある差が定まると研究全体の見通しが良くなる
 - 実現可能性、倫理性、科学的な妥当性は特に見えやすくなる
 - サンプルサイズの計算も研究仮説設定の一環

難しいと敬遠されがちな統計ですが、
統計の視点があると計画全体が俯瞰しやすくなります

