

臨床研究と生物統計

令和5年度 厚生労働省 臨床研究総合促進事業
医師研修
神戸大学医学部附属病院

2023/12/17

臨床研究のデザイン

令和5年度 厚生労働省 臨床研究総合促進事業
医師研修
神戸大学医学部附属病院

2023/12/17

臨床研究の目的

▶研究のゴール

薬事承認 / ガイドライン収載 / 論文発表 / 学位取得 / 学会発表

▶研究の内容における目的

- ・ 因果関係を調べたい（新しい治療方法は有効なのか）
- ・ 予後に関連ある因子を調べたい。
- ・ 診断方法を確立したい（感度・特異度など）

サウナは体にいいのだろうか？

サウナをして効果のある人はどういう人だろう。

臨床研究の目的

▶研究のゴール

薬事承認 / ガイドライン収載 / 論文発表 / 学位取得 / 学会発表

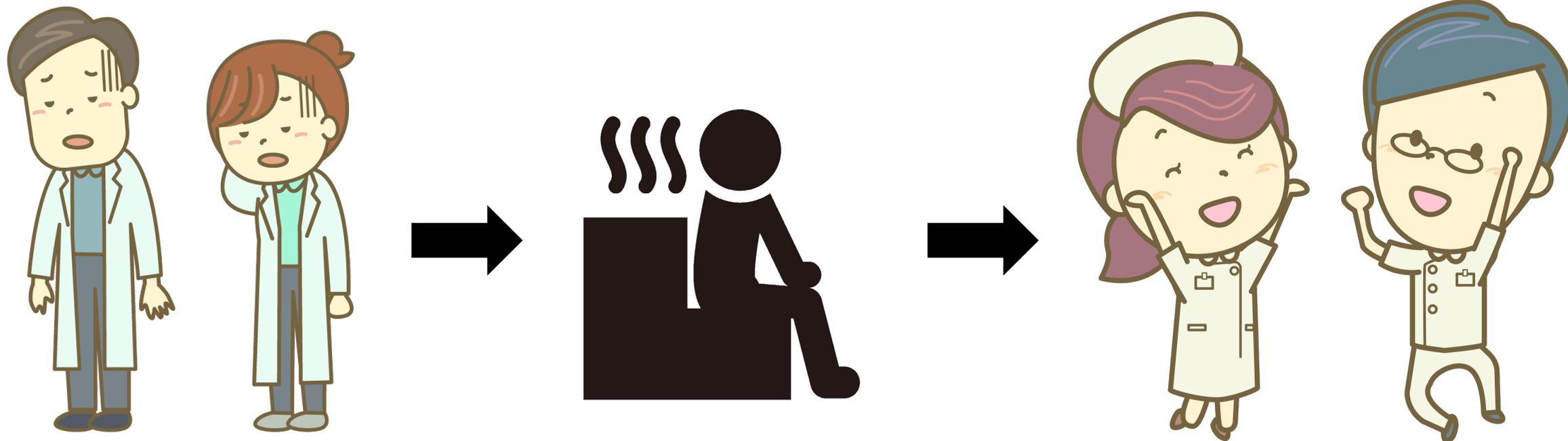
▶研究の内容における目的

- ・ 因果関係を調べたい（新しい治療方法は有効なのか）
- ・ 予後に関連ある因子を調べたい
- ・ 診断方法を確立したい（感度・特異度など）

“有効”であるとは

サウナが有効であることを示しているのでしょうか？

サウナを利用した人はストレスのない生活ができます。
サウナに通ってストレスフリーな生活を！



“有効”であるとは

サウナを利用した人はストレスのない生活ができます。
サウナに通ってストレスフリーな生活を！

サウナを利用しない人も
ストレスのない生活を送っている。



サウナの利用しない場合と比較して、
サウナを利用した場合の方がよいことを示す。

比較対照の種類

臨床研究以前の期間



①臨床研究として
サウナを利用した対象者の
ストレス

臨床研究の期間

比較対照の種類_同じ対象者



① 研究対象者でサウナを利用しない時期のストレス

臨床研究以前の期間



② 臨床研究としてサウナを利用した対象者のストレス

臨床研究の期間

比較対照の種類_既報

臨床研究以前の期間



②これまでの報告での
サウナを利用しない場合の
ストレス
(論文、手元データなど)



①臨床研究として
サウナを利用した対象者の
ストレス

臨床研究の期間

比較対照の種類_同じ研究の中

臨床研究以前の期間



①臨床研究として
サウナを利用した対象者の
ストレス

臨床研究の期間



③臨床研究として同じ時期に
サウナを利用しない対象者の
ストレス

比較対照の種類



単群試験 前後比較試験

- ① 研究対象者でサウナを利用しない時期のストレス

臨床研究以前の期間



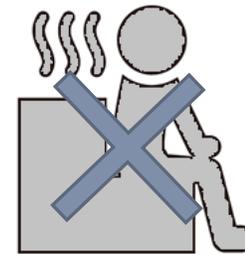
単群試験 ヒストリカル コントロール

- ② これまでの報告でのサウナを利用しない場合のストレス
(論文、手元データなど)



- ③ 臨床研究としてサウナを利用した対象者のストレス

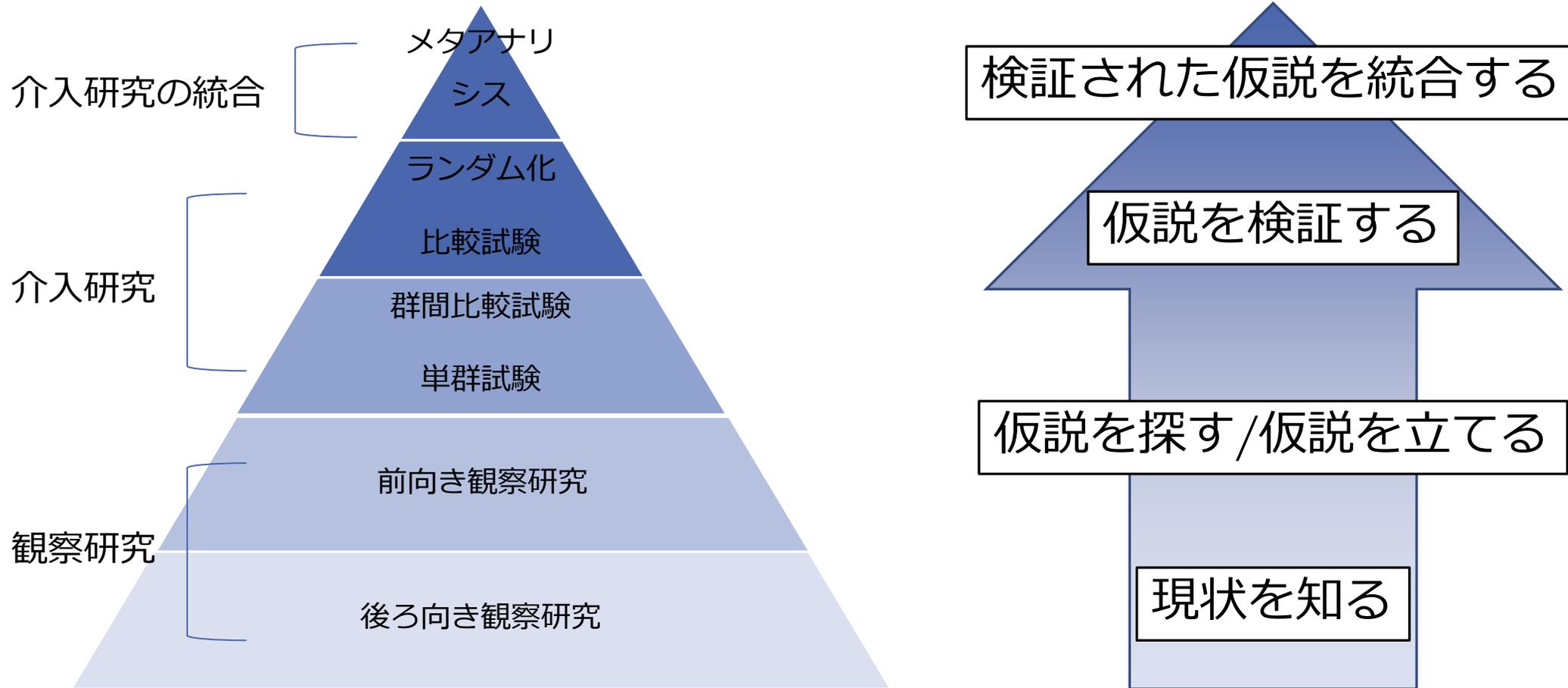
臨床研究の期間



並行群間比較試験

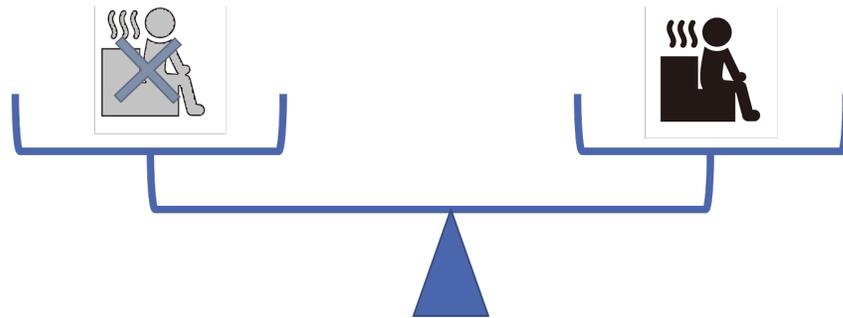
- ③ 臨床研究として同じ時期にサウナを利用しない対象者のストレス

比較の対象とエビデンスレベル

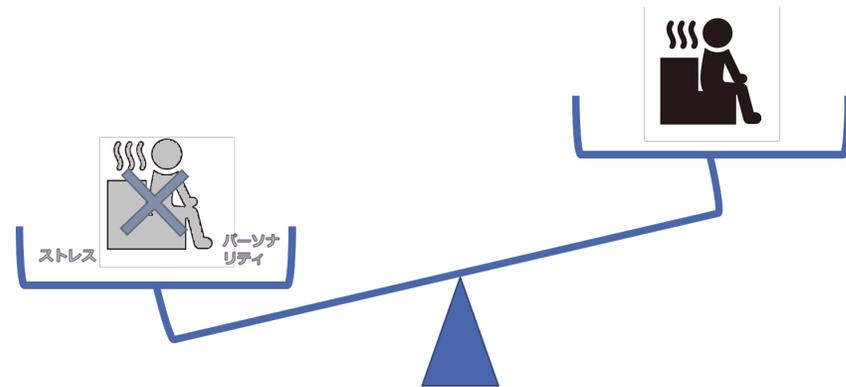


適切に比較をするために、

適切な比較



不適切な比較

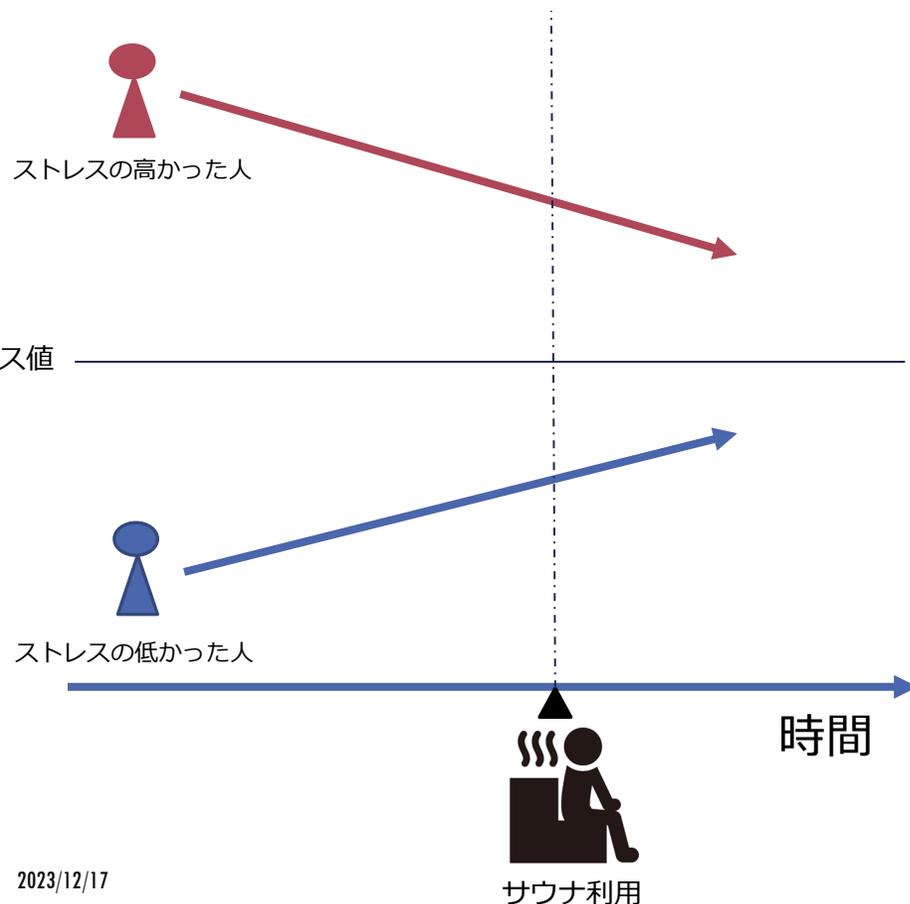


比較の背景がそろっている必要がある。

比較の背景をそろえる。

- ・ 個人で揃える ⇒ 同じ人で比較する。
前後比較試験（時間的な要素はそろわない）
- ・ 集団で揃える ⇒ いろいろな人がいるけれど、
サウナ利用ありとサウナ利用なしの
グループの平均としてそろえる。
ランダム化比較試験

前後比較の場合の問題点



- ・ 「平均への回帰」という現象の発生

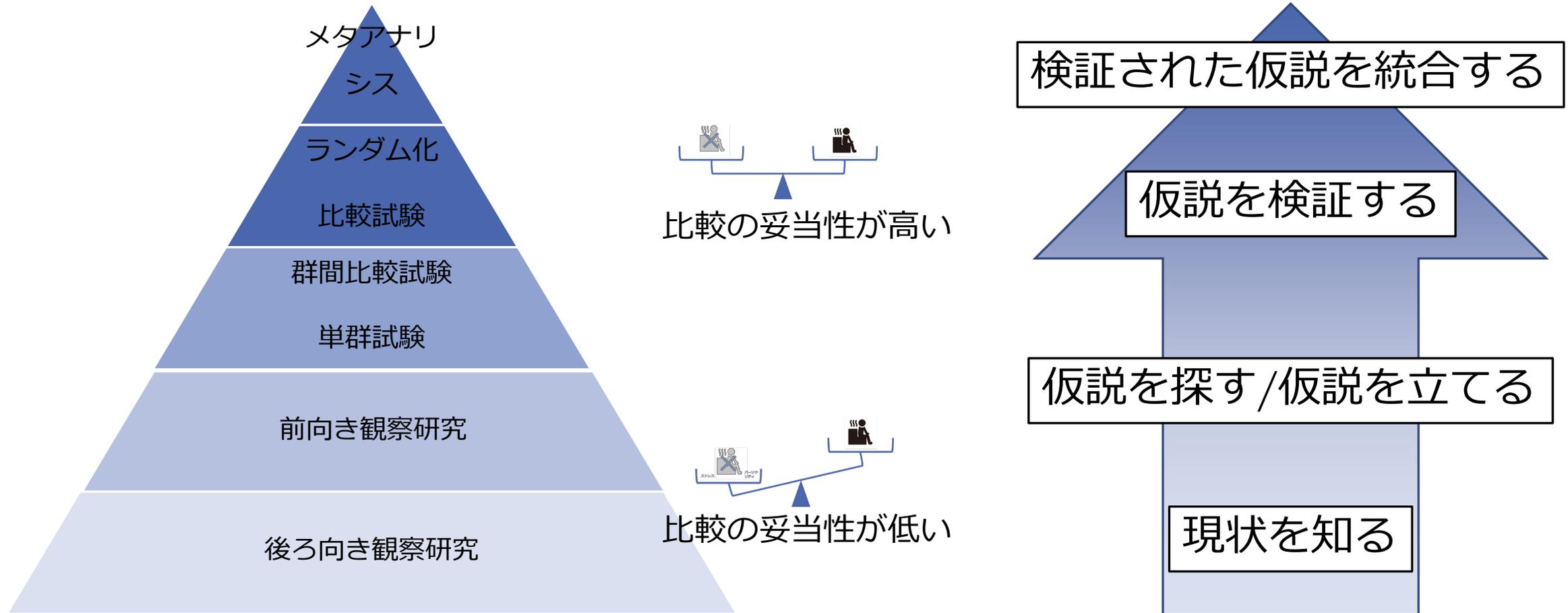
⇒ 「ある時点で特別に良かったり、悪かったりした値は、次に測定するときには平均的な値に近づく」という現象

前後比較試験での問題点：
サウナ利用前後で比較をしているのか、
「平均への回帰」を見ているのか
区別がつかない。

研究デザインと比較の妥当性

	個人毎の背景	時間的要因	集団としての背景	平均への回帰
単群試験 (前後比較)	○	×	△ (時間要素が×)	あり
単群試験 (ヒストリカル コントロール)	×	×	×	なし
並行群間比較試験 (非ランダム化)	×	○	△ (確実性に乏しい)	あるが群間で同じ
並行群間比較試験 (ランダム化)	×	○	○	あるが群間で同じ

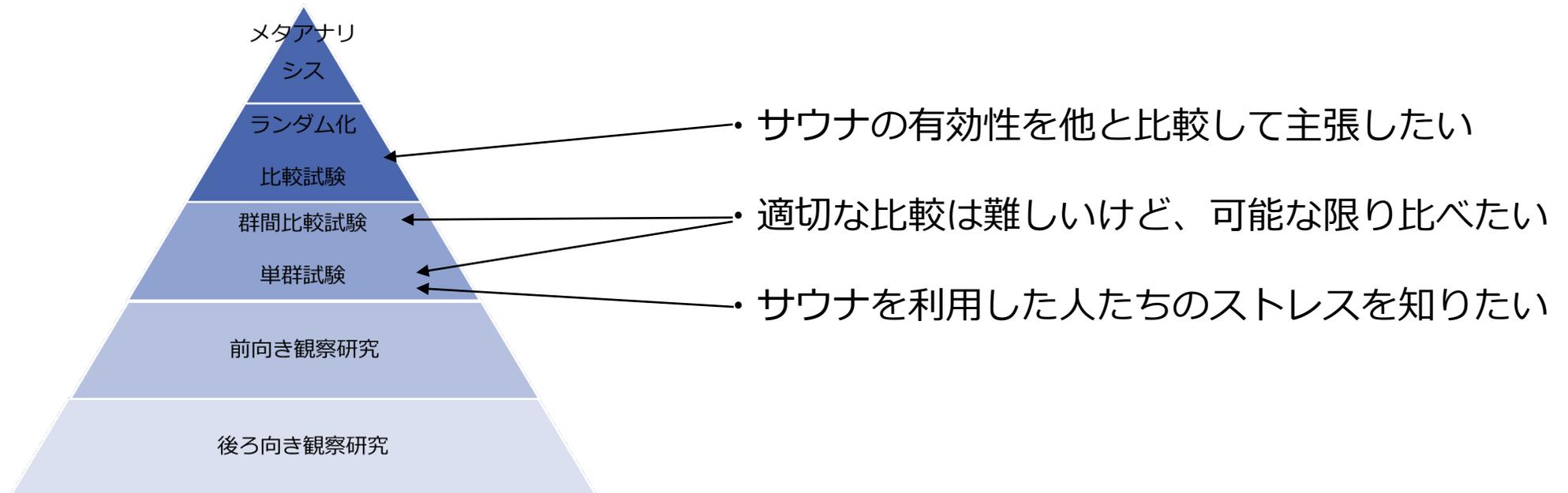
比較の対象とエビデンスレベル



研究デザインの選択

▶研究のゴール

薬事承認 / ガイドライン収載 / 論文発表 / 学位取得 / 学会発表



まとめ

- ▶ 治療方法の有効性を検討するには、それ以外の方法の場合と有効性を比較する必要がある。
- ▶ 有効性を比較する場合には比較の妥当性を考える必要がある。
- ▶ 比較の妥当性は研究デザインによって異なるためその研究の位置づけから適切な研究デザインを選択する必要がある。

サンプルサイズ

令和5年度 厚生労働省 臨床研究総合促進事業
医師研修
神戸大学医学部附属病院

2023/12/17

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること
- ▶ 間違いを制御したサンプルサイズであること
- ▶ 集積可能なサンプルサイズであること

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること
- ▶ 間違いを制御したサンプルサイズであること
- ▶ 集積可能なサンプルサイズであること

対象者の不利益を小さくする

「現在の最も有効な治療法は標準治療」

⇒ 研究に参加することで標準治療を受けられない。

⇒ 研究に参加することで新しい治療に因る未知の有害事象の可能性。

このような不利益を被る患者さんを最小限にしなければならない。

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること
- ▶ 間違いを制御したサンプルサイズであること
- ▶ 集積可能なサンプルサイズであること

効果の差とサンプルサイズ

小さな差を研究で差があるという
⇒多くのサンプルサイズ



サウナを
利用しない群

ストレスある人
25%



サウナを
利用する群

ストレスある人
20%

必要サンプルサイズ：各群約1100人

大きな差を研究で差があるという
⇒少ないサンプルサイズ



サウナを
利用しない群

ストレスある人
40%



サウナを
利用する群

ストレスある人
15%

必要サンプルサイズ：各群約60人

治療法の効果の大きさ（差）によって
必要なサンプルサイズが異なる。

少ないサンプルサイズでの研究

小さな差を研究で差があるという
⇒多くのサンプルサイズ



サウナを
利用しない群

ストレスある人
25%

必要サンプルサイズ：各群約1100人



サウナを
利用する群

ストレスある人
20%



ある臨床研究



サウナを
利用しない群

ストレスある人
25%



サウナを
利用する群

ストレスある人
20%

(とりあえず)各群 60人

オッズ比 0.75(95% 信頼区間0.29 to 1.93)
P = 0.662

少ないサンプルサイズでの研究

研究結果における判断

- オッズ比が1より小さい
⇒サウナ利用の有無で
ストレスに差があった。
- 信頼区間が1をまたぐ
⇒サウナ利用の有無で
ストレスに差はなかった
- p値が0.05（有意水準より大きい）
⇒サウナ利用の有無での
ストレスの差については
どちらとも判断ができない。

ある臨床研究



サウナを
利用しない群

ストレスある人
25%



サウナを
利用する群

ストレスある人
20%

各群 60人

オッズ比 0.75(95% 信頼区間0.29 to 1.93)
P = 0.662

少ないサンプルサイズでの研究

研究結果における判断

- オッズ比が1より小さい
⇒サウナ利用の有無で
ストレスに差があった。
- 信頼区間が1をまたぐ
⇒サウナ利用の有無で
ストレスに差はなかった
- p値が0.05（有意水準より大きい）
⇒サウナ利用の有無での
ストレスの差については
どちらとも判断ができない。

ある臨床研究



サウナを
利用しない群

ストレスある人
25%



サウナを
利用する群

ストレスある人
20%

各群 60人

オッズ比 0.75(95% 信頼区間0.29 to 1.93)
P = 0.662

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること（=期待する効果に基づいている）
- ▶ 間違いを制御したサンプルサイズであること
- ▶ 集積可能なサンプルサイズであること

期待する効果

- ▶ 連続量の場合（例：血圧、検査値など）

 - 群間の差⇒差が小さいほど必要サンプルサイズは多い

 - 標準偏差⇒標準偏差（ばらつき）が大きいほど必要サンプルサイズは多い

- ▶ カテゴリーの場合（例：有効/無効など）

 - 各群の割合⇒50%に近くなると必要サンプルサイズは多い

 - 割合の差⇒差が小さいほど必要サンプルサイズは多い

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること（=期待する効果に基づいている）
- ▶ 間違いを制御したサンプルサイズであること
- ▶ 集積可能なサンプルサイズであること

臨床研究で起こり得る間違い

		本当の世界	
		効果に差がある	効果に差がない
研究の判断	効果に差があると判断する	期待する結果であり正しい判定	Type1 エラー (α エラー)
	効果に差がないと判断する	Type2 エラー (β エラー)	期待する結果ではないが正しい判定

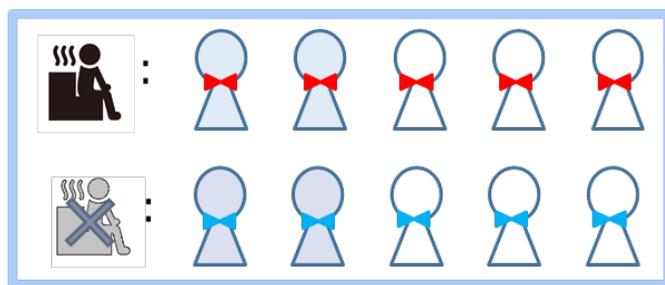
Type1 エラー、Type2 エラーを許容可能な小ささにする。
(ゼロにするのではない)

臨床研究で起こり得る間違い

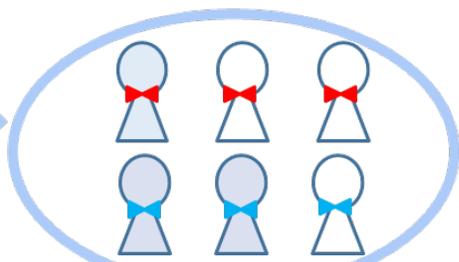
		本当の世界	
		効果に差がある	効果に差がない
研究の判断	効果に差があると判断する	期待する結果であり正しい判定	Type1 エラー (α エラー)
	効果に差がないと判断する	Type2 エラー (β エラー)	期待する結果ではないが正しい判定

Type1エラー

効果は等しい
本当の世界A

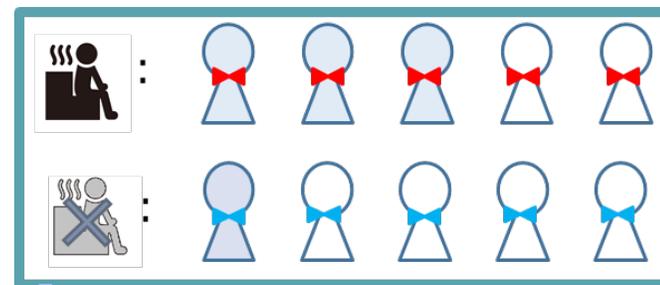


研究結果
(差がある)

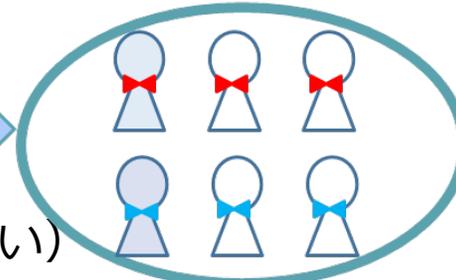


Type2エラー

効果は異なる
本当の世界B



研究結果
(差がない)



臨床研究で起こり得る間違い

		本当の世界	
		効果に差がある	効果に差がない
研究の判断	効果に差があると判断する	期待する結果であり正しい判定	Type1 エラー (α エラー)
	効果に差がないと判断する	Type2 エラー (β エラー)	期待する結果ではないが正しい判定

➤ Type 1 エラー (α エラー)

通常は有意水準として両側 5%。緩める場合でも両側10%まで。

(論文報告時の有意水準になるので、そのことも意識しておきましょう。)

➤ Type2エラー (β エラー)

多くの場合は20% (1-Type2エラー(=検出力)として80%)

緩める場合で検出力として70%まで。

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること（=期待する効果に基づいている）
- ▶ 間違いを制御したサンプルサイズであること（=エラーの大きさを考える）
- ▶ 集積可能なサンプルサイズであること

サンプルサイズに求められるもの

▶ 対象者の不利益を小さくするサンプルサイズであること

▶ 判断可能なサンプルサイズであること (=期待する効果に基づいている)

▶ 間違いを制御したサンプルサイズであること (=エラーの大きさを考える)

▶ 集積可能なサンプルサイズであること

主要評価項目とその解析方法が決まらなければ
効果の大きさも決まらない。

サンプルサイズに求められるもの

- ▶ 対象者の不利益を小さくするサンプルサイズであること
- ▶ 判断可能なサンプルサイズであること（=期待する効果に基づいている）
- ▶ 間違いを制御したサンプルサイズであること（=エラーの大きさを考える）
- ▶ 集積可能なサンプルサイズであること

集積可能であること

研究開始後に集積できなければ、

研究体制の変更

少ない症例数で中止

延々と期間延長

研究開始前にできること

適格基準にあう対象者数の推定

実施体制の構築

主要評価項目の再検討

研究の位置づけの再検討

研究デザインの再検討

まとめ

- ▶ 対象者の不利益を小さくするサンプルサイズである必要がある。
- ▶ サンプルサイズは、期待する効果の大きさと、判断を間違える確率を設定することで算出される。
- ▶ 期待する効果に合わないサンプルサイズでの研究は、適切な判断を行えない。
- ▶ 研究開始前に集積可能かどうかを十分に検討し、困難が予想される場合には研究計画を再検討する。