

臨床研究と生物統計

令和7年度 厚生労働省 臨床研究総合促進事業
臨床研究・治験従事者研修
神戸大学医学部附属病院

2025/12/14

動画の構成

- ランダム化比較試験
- サンプルサイズの考え方

ランダム化比較試験

令和7年度 厚生労働省 臨床研究総合促進事業
臨床研究・治験従事者研修
神戸大学医学部附属病院

2025/12/14

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

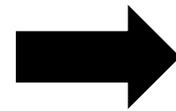
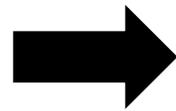
- ・ 因果関係を調べたい（新しい治療方法は有効なのか）
- ・ 予後に関連ある因子を調べたい
- ・ 診断方法を確立したい（感度・特異度など）

介入研究における
研究目的

治療法が“有効”であるとは

【例：睡眠】

睡眠前にホットミルクを飲むと
睡眠の質が改善するかを検討したい。



介入：
ホットミルクを飲む

ホットミルクは睡眠の質に有効か

- ハナコさんは睡眠の質に不満がありました。
- ハナコさんは睡眠前にホットミルクを飲みました。
- その日はぐっすり眠りました。

ホットミルクは睡眠の質
に効果があります。



ホットミルクは睡眠の質に有効か

- ハナコさんは睡眠の質に不満がありました。
- ハナコさんは睡眠前にホットミルクを飲みました。
- その日はぐっすり眠りました。

ハナコさんには効果
があった？
他の人には？
たまたま？

ホットミルク？は睡眠の質
に効果があります。



ホットミルクは睡眠の質に有効か

- 睡眠の質に不満がある人を100人集めました。
- 100人は睡眠前にホットミルクを飲みました。
- 50人はぐっすり眠りました。

ホットミルクは睡眠の質
に効果があります。



ホットミルクは睡眠の質に有効か

- 睡眠の質に不満がある人を100人集めました。
- 100人は睡眠前にホットミルクを飲みました。
- 50人はぐっすり眠りました。

ホットミルク？は睡眠の質
に効果が？あります。

不眠の要素がなくなっ
たのでは？
たまたま？



有効性を示すために必要なこと

▶適切な比較が必要

- 比べたい要素（介入）以外の結果に影響を与える要因が同じ状態で介入（ホットミルク）がある状態とない状態を比較する。

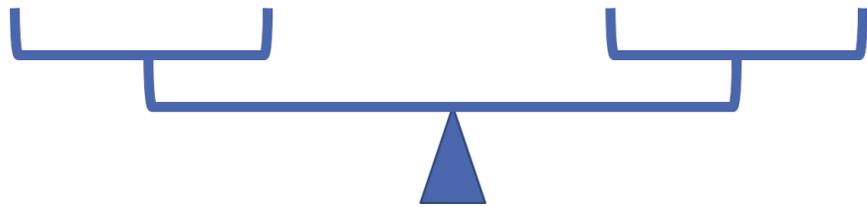
この状況を実験的に作り出そうとするものが

「ランダム化並行群間比較試験」

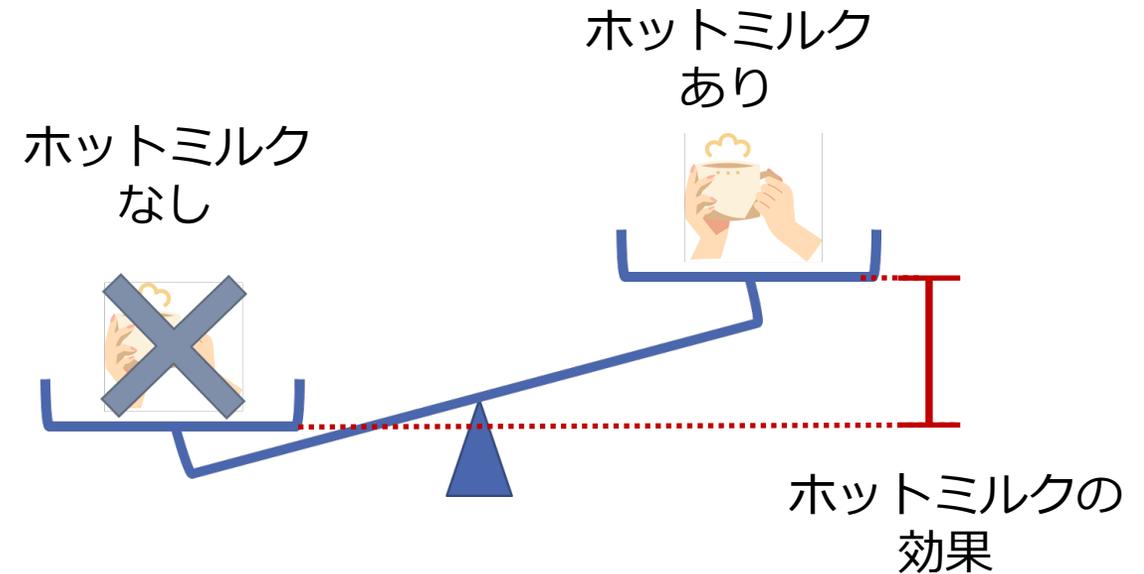
Randomized control trial (RCT)

比較をして検討する-適切な比較-

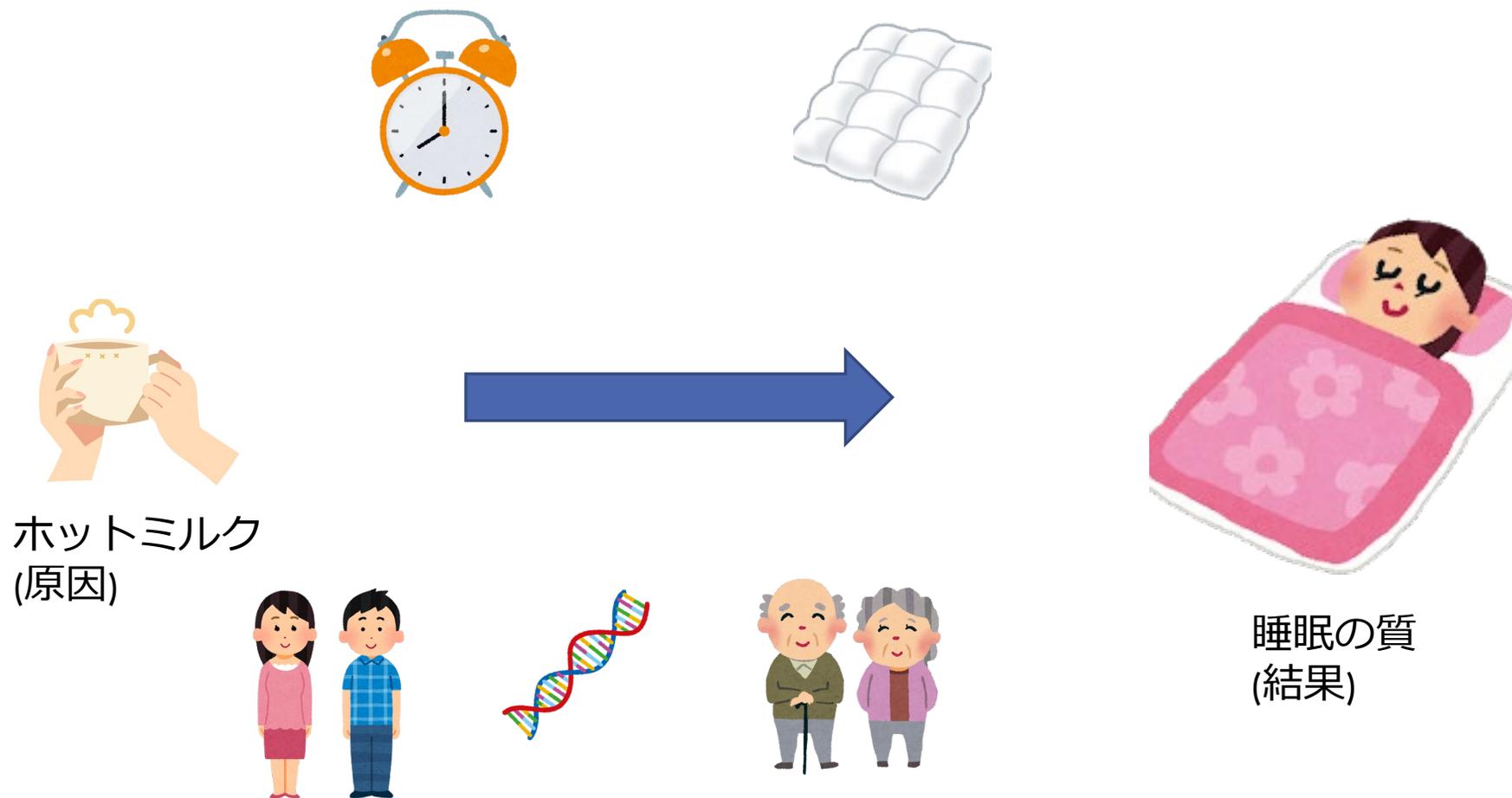
比較前



比較後

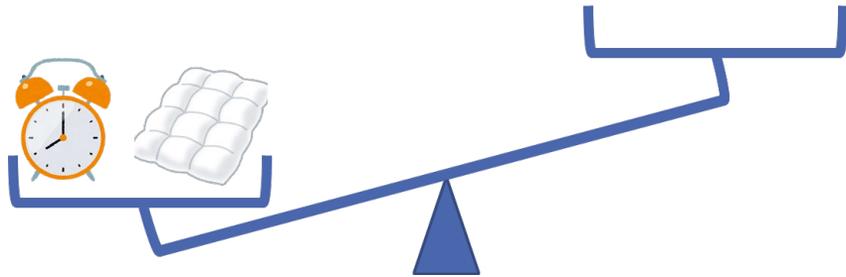


睡眠の質に影響のある要素



比較をして検討する-不適切な比較-

比較前



比較後

ホットミルク
なし



ホットミルク
あり



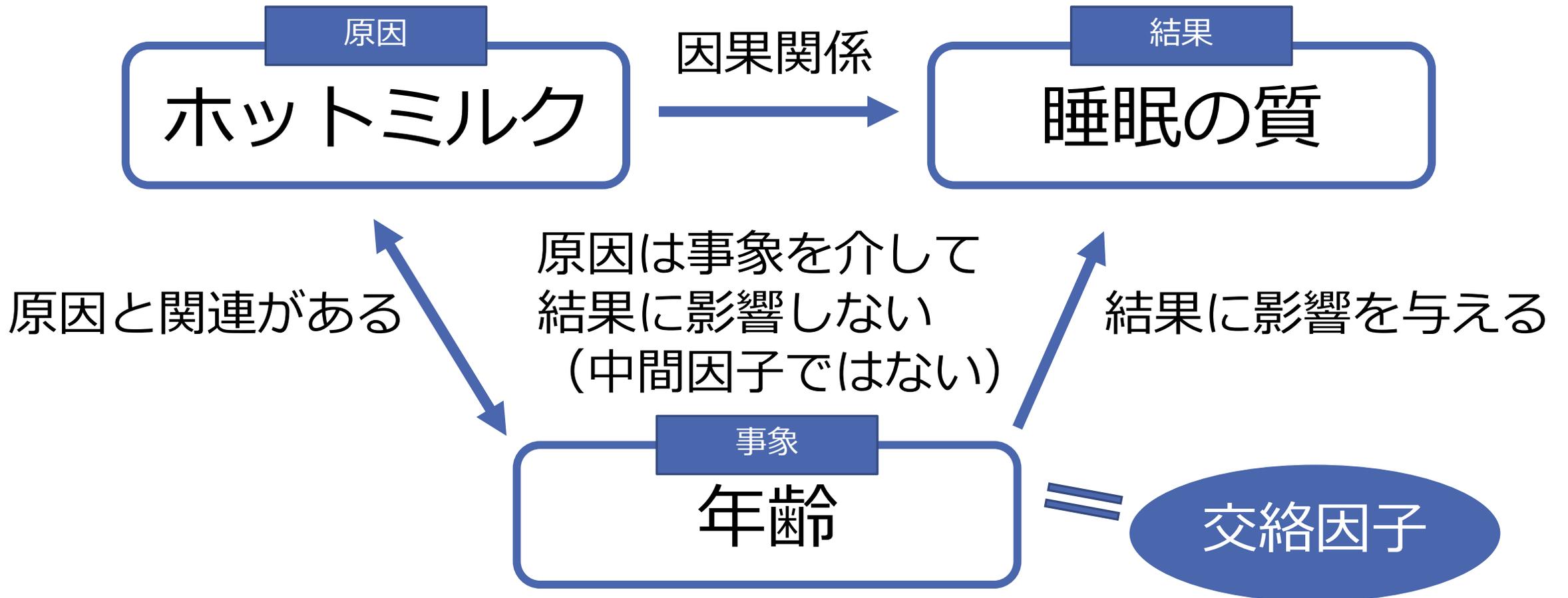
ホットミルクの
効果？

比較を邪魔する要素

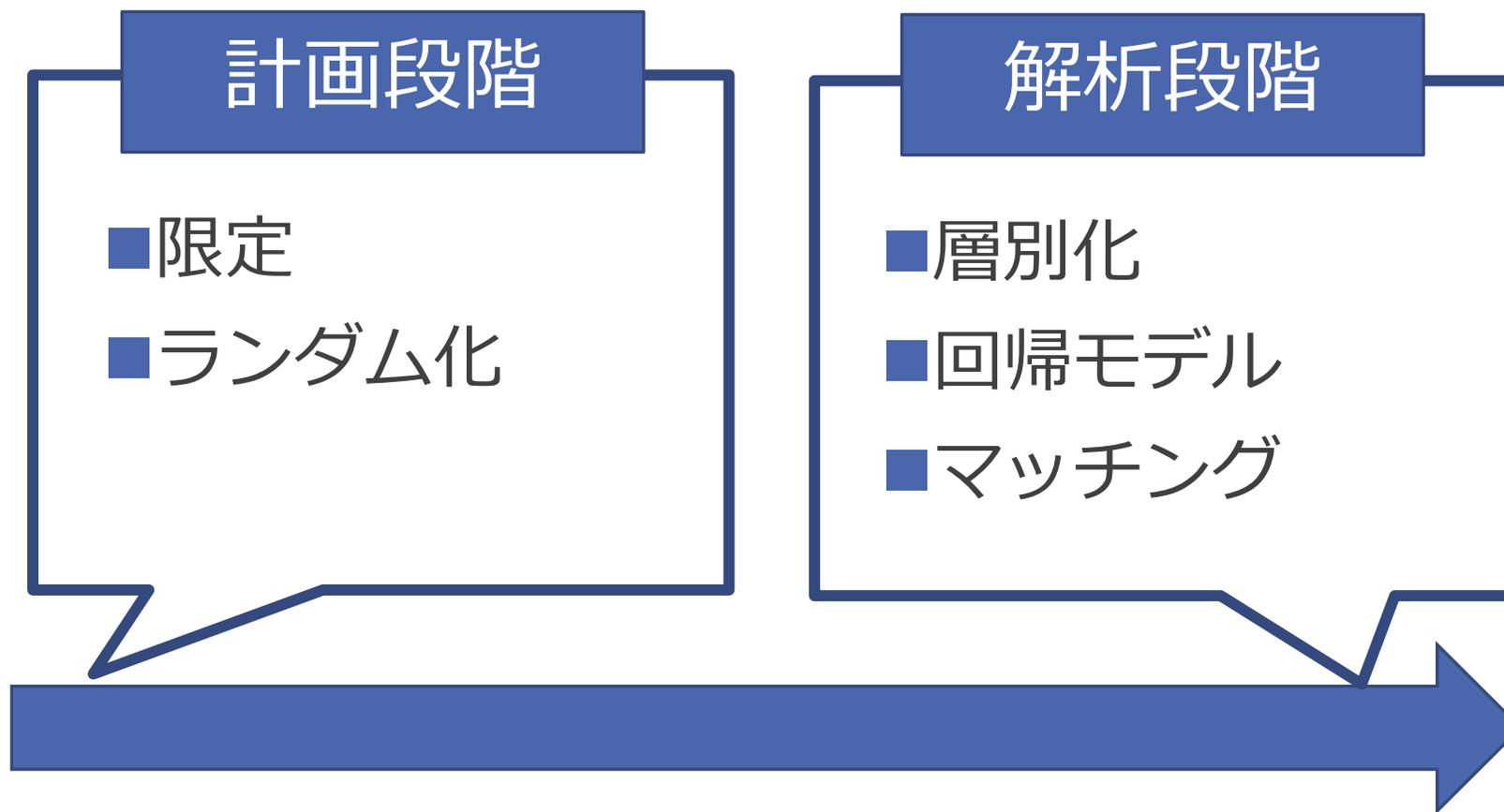
- 交絡

検討したい要素（介入）と結果との関係が、別の因子によって説明されてしまう状態のこと

比べたい要素以外の結果に影響を与える要因-交絡-



交絡の調整・制御の方法



解析段階での交絡制御

▶ 層別化（層化）

交絡因子によりグループにわけて、そのグループごとに解析を行い結果を統合する、

▶ 回帰モデル（多変量解析）

統計モデルに変数として交絡因子を含め、それぞれ因子の影響をみる。

▶ マッチング

比較したい治療法間で性別や年齢などの交絡因子が合う様に対象者を選ぶ。

解析段階での交絡制御

▶ 層別化（層化）

交絡因子によりグループにわけて、そのグループごとに解析を行い結果を統合する。

▶ 回帰

統計

関連があることが分かっている因子は制御可能

関連があることが分かっていない因子、測定されていない因子

は制御できない

みる。

▶ マッチング

比較したい治療法間で性別や年齢などの交絡因子が合う様に対象者を選ぶ。

計画段階での交絡調整①

▶限定

交絡因子をもつ集団を除外する

→適格性（選択基準・除外基準）

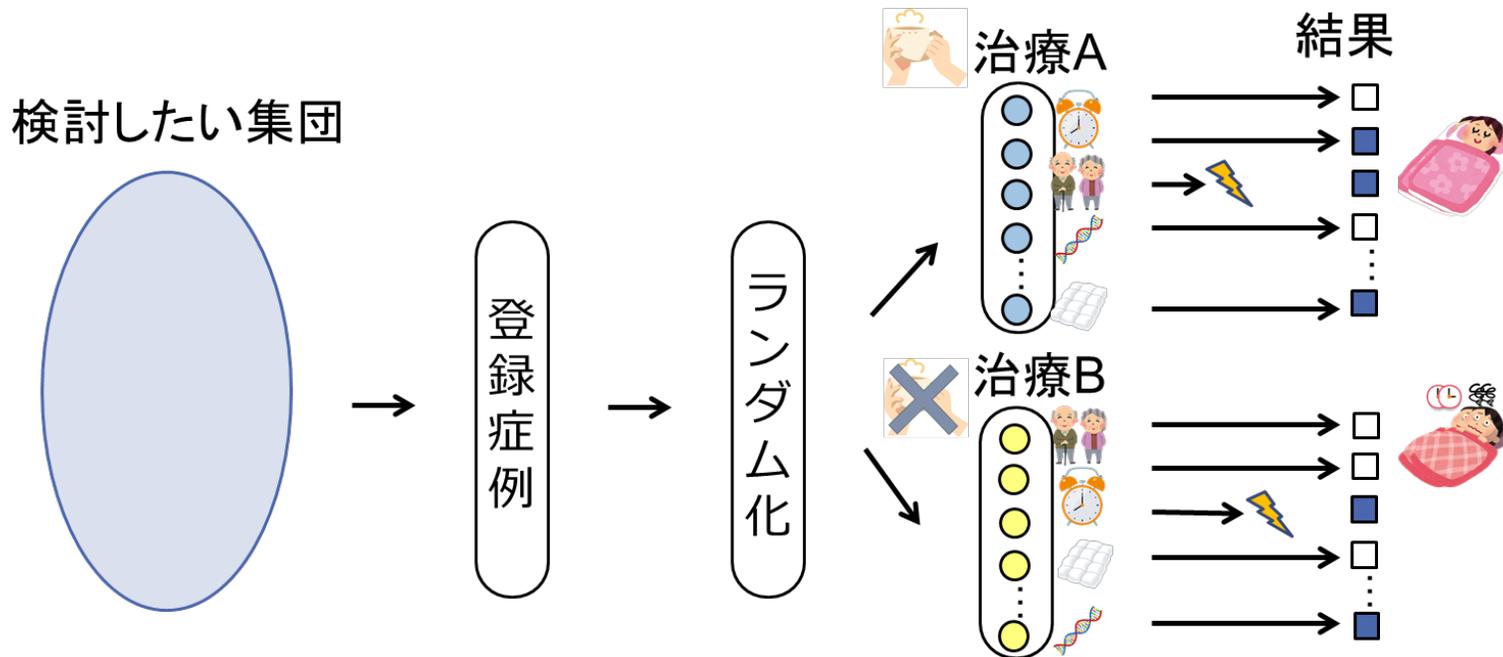
結果を当てはめられる集団が限定的である。

分かっている測定できる因子しか制御できない。

計画段階での交絡調整②

▶ ランダム化（無作為化）

1つの集団をランダムにより複数のグループに分割し、異なる治療を施し追跡した後の結果をグループ間で比べる方法



計画段階での交絡調整②

▶ ランダム化（無作為化）

1つの集団をランダムにより複数のグループに分割し、異なる治療を施し追跡した後の結果をグループ間で比べる方法

平均的に均一なグループを作ることができる。

分かっているが測定できない交絡因子のバランスを保つことができる。

分かっていない交絡因子のバランスを保つことができる。

比較を邪魔する要素

- 交絡

検討したい要素（介入）と結果との関係が、別の因子によって説明されてしまう状態のこと

➡ ランダム化並行群間比較試験とすることで、ある程度制御可能

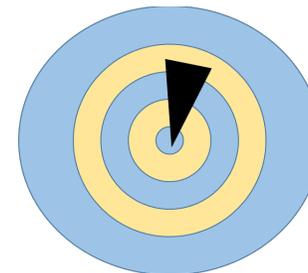
ランダム化比較試験において 交絡以外の比較を邪魔する要因

- 系統誤差(バイアス)

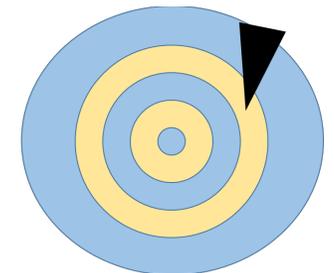
治療効果の推定が一定の向きをもって本当の値とずれてしまうこと

- 脱落

ランダム化比較試験として交絡因子のバランスをとっても、脱落があるとそのバランスを崩すことになる。



見たいところ



バイアスがある状態

系統誤差(バイアス)の種類

- ▶ 選択バイアス
- ▶ 割付バイアス
- ▶ 情報バイアス

選択バイアス SELECTION BIAS

患者さんを登録する際に生じるバイアス

- 例) 患者さんを限定してしまったり、想定している対象者とは異なる対象者のいる施設で患者さんを集める。
効果が出そうな人を集める。
→ 本当に検討したい対象者からずれる。

選択バイアスを避けるために

- ▶ 選択基準・除外基準で適切な対象者を検討できる様にする。
- ▶ 割付情報の秘匿：
次に来た患者さんがどの治療法になるかを分からない状態で実施する。
- ▶ 盲検化：
研究者にも患者さんにもどの治療法かを分からないようにする。

割付バイアス ALLOCATION BIAS

ランダムに割り付けた結果として、
結果に影響を与える要因が
グループ間で偶然に偏ってしまうために結果が本当の値からずれること。

例) 年齢が結果に影響する場合に、偶然にグループ間で年齢分布が偏る。
→検討したい介入以外の効果を見ることになる。

割付バイアスを避けるために

- 割付方法を工夫する。(研究計画時)
 - ブロック割付の利用。
 - 層別ランダム化、最小化法。
- 解析手法で対応する。(解析時)
 - 偏りの疑いがある項目で調整した評価指標の推測を行う。

情報バイアス INFORMATION BIAS

データを測定する際に生じるずれ。

測定者によるもの・・・測定者によって測定の方法に特徴がある。
研究対象者によるもの・・・「よい結果になるだろ」と思って評価する。
測定手法によるもの・・・機器によってばらつきが大きい。 など

例) 主観的なスコアを使う研究で、グループを知っているため、
「効果があるだろう」と思ってスコアを付ける。
→真の値からずれたスコアになる。

情報バイアスを避けるために

- ▶ 客観的な評価項目を使う。
評価項目が主観的な場合は、効果の判定のための第三者の評価者が行う。
- ▶ 研究者等や対象者どちらの治療を行っているかをわからないようにする。
 - 盲検化
 - プラセボ（見た目は同じだが効果のない薬）の使用

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、適切に比較する必要がある。
(治療法以外の差が生じていないか)
- ▶ ランダム化は適切に比較するために交絡を制御するよい方法である。
- ▶ 臨床研究ではランダム化のみでは比較の適切性が担保されないため研究計画時には十分に検討する必要がある。

サンプルサイズの方

令和7年度 厚生労働省 臨床研究総合促進事業
臨床研究・治験従事者研修
神戸大学医学部附属病院

2025/12/14

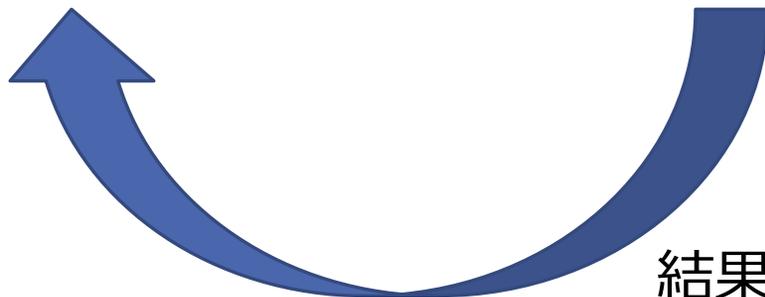
検討したい集団
= 母集団



臨床
研究



母集団を反映する症例数で
臨床研究を実施し、

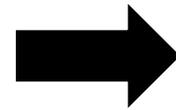
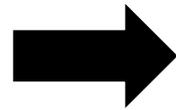


結果から正しい判断を下す。

治療法が“有効”であるとは

【例：睡眠】

睡眠前にホットミルクを飲むと
睡眠の質が改善するかを検討したい。



介入：
ホットミルクを飲む

臨床研究における判断

睡眠前にホットミルクを飲むと
睡眠の質が改善するかを検討したい。



- 研究における判断① ホットミルクは睡眠の質を改善する。
- 研究における判断② ホットミルクは睡眠の質を改善しない。
- 研究における判断③ ホットミルクが睡眠の質を改善するか分からない。

サンプルサイズ

この人数を調べたら、**高い確率**で
母集団を反映した
研究結果になるだろう。

検出力

母集団を反映する症例数で
臨床研究を実施し、

検討したい集団
= 母集団

調査したい本当の値
= 真の値

臨床研究

臨床研究の結果

効果量

ホットミルクには
これだけの睡眠改善の
効果があるだろう

有意水準

研究の結果から
ホットミルクには睡眠
改善の効果がある/ない
と**判断**する。

結果から正しい判断を下す。

効果量

- 「睡眠の質」を評価する候補
 - ・ 自覚的な眠りの満足度
 - ・ 入眠までの時間
 - ・ 睡眠時間
 - ・ 中途覚醒の回数
 - ・ 日中の眠気

⇒ 最も検討したい項目を「主要評価項目」として設定する。



睡眠の質？

主要評価項目に求められるもの

- ▶ 検討したいものを最も代表する項目であること
- ▶ 測定可能であること
- ▶ 客観的な指標であること
- ▶ 評価が確立されている指標であること

期待する効果量

▶例えば「入眠までの時間」を評価した場合

ホットミルクあり
入眠までの時間
・・・30分



ホットミルクなし
入眠までの時間
・・・60分

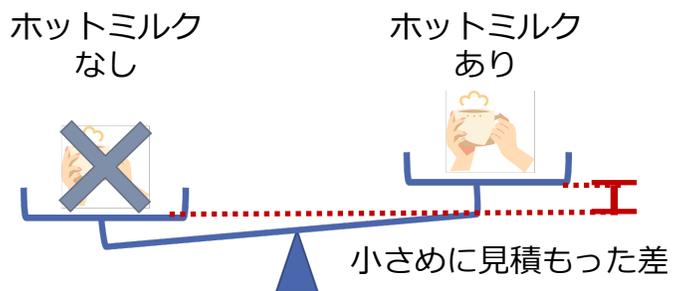


差の30分が効果量

この差が臨床的に意味がある差か、ということを考え、
この差を検出するために必要なサンプルサイズを設計する。

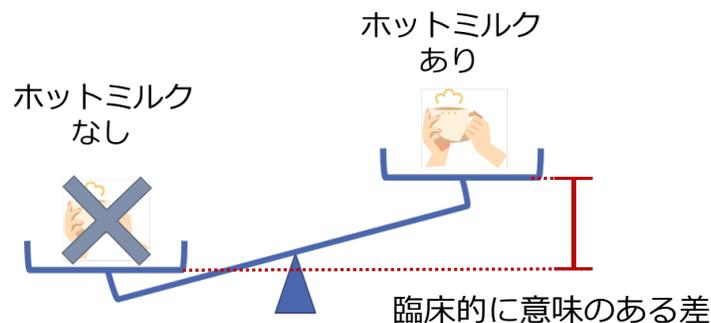
想定する効果量とサンプルサイズの関係

小さめに効果量を見積もる場合



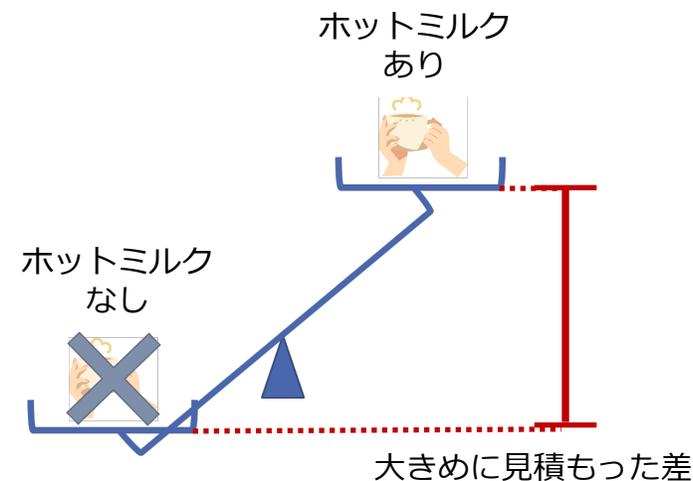
- 症例数：多くなる。
- 差を検出できても意味のない差であり、ホットミルクを試みる価値はないかもしれない。

適切に効果量を見積もる場合



- 症例数：基準。
- 差を検出できれば、ホットミルクを試みる価値がある。
- 差を検出できなければ、ホットミルクを試みる価値はない。

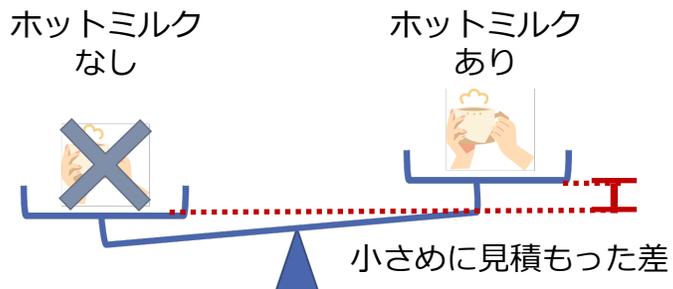
大きめに効果量を見積もる場合



- 症例数：少なくなる。
- 差を検出できれば、ホットミルクを試みる価値がある。
- 差を検出できなければ、ホットミルクを試みる価値は不明。

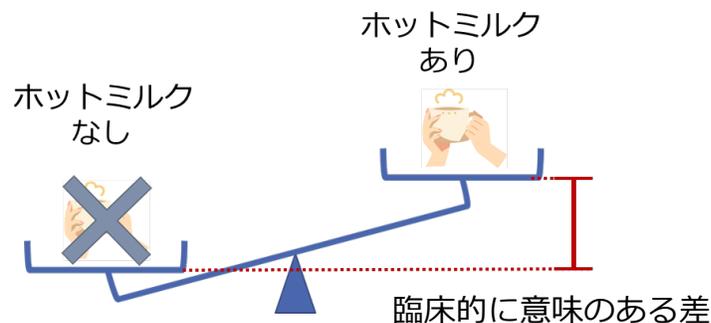
想定する効果量とサンプルサイズの関係

小さめに効果量を見積もる場合



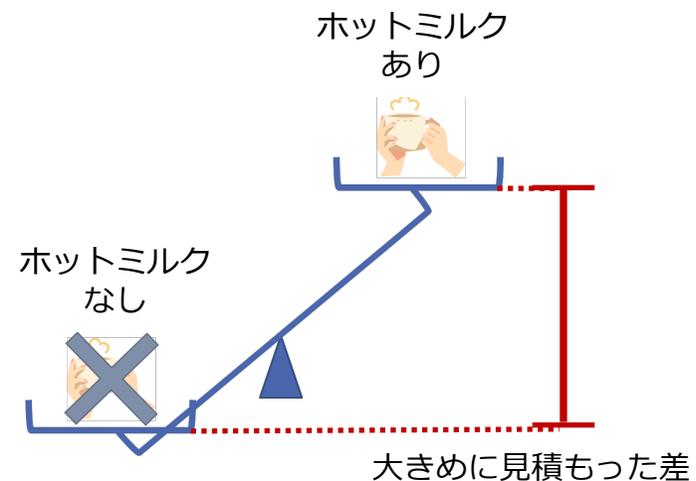
小さな差を検出できるが、労力をかけて、意味のない差を示すことになる。

適切に効果量を見積もる場合



どの様な場合も臨床的に意味のある判断ができる。

大きめに効果量を見積もる場合



差が検出できた場合には臨床的に意味のある判断ができる。差が検出できない場合には、判断ができない。

サンプルサイズ

この人数を調べたら、**検出力**
高い確率で
母集団を反映した
研究結果になるだろう。

母集団を反映する症例数で
臨床研究を実施し、

検討したい集団
= 母集団

調査したい本当の値
= 真の値

臨床研究

臨床研究の結果

有意水準

効果量

ホットミルクには
これだけの睡眠改善の
効果があるだろう

研究の結果から
ホットミルクには睡眠
改善の効果がある/ない
と判断する。

結果から正しい判断を下す。

検出力と有意水準

▶ 検出力： $1 - \beta$ エラー

▶ 有意水準： α エラー

臨床研究における判断

研究における判断① ホットミルクは睡眠の質を改善する。

研究における判断② ホットミルクは睡眠の質を改善しない。

臨床研究で起こり得る間違い

		母集団	
		効果に差がある (ホットミルクは睡眠の質を改善する。)	効果に差がない (ホットミルクは睡眠の質を改善しない)
研究の判断	p値 ≤ 有意水準 効果に差があると判断する (ホットミルクは睡眠の質を改善する)	期待する結果であり正しい判定 ・ 検出力 (1-βエラー)	本当は効果がないのに 効果があると判断する間違い ・ αエラー
	p値 > 有意水準 効果に差がないと判断する (ホットミルクは睡眠の質を改善しない)	本当は効果があるのに 効果がないという間違い ・ βエラー	期待する結果ではないが 正しい判定

βエラー：一般に20%
(検出力として80%)

臨床研究で起こり得る間違い

		母集団	
		効果に差がある (ホットミルクは睡眠の質を改善する。)	効果に差がない (ホットミルクは睡眠の質を改善しない)
研究の判断	p値 ≤ 有意水準 効果に差があると判断する (ホットミルクは睡眠の質を改善する)	期待する結果であり正しい判定 ・ 検出力 (1-βエラー)	本当は効果がないのに 効果があると判断する間違い ・ αエラー
	p値 > 有意水準 効果に差がないと判断する (ホットミルクは睡眠の質を改善しない)	本当は効果があるのに 効果がないという間違い ・ βエラー	期待する結果ではないが 正しい判定

αエラー：一般に両側5%

臨床研究で起こり得る間違い

		母集団	
		効果に差がある (ホットミルクは睡眠の質を改善する。)	効果に差がない (ホットミルクは睡眠の質を改善しない)
研究の判断	p値 ≤ 有意水準 効果に差があると判断する (ホットミルクは睡眠の質を改善する)	期待する結果であり正しい判定 ・ 検出力 (1-βエラー)	本当は効果がないのに 効果があると判断する間違い ・ αエラー
	p値 > 有意水準 効果に差がないと判断する (ホットミルクは睡眠の質を改善しない)	本当は効果があるのに 効果がないという間違い ・ βエラー	期待する結果ではないが 正しい判定

1

2

臨床研究で起こり得る間違い

	効果に差がある (ホットミルクは睡眠の質を改善する。)	効果に差がない (ホットミルクは睡眠の質を改善しない)	
研究の判断	<p>p値 ≤ 有意水準 効果に差があると判断する (ホットミルクは睡眠の質を改善する)</p>	<p>期待する結果であり正しい判定 ・ 検出力 (1-βエラー)</p>	<p>本当は効果がないのに 効果があると判断する間違い ・ αエラー</p> <p>1</p>
研究の判断	<p>p値 > 有意水準 効果に差がないと判断する (ホットミルクは睡眠の質を改善しない)</p>	<p>本当は効果があるのに 効果がないという間違い ・ βエラー</p>	<p>期待する結果ではないが 正しい判定</p> <p>2</p>

これらのエラーを含んだうえで
研究結果に基づき何らかの判断を下すために、サンプルサイズを検討する。

サンプルサイズ

この人数を調べたら、**高い確率**で
母集団を反映した
研究結果になるだろう。

検出力

母集団を反映する症例数で
臨床研究を実施し、

検討したい集団
= 母集団

調査したい本当の値
= 真の値

臨床研究

臨床研究の結果

効果量

ホットミルクには
これだけの睡眠改善の
効果があるだろう

有意水準

研究の結果から
ホットミルクには睡眠
改善の効果がある/ない
と**判断**する。

結果から正しい判断を下す。

サンプルサイズの計算

R 2群の平均値の比較のためのサンプルサイズの計算

2群間の平均値の差	30	効果量
2群共通の標準偏差(SD)	45	
α エラー (0.0-1.0)	0.05	α エラー(有意水準)
検出力(1- β エラー) (0.0-1.0)	0.80	検出力
グループ1と2のサンプルサイズの比 (1:X)	1	

方法

両側

One-sided

計算に必要な情報

主要評価項目： 入眠までの時間
効果量： 30分
 α エラー： 5%
検出力： 80%

計算されたサンプルサイズ

2群間の平均値の差	仮定
標準偏差	30
α エラー	45
検出力	0.05
N2とN1のサンプルサイズの比	両側検定
必要サンプルサイズ	0.8
N1	1
N2	計算結果
	36
	36

サンプルサイズ

この人数を調べたら、
高い確率で
母集団を反映した
研究結果になるだろう。

検出力

母集団を反映する症例数で
臨床研究を**実施し**、

検討したい集団
= 母集団

調査したい本当の値
= 真の値

臨床研究

臨床研究の結果

効果量

ホットミルクには
これだけの睡眠改善の
効果があるだろう

有意水準

研究の結果から
ホットミルクには睡眠
改善の効果がある/ない
と**判断**する。

結果から正しい判断を下す。

症例を集積するために

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

- ▶ サンプルサイズの検討には、効果量、検出力、 α エラーが必要。
- ▶ 効果量は「どれくらいの差なら臨床的に意味があるか」を考える。
- ▶ 研究結果から判断できる臨床研究にするために、効果量や判断の間違いの程度を想定して必要なサンプルサイズを検討する。