

倫理審査委員が知っておくべき 生物統計学の基礎知識

国立がん研究センター
水澤 純基

2025.8.23 (土)

2025年度倫理審査委員会・治験審査委員会委員養成研修

今日の話題

1 交絡とランダム化

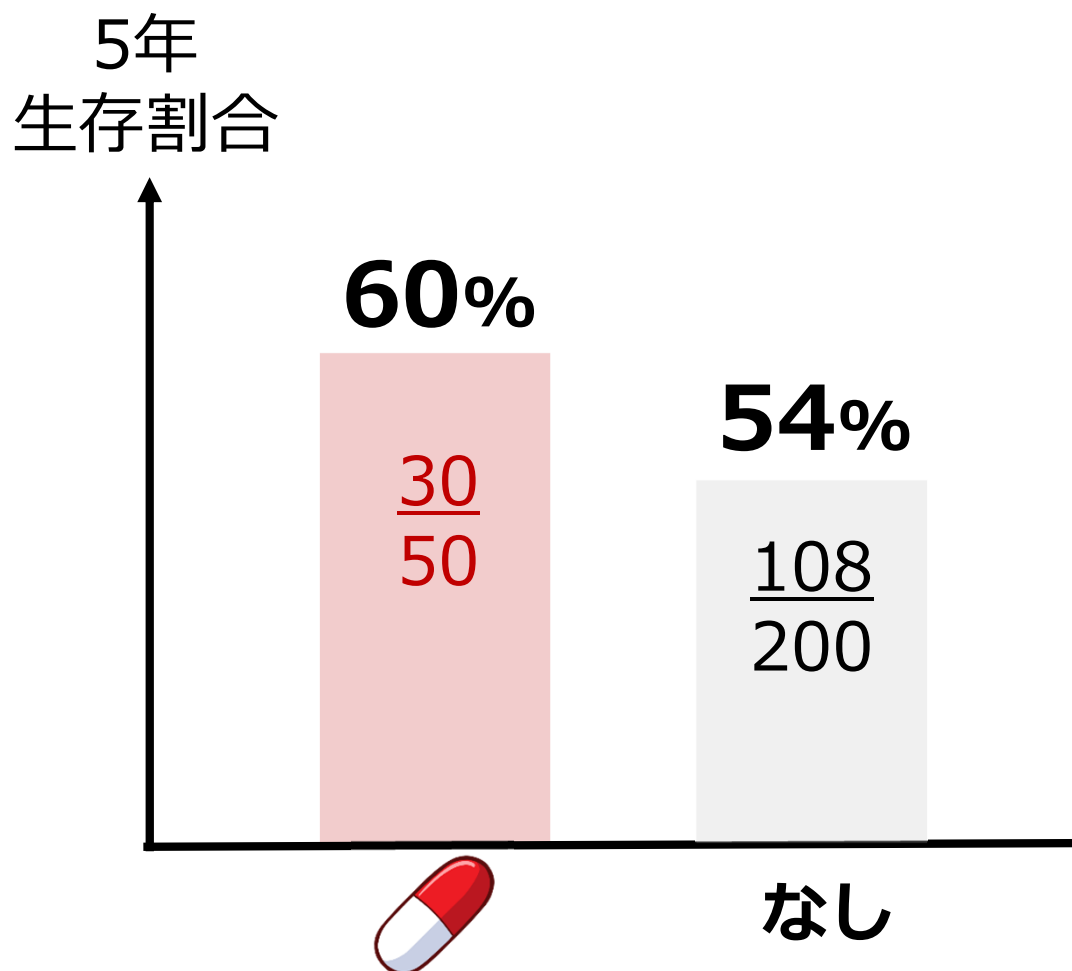
2 研究目的と研究デザイン

3 統計モデルで交絡を調整するとは

4 倫理審査Q&A

よくある研究

当院の80歳以上の胃切除後の胃がん患者のうち、
術後化学療法を受けた患者 (50例)と**受けなかった患者** (200例)の
胃切除後5年時点で生存している患者の割合 (5年生存割合) を評価



術後化学療法の方が
5年生存割合が改善



80歳以上の高齢者にも
術後化学療法を
施行した方が良い

注：仮想例

術後の体調で分けた場合の5年生存割合

術後の体調**良好**の
患者さん



5年
生存割合

67%

$\frac{20}{30}$

66%

$\frac{33}{50}$



なし

術後の体調**不良**の
患者さん



5年
生存割合

50%

$\frac{10}{20}$

50%

$\frac{75}{150}$

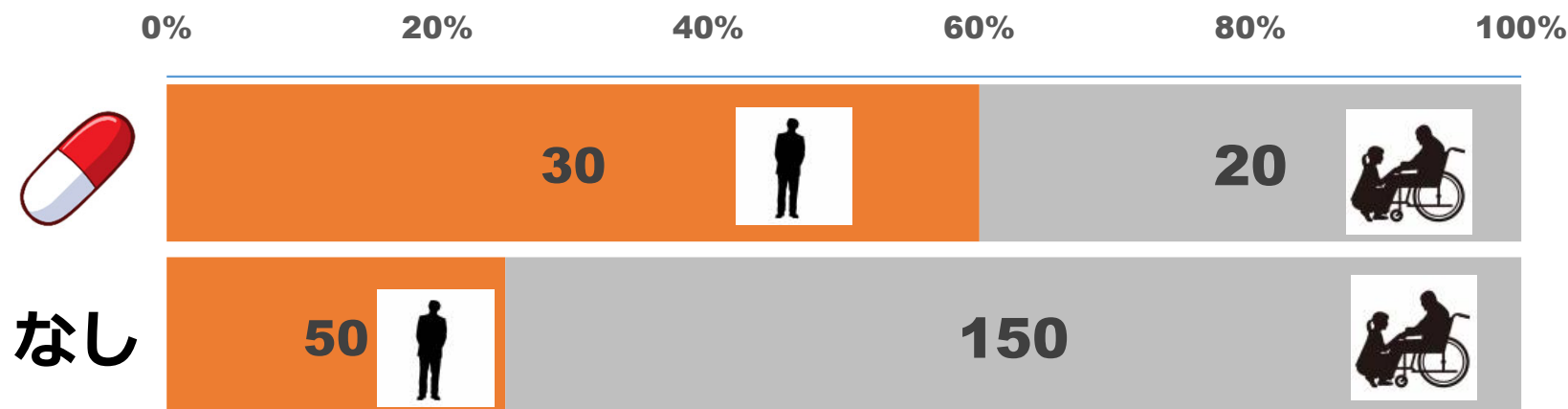


なし

全体の結果と術後の体調で分けた結果が異なるが、どちらが正しい？

妥当な比較のためには？

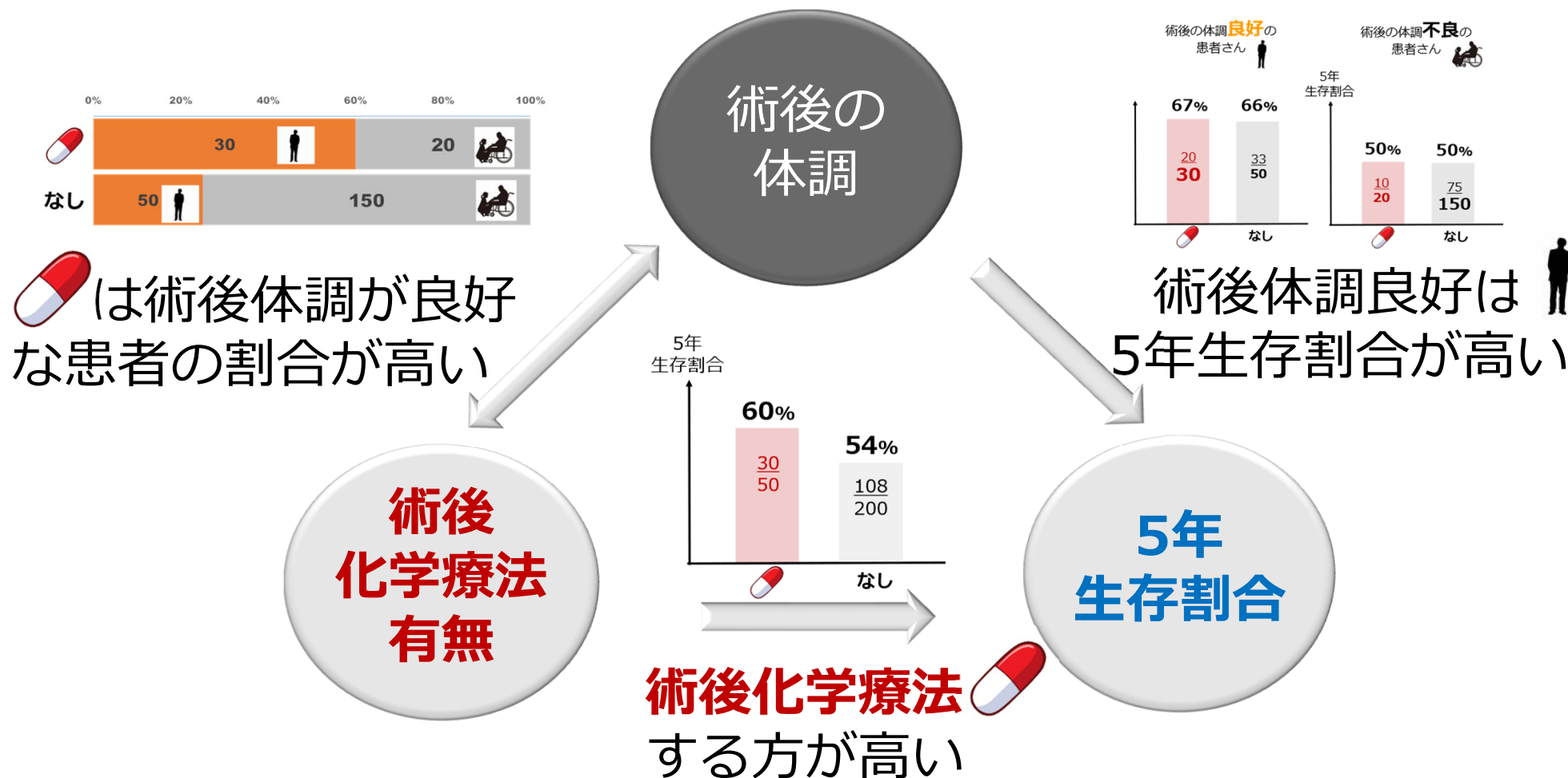
治療法以外の5年生存割合に影響する要因の条件が同じにする



-  はなしと比べ 「術後の体調良好」 の割合が高い
 -  は60%、**「なし」** は25%
- 術後の体調によって5年生存割合が異なる
 -  は67%、 は50%

交絡という現象

比較したいもの（術後化学療法有無）と研究で評価したい項目（5年生存割合）に関連する第3の因子（術後の体調）によって、見かけ上の関連が生じてしまう現象のこと



交絡がないことを保証するには

- 比較したいもの（治療）間で研究で評価したい項目（生存割合）に関係する背景因子を揃えてあげれば良い
 - 年齢やがんの進行度
 - 全身状態
 - 生存割合に起因する遺伝子その他（未知の因子も含めて）

因子がたくさんある・未知の因子があるために
全てを考慮できない



ランダムに決める

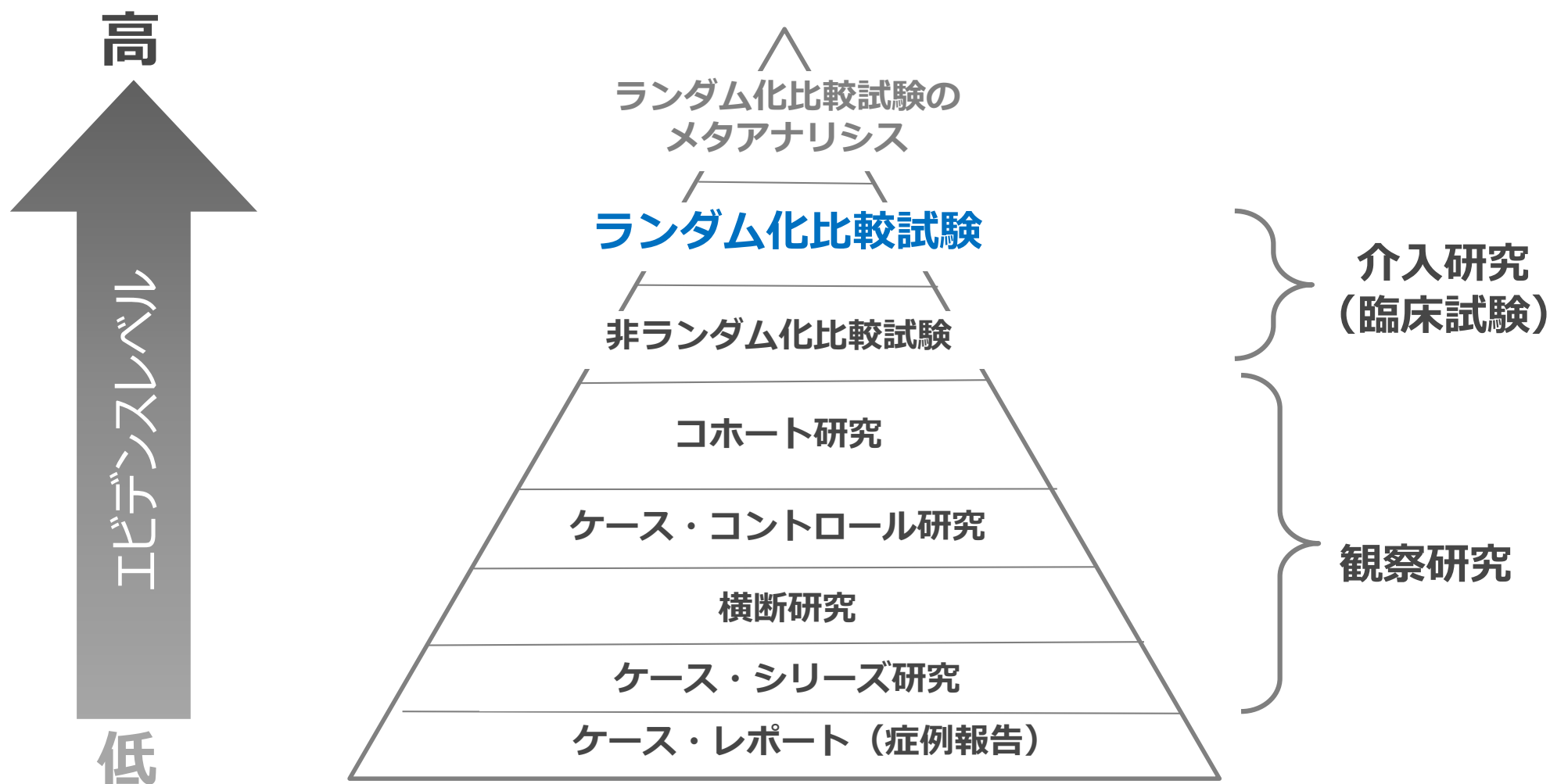
ランダム化 randomization

- 医師あるいは患者の意思によらず、確率に基づいて各治療を決める
- 予見による患者選択の偏りの防止
 - 状態の良い患者は術後化学療法を行う、などを防ぐ
- 比較可能性（内的妥当性）が担保される
 - 治療法以外は等しい集団 → 効果に差があれば治療法の違い



- 多くの患者をランダム化する（ランダム化試験に登録する）ことで平均的に群間で患者背景のバランスを取ることができる
- 結果として正しい治療法の評価を行うことが出来るので、**ランダム化試験は信頼性の高い試験**と言える

伝統的な臨床研究に関するエビデンスピラミッド



ランダム化比較試験は**交絡をデザイン段階で除去できる**ため、
エビデンスレベルが高い

今日の話題

1

交絡とランダム化

2

研究目的と研究デザイン

3

統計モデルで交絡を調整するとは

4

倫理審査Q&A

臨床研究の分類

臨床研究＝人を対象とした医学系研究

観察研究(非介入研究)

介入研究＝臨床試験

治験

医師主導治験

特定臨床
研究

先進医療下の臨床試験

“努力義務
研究”

生命・医学系
指針研究

医薬品医療機器等法
(薬機法)、GCP省令
に基づき実施

臨床研究法
に基づき実施

+ 先進医療通知
に基づき実施

人を対象とする生命科学・
医学系研究に関する倫理指針
に基づき実施

観察研究が行われる状況

1. 介入が倫理的に不適切

- 喫煙と肺がん発症の関係を調べたい場合、**喫煙有無**のランダム化は非倫理的

2. 介入が不可能

- 乳がん患者の中で**閉経有無**や**バイオマーカー有無**と再発の関係

3. 探索的，仮説生成的な研究

- **日常的なアスピリン使用の有無**と大腸がん発症の関係を調べる
- 診療データベースから大腸がんの**予後に関わる因子**を探索する

4. より現実的な状況での介入のリスク-ベネフィット評価

- **長期的な影響**や**稀なアウトカム**の調査
- **通常の臨床試験では適格外**の高齢者や併存疾患のある患者に対する影響

臨床研究を分類する3つの観点

1. 実験的 (experimental) か観察的 (observational) か

- 実験的：被験者に対して介入するのが実験的 = 介入研究 = 臨床試験
- 観察的：介入しない研究は観察研究

2. 前向き (prospective) か後向き (retrospective) か

- 前向き：研究開始時点からデータ収集する
- 後向き：診療記録など現存する情報からデータ収集する

3. 縦断的 (longitudinal) か横断的 (cross sectional) か

- 縦断的：介入/処理に関連した時間的な調査をする
- 横断的：ある特定の時点で調査対象を観察する

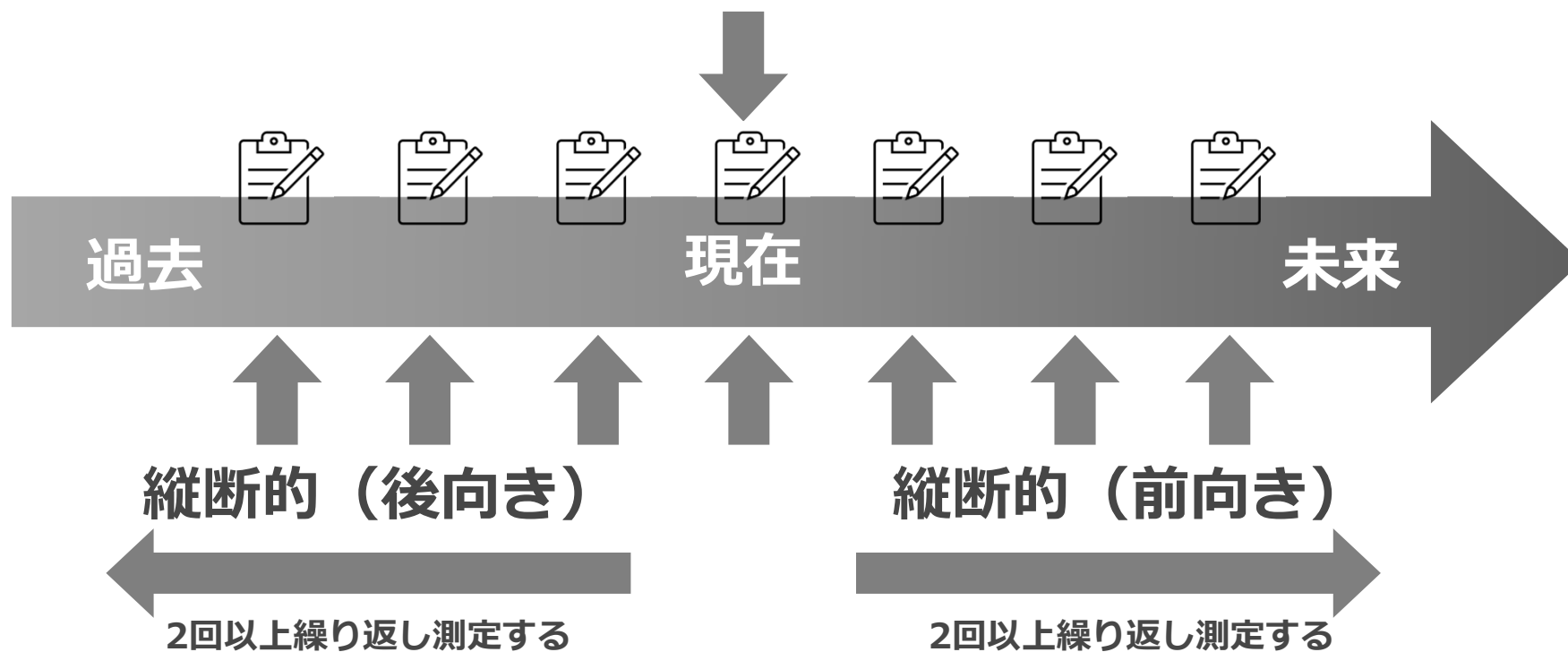
図による観察研究の分類

横断研究

ある特定の時点で一度だけ調査する

横断的

(前向き、後向き)



- ・ ケースコントロール研究
- ・ (後向きコホート研究)

前向きコホート研究

※注：前向きと後向きの定義は期間の時間的順序を意味する場合と、原因と結果の記録の順序を意味する場合がある。ここでは期間の時間的順序の意味

横断研究

- ある時点で一度だけ調査する研究
 - 例：2024年7月時点のCOVID-19の有病率の調査
- 調査したい集団からランダムに抽出される
 - 緑茶摂取と乳がん罹患との関連を調査したい場合

仮想データ	乳がんあり	乳がんなし	合計
緑茶摂取あり	5人	395人	400人
緑茶摂取なし	10人	590人	600人
合計	15人	985人	1000人

- 稀なアウトカムや原因については該当例がほとんどいない可能性あり
- 因果関係（緑茶摂取→乳がん罹患を減らす）は議論できない場合が多い
 - 乳がん罹患してから、緑茶摂取し始めたかもしれない

ケース・コントロール研究

- 現在の「結果」から過去の「要因」に遡ってデータを収集し、解析
 - 緑茶摂取状況が乳がん罹患有無で異なるかどうかを調査したい場合
 - 乳がんあり/なしの人を特定し、その人の過去の緑茶摂取状況を調べる

仮想データ	乳がんあり (ケース)	乳がんなし (コントロール)	
緑茶摂取あり	50人	100人	
緑茶摂取なし	150人	700人	
合計	200人	800人	

- 稀なアウトカムでも「あり」の人を集めやすく実施が比較的容易
- 時間・費用もそれほどかからない
- 過去の「要因」の特定にバイアスが生じやすい（後述）

ケース・コントロールの弱点

- **理想的なコントロールの選択が困難：選択バイアスの問題**
 - 研究対象集団を代表していて、疾患発症有無とその原因以外はケース集団と似ていることが理想的
 - 乳がん罹患者を国立がん研究センター中央病院（NCCH）を受診した患者とした場合、罹患していない方はどうやって選択する？
 - NCCHで別の癌で受診した方？
 - NCCHにがん健診に来ている方？
- **様々なバイアスへの対処が困難：情報バイアスや誤分類の問題**
 - 過去の緑茶摂取状況は正確に調査可能か？
 - 乳がん罹患者は真剣に思い出し、非罹患者は真面目に回答しない
 - 「緑茶摂取あり」をどう定義するか？
 - 例えば「コップ1杯・週4日以上を3か月間」と決めて、正確に分類できているか？

後向きのデータ収集や定義はバイアスが入りやすい

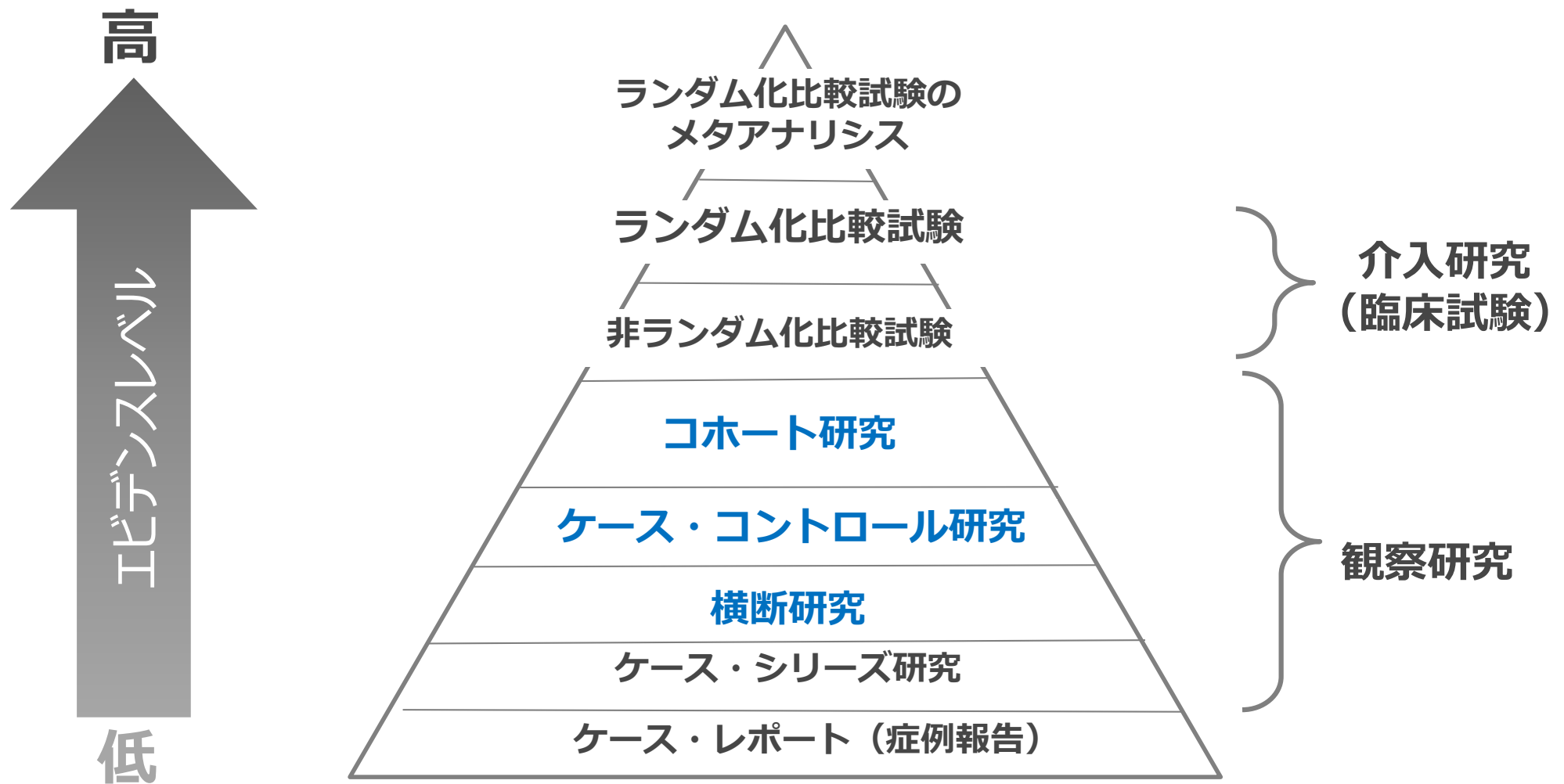
コホート研究

- 現在の「要因」有無の人を特定し、「結果」がどうなるか追跡
 - 乳がん罹患有無は緑茶摂取有無の2群間で異なるかを調査したい場合
 - 乳がんあり/なしの人を特定し、その人の過去の緑茶摂取状況を調べる

仮想データ	乳がんあり	乳がんなし	合計
緑茶摂取あり	10人	390人	400人
緑茶摂取なし	15人	585人	600人

- 要因→結果の順に観察できるため因果の順番と同じであり、複数の「要因（緑茶以外）」・複数の種類の疾患（乳がん以外）の検討が同時に可能
- 時間も費用もかかる
- 特に稀な疾患については膨大な対象者数が必要となり適さない

再掲：伝統的な臨床研究に関するエビデンスピラミッド



- 一般的に、「縦断的」・「前向き」・「介入あり」の方がエビデンスレベルが高い
 - バイアスが生じにくく、因果関係を適切に評価可能
- ただし、上に行くほど時間やコストもかかる

※注：現在は単純にこのように分類は出来ないとされているが、基本的な考え方として提示している

今日の話題

1

交絡とランダム化

2

研究目的と研究デザイン

3

統計モデルで交絡を調整するとは

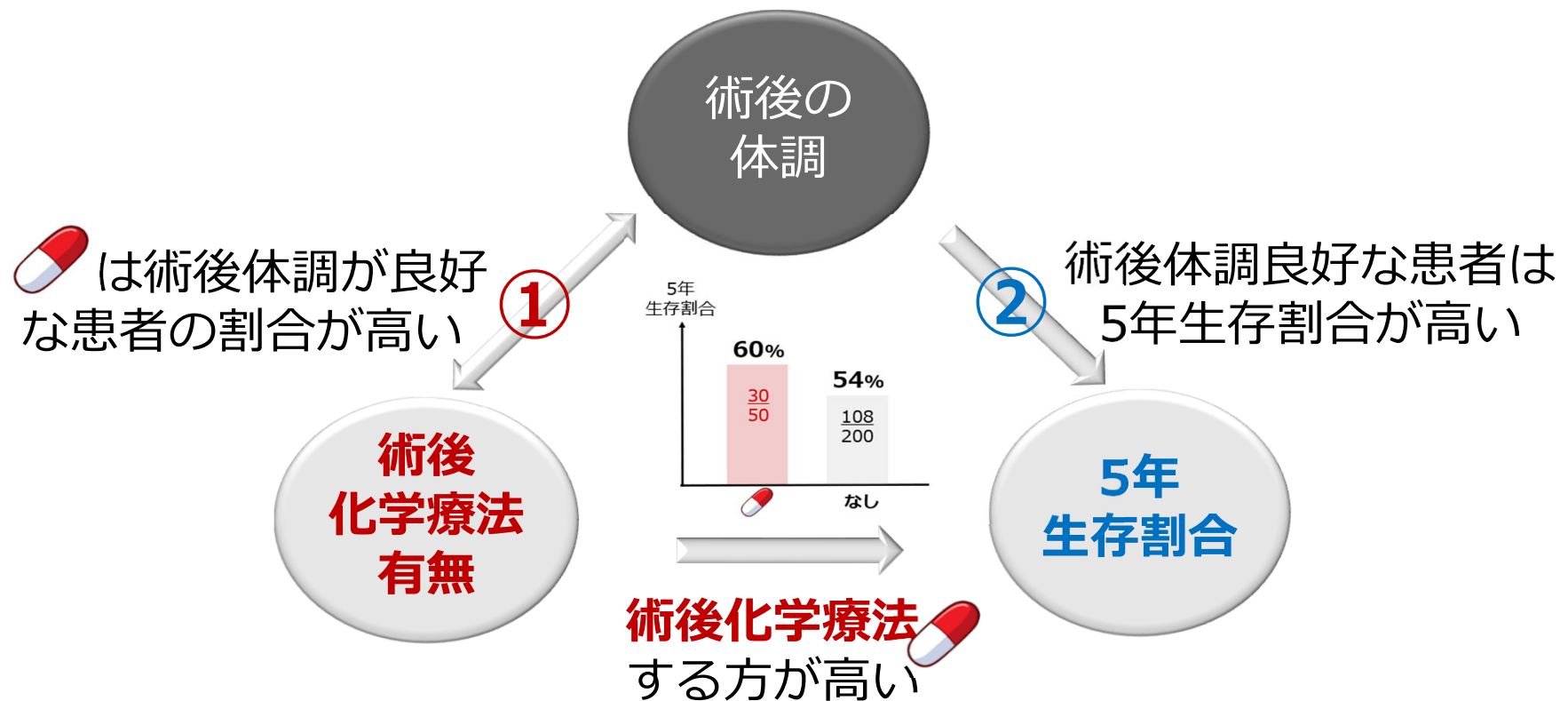
4

倫理審査Q&A

交絡を除去するには？

- 少なくともどちらか一方の関係を除去すれば良い

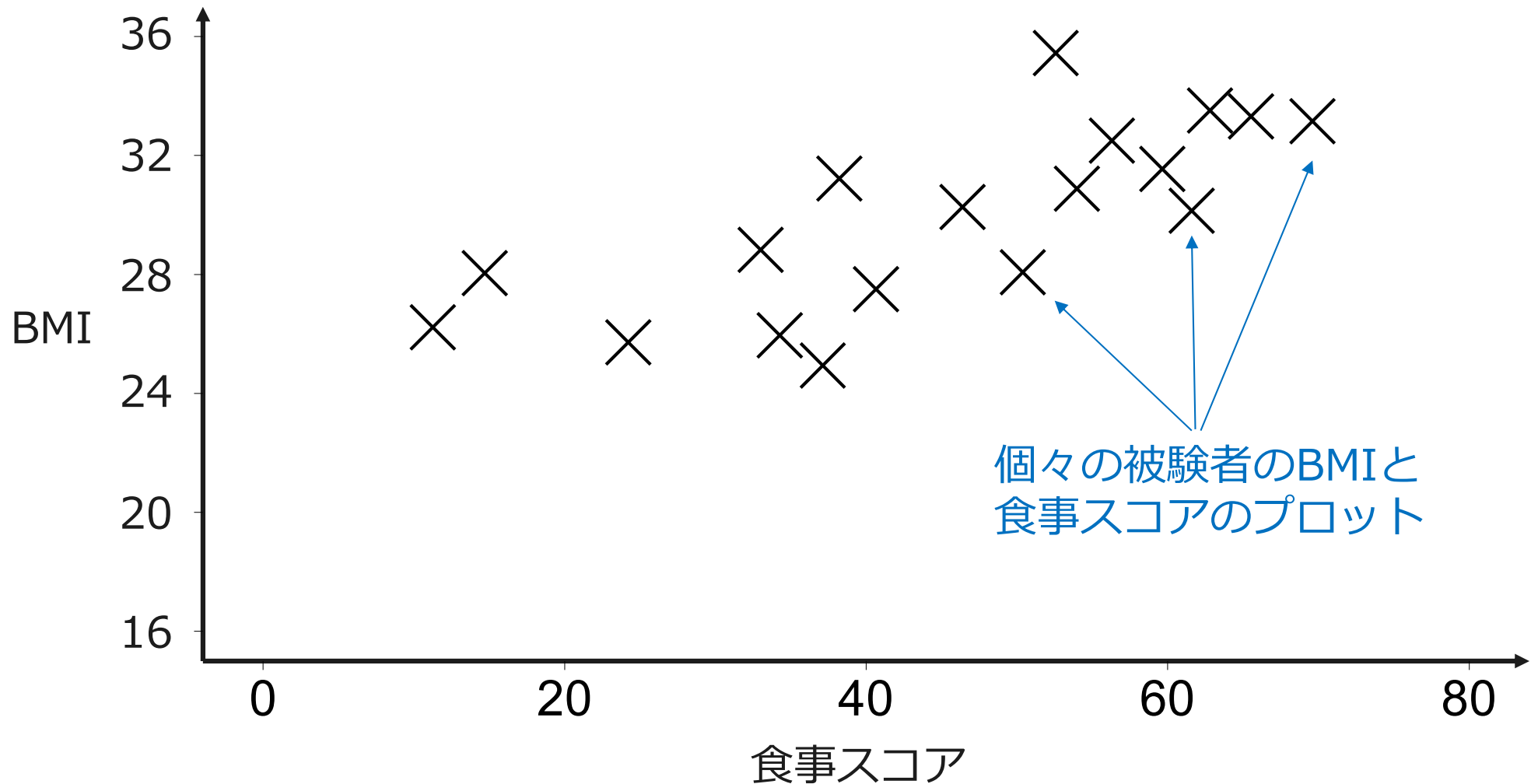
- ① **治療法**と**交絡因子**の関係 ⇨ 背景因子を治療群間で揃える
→ **ランダム化**を用いる
- ② **5年生存割合**と**交絡因子**の関係 → **統計モデル**を用いる



ランダム化が出来ない観察研究の多くは**統計モデル**による交絡調整が行われる

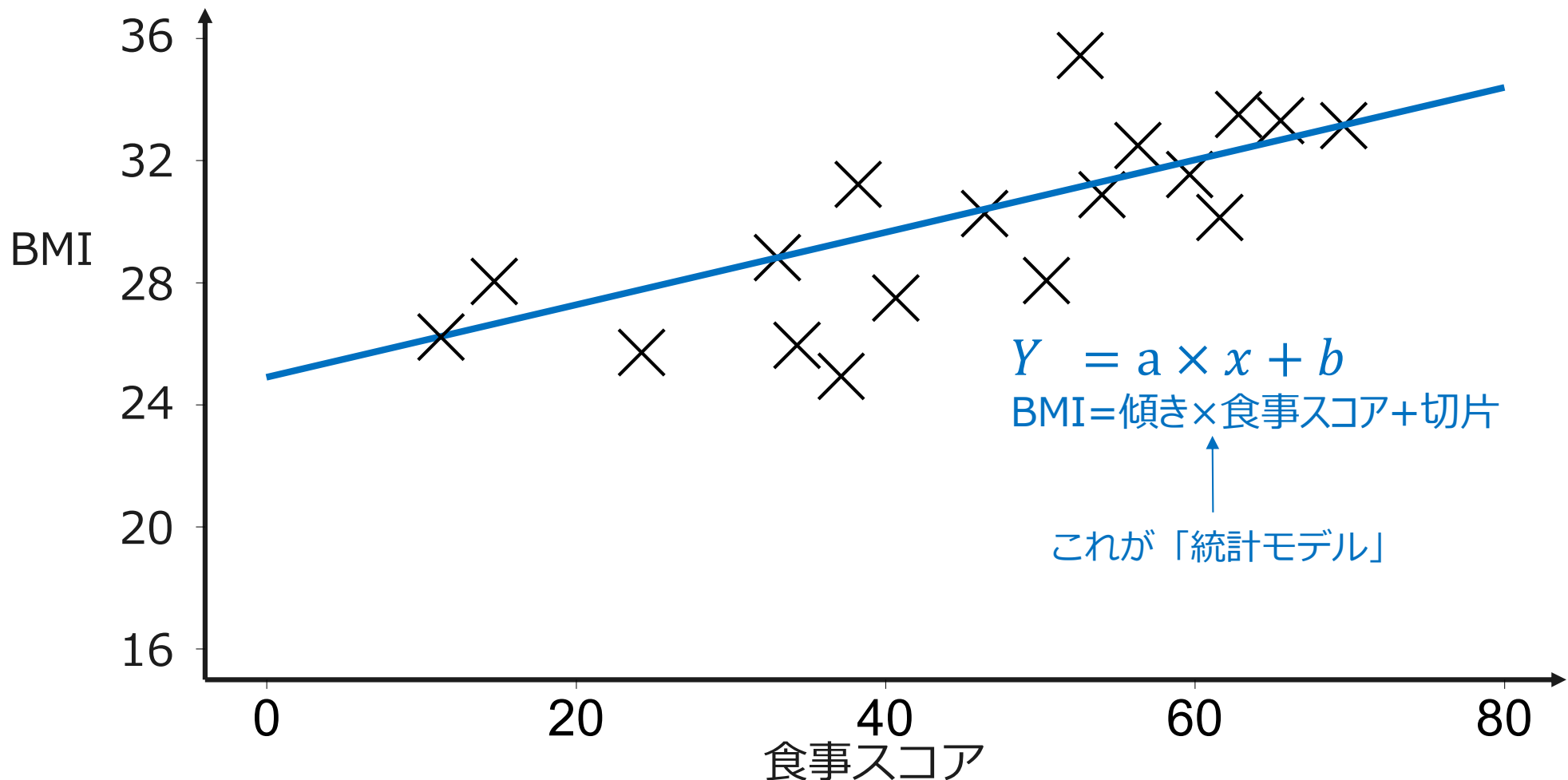
統計モデルを用いた方法

- 研究者が何をどれだけ摂取したかに基づく食事スコアを開発した
- 食事スコアによってBMI（肥満度）が予測できるかを検討したい



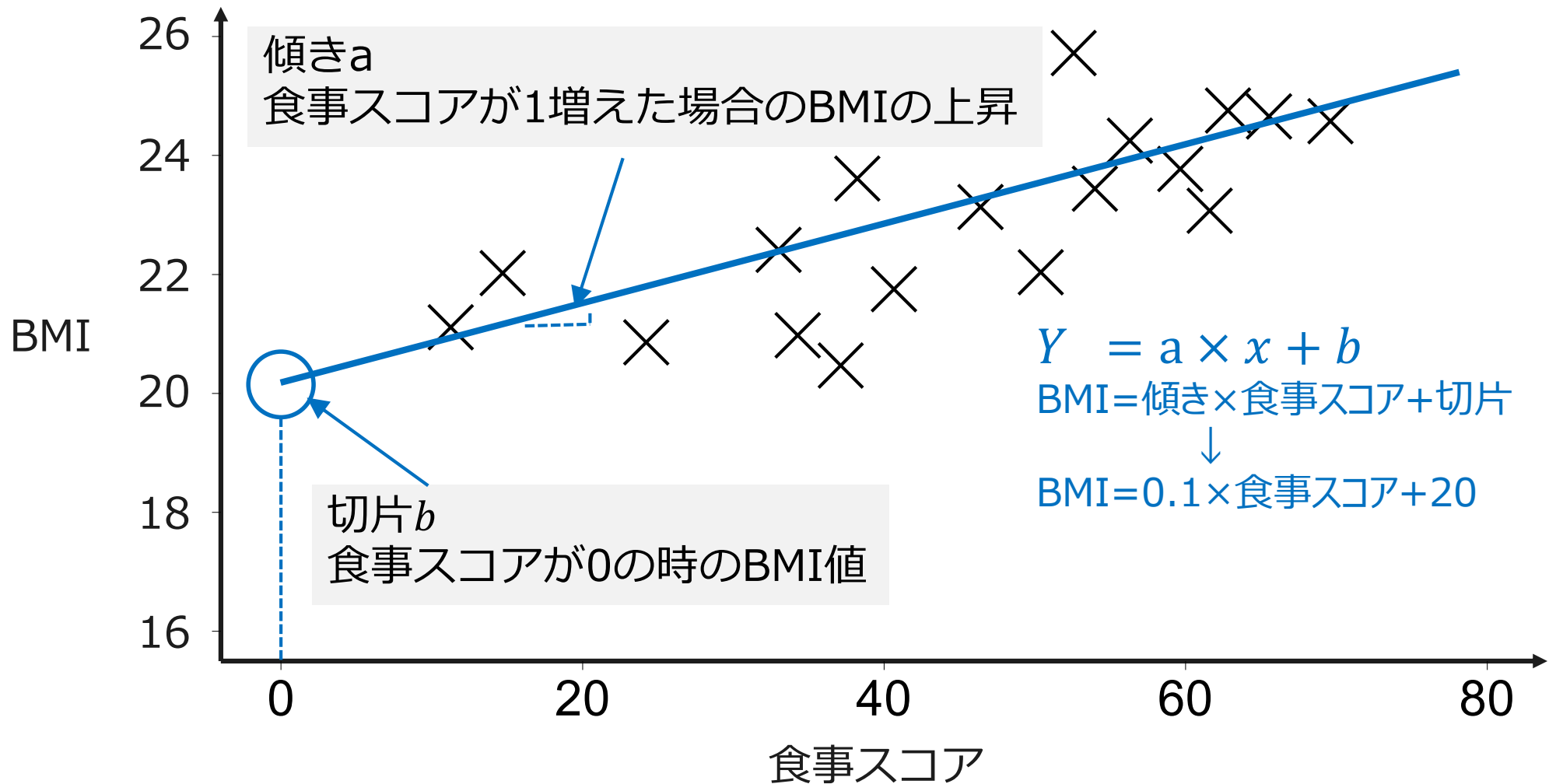
統計モデルを用いた方法-単変量回帰-

- 食事スコア(x)が増えればBMI(Y)が増える関係がありそうなので、**単純な直線の式**で表してみる
- このプロットに**尤も適切と思われる直線**を当てはめて、その時の**切片**と**傾き**を推定する
 - この条件に当てはまる計算式が決まっているのだが、今回は省略



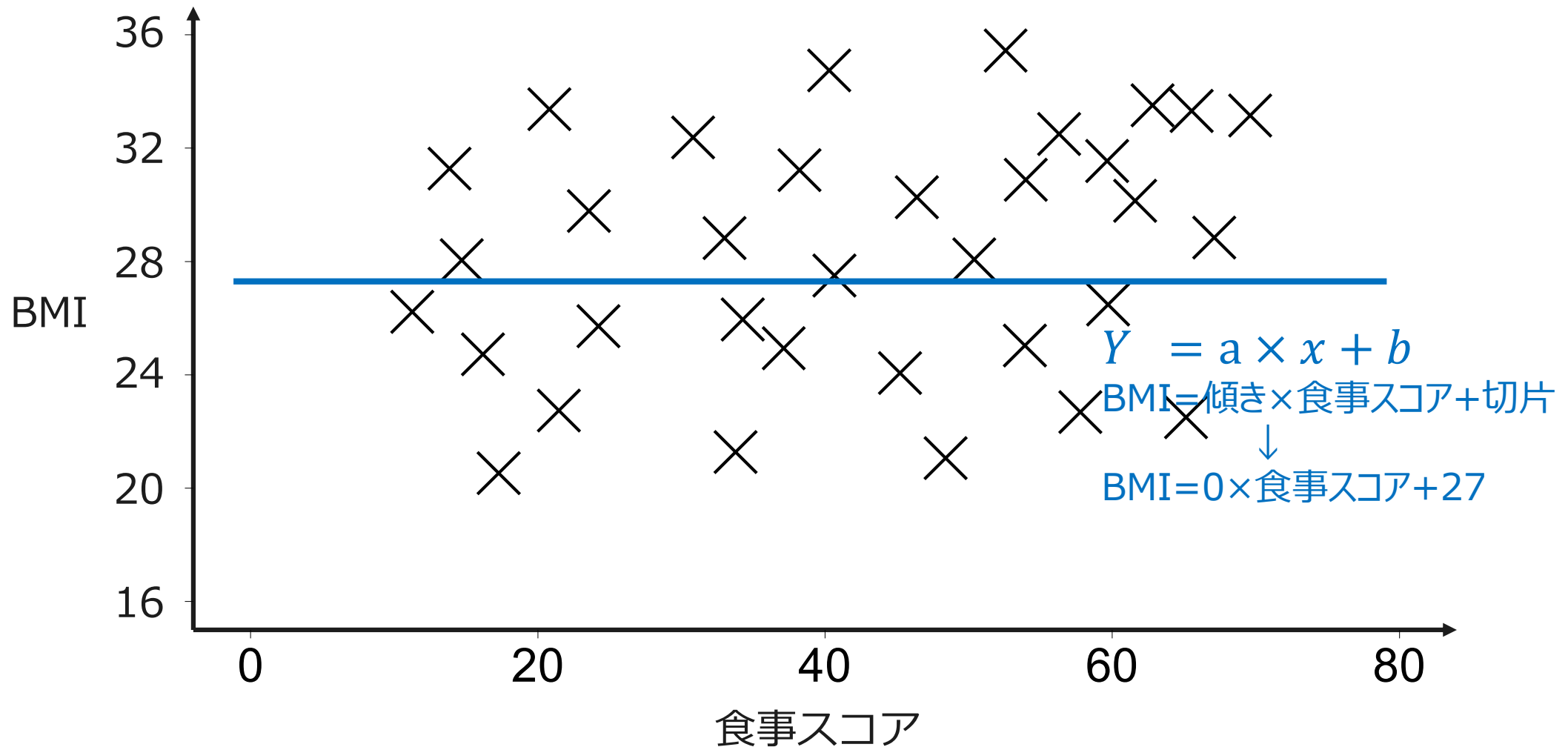
統計モデルを用いた方法-単変量回帰-

- 実際に今回のデータに統計モデルを当てはめたら $BMI = 0.1 \times \text{食事スコア} + 20$ だった
- 食事スコアが10なら、 $BMI = 0.1 \times 10 + 20 = 22$ となる



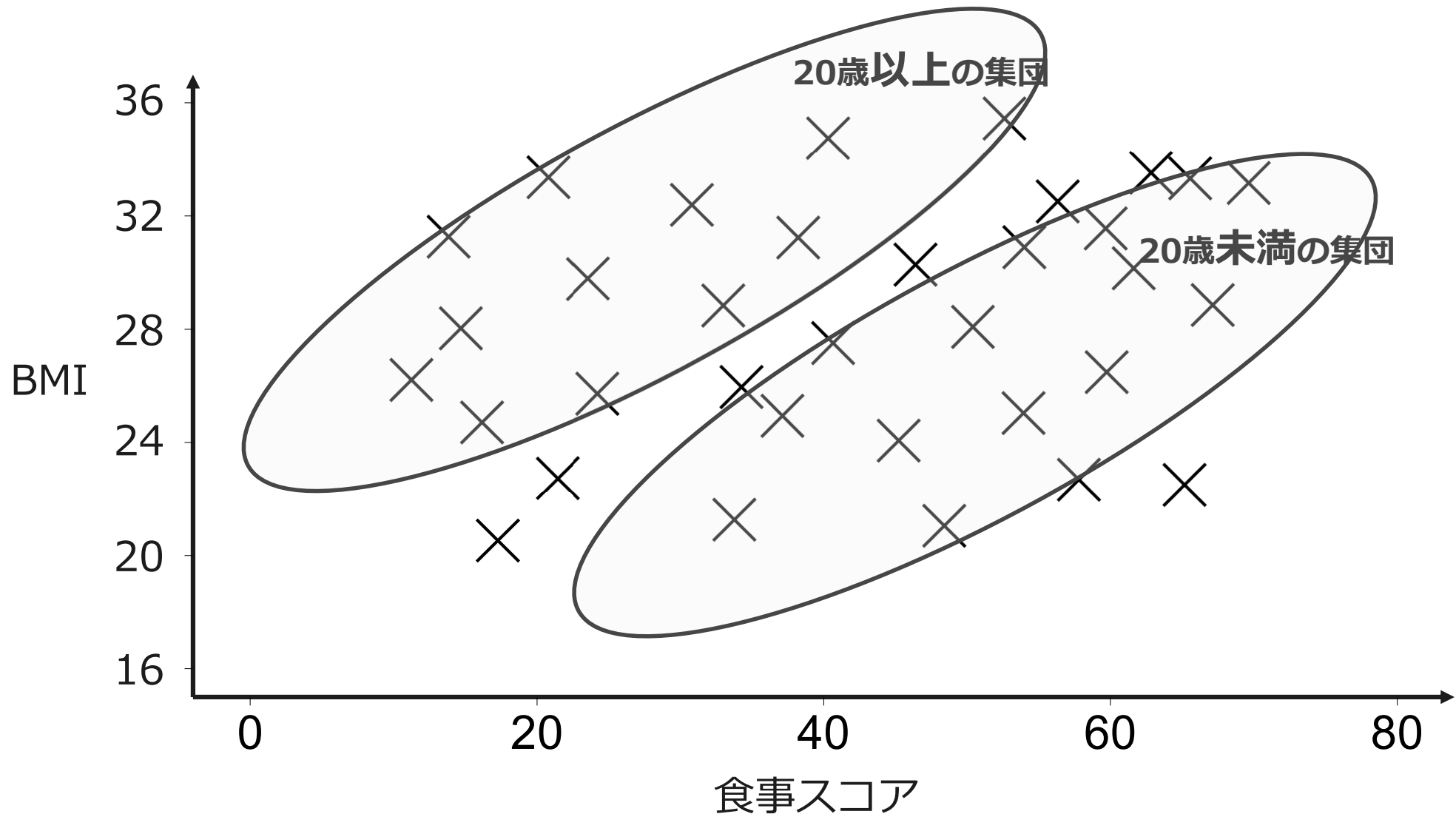
もっとデータを増やしてみた

- 先ほどのプロットと異なり、食事スコアとBMIには関連がなさそう
- このまま単変量解析をすると食事スコアの係数 (a) が0 (=関連なし) になる



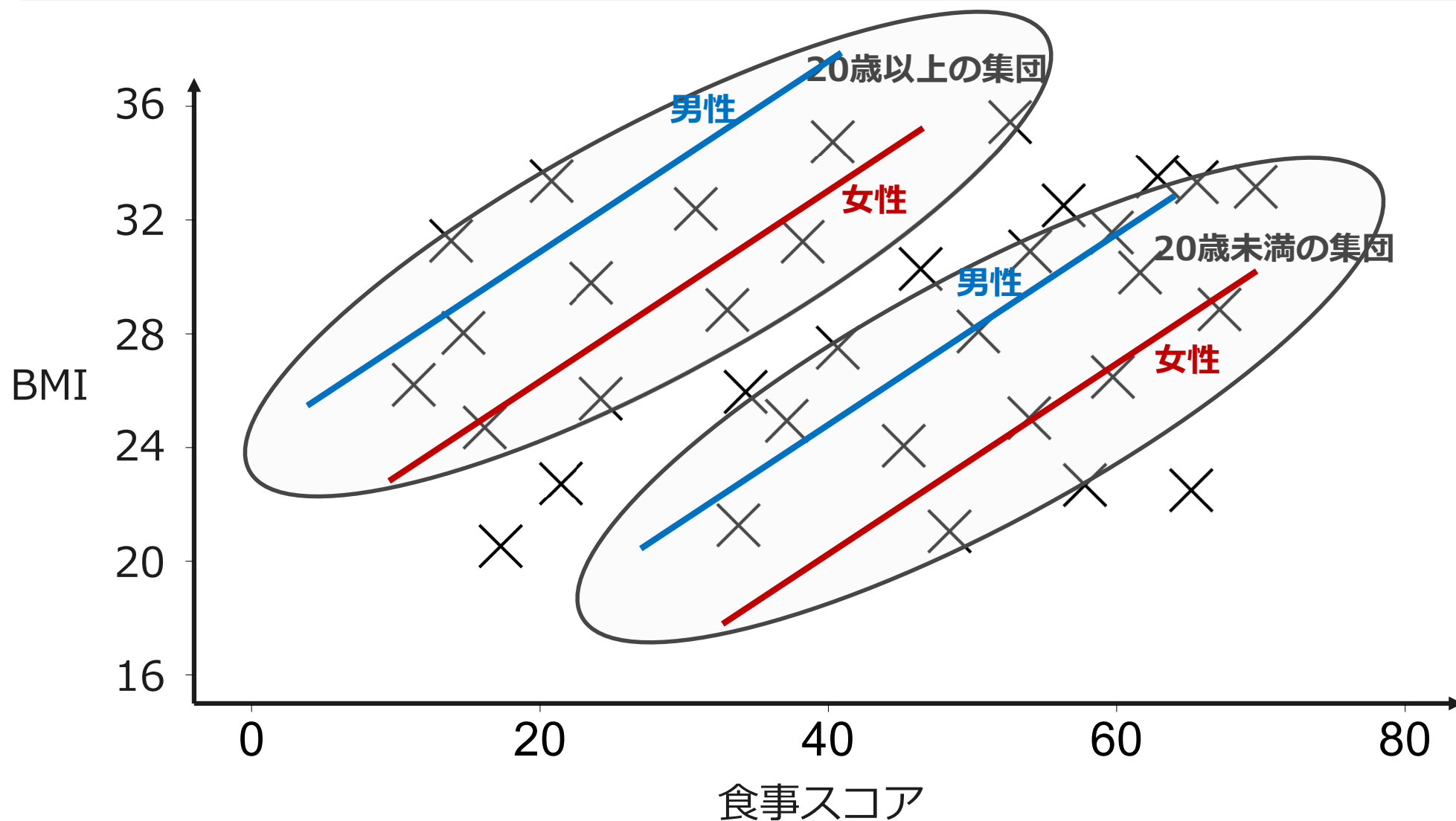
もっとデータを増やしてみた

よく調べてみると、年齢によって集団が分かれていることがわかった



さらに性別でも違いがありそうなことがわかった

「食事スコア」と「BMI」に関連する「年齢」と「性別」により「交絡」が生じている



統計モデルを用いた方法-多変量回帰-

統計モデルを用いて「年齢」と「性別」の影響を調整する

単変量解析：xは食事スコアのみ

$$Y = a \times x + b$$

BMI=傾き×食事スコア+切片



$$\text{BMI} = 0.1 \times \text{食事スコア} + 20$$

多変量解析：xは食事スコアに加えて、年齢、性別も加える

$$Y = a1 \times x1 + a2 \times x2 + a3 \times x3 + b$$

BMI=傾き1×食事スコア+傾き2×年齢（20歳以上=1、20歳未満=0）+傾き3×性別（男性=1、女性=0）+ 切片



$$\text{BMI} = 0.07 \times \text{食事スコア} + 0.1 \times \text{年齢} + 0.3 \times \text{性別} + 25$$

- 食事スコア、年齢、性別が独立なので別々に検討できる
- 食事スコア20、20歳以上、男性の時は $\text{BMI} = 0.07 \times 20 + 0.1 \times 1 + 0.3 \times 1 + 25 = 26.8$
- プロトコルに「多変量解析」と記載があれば、このような事をして交絡の調整が計画されている

今日の話題

1

交絡とランダム化

2

研究目的と研究デザイン

3

統計モデルで交絡を調整するとは

4

倫理審査Q&A

Q1. 予定対象者数の記載の確認項目は？

Q. ランダム化比較試験などの介入研究では予定対象者数の設定根拠が重要なのは理解していますが、観察研究でも同様でしょうか。電子カルテなどの情報を調べる研究では、根拠が記載されていることは少ないです



A. 原則としては観察研究でも必要ではあるがケースバイケース

- 被験者数が少なすぎると**研究から得られるものが何もない**ということが生じてしまう
 - 最低限の目的を達成しないと、研究者も無駄に労力と時間をかけることになり、貴重なデータを提供したことになる被験者に対しても説明ができない
- **なぜ予定している被験者数で十分と考えるのか**記載があると良い
- 一方でデータベース研究で因果効果を評価する研究であれば、サンプルサイズ計算に時間をかけずに、さっさとやった方が良いという意見もある

観察研究の予定対象者数の記載例

- ケース・コホート研究で曝露有無とアウトカムの比較をする場合

先行研究より、○○群と▲▲群の××発生割合をそれぞれX%、Y%と仮定する。
有意水準両側 $\alpha=5\%$ 、検出力=80%とすると、必要被験者数は■ ■となる。

通常は仮説検定を行うので、ランダム化比較試験などの同様の記載がなされる

- アウトカム（有害事象や予後）のリスク因子を探索する研究の場合

当院で○○療法を受けた▲▲癌の患者は約年間xx例である。20yy/mm～20yy/mmで合計■ ■例を対象する。なお ■ ■例であれば、見込みのリスク比=x.xxを有意水準5%で解析した場合、検出力80%以上が期待できる。

通常はCox回帰やロジスティック回帰という多変量解析の手法が用いられる。多変量解析に対する予定対象者数の厳密な手法は大変難解で多くの仮定が必要のため、厳密な計算根拠を求められることは考えにくい。ただし、データを集めても意味がないほど無謀な計画ではないことを示すため、予定対象者数で見込みの効果（ハザード比、リスク比、オッズ比等）に対して一定程度の検出力（例えば80%以上）が担保されているかの記載があると良い

観察研究の予定対象者数の記載例

- 記述的研究としてアウトカムの発症頻度を調べる場合

〇〇の発症頻度が◆◆%程度と仮定すると、95%信頼区間の幅が±▲▲以内となるためには、■例の登録が必要である

ある程度しっかりと調査を行う場合、対象とするアウトカム（有害事象発症割合など）を一定程度の精度で推定することが望ましいため、精度ベースの例数設計がされていると良い

- 注目する因子の実施可能性を検討する研究の場合

侵襲を伴う〇〇についての測定が可能かどうかの研究である。データ収集により分布の形状およびバラツキを評価する。適格規準を満たす患者は月あたり◆◆例であることから、□□例の登録とする。

理想的には精度ベースでの例数設計がされていると良いが、初めて測定するバイオマーカー値など、注目する因子の分布が全くわからない場合、研究者の資源と被験者に対して倫理的に許容可能な範囲（数例～10例程度？）で実施することはあり得る。この場合は、統計的な例数設計の根拠は求めない

Q2. 解析手法は確認しないといけないのか？

Q. 観察研究の場合、データ解析を進めながら別の解析を行うこともあるようで、「統計解析手法は定めない」と記載されているものもあります。どこまで解析手法の記載を求めるべきでしょうか。



A. 一部の記述的研究を除いて観察研究でも確認必要ですが、専門的な事は統計担当に任せれば良い

- しっかりと計画された研究であれば、観察研究であっても一定レベルの統計解析手法は定まっているので、記載されていることが原則
- 解析手法が決まってないということは、**材料だけ買ってきて調理の仕方がわからないのと同じなので、データ収集しても何も得られず、手法を間違えると誤った結論を導くことになる**
- 専門的に勉強をしていないと難しいため、審査は**統計担当に任せれば良い**

基本的な統計解析手法

統計手法はアウトカムの型（連続値/2値データ/生存時間）によって、通常用いる手法がおおよそ決まっています

アウトカムの型	連続量	2値(0 / 1)	生存時間
アウトカムの例	臨床検査値、 治療後値-治療前値	腫瘍縮小効果、 有害事象発生割合	全生存期間 無増悪生存期間
データの要約法	ヒストグラム・散布図	分割表	Kaplan-Meier曲線
群間比較の 検定手法	t検定、 Wilcoxon検定	カイ2乗検定、 Fisherの 直接確率検定	ログラंक 検定
統計モデル	重回帰分析	ロジスティック回帰	Cox回帰

原則はありますが料理本のように用いるのも危険なので、専門的な事は専門家に任せれば良いと思います

予後因子/リスク因子探索を目的とした研究の統計記載例

記載例1：「XX癌に対する予後因子を明らかにする。本研究は探索的な解析であるため統計記載は行わない。」

- 探索的 = 適当でよいというわけではない。プロトコールから（願わくは第3者が）実際に解析を再現できるような記載が望ましい
- 実際にどのように解析するかが全く書かれておらず、科学的に妥当かが判断できないため、解析法を記載するように意見し、再審査を求める

記載例2：「XX癌に対する予後因子を明らかにする。全生存期間についてはCox回帰を用いて単変量解析・多変量解析による検討を行う。」

- “Cox回帰”は特定されているが、どのような因子について検討するのか記載がない
- 多変量解析も様々な手法があるが、どんな手法を用いる予定なのか
- サンプルサイズと予後の記載がない場合、解析がまともに出来るのか判断できない
- これからデータ収集するような研究であれば、再審査を求める

統計記載の例

記載例3：「全生存期間についてはCox回帰を用いて単変量解析・多変量解析（①）による検討を行う。単変量解析はプロトコールX.Xに示す10因子（②）について検討を行う。連続量の変数については中央値で2水準にする。多変量解析はプロトコールX.Xに示す全ての因子を含める（③）。また、 $\alpha=20\%$ として変数増減法（④）による検討も行う。」

① 解析法が特定されている

② ①で特定した解析法に対してどのような因子（年齢、性別、病期など）について検討するのかが特定されている。これをデータ収集する（されている）ということもわかる

③と④ ②と同様、多変量解析についても特定されている。

このくらい記載があると、少なくとも解析手法が妥当かどうか（≡科学的に妥当な研究計画か）が判断できる

まとめ

- ランダム化試験は信頼できる結果を得る最良の方法
 - 予見による患者選択の偏りの防止、交絡の除去
- デザインごとの観察研究の目的と限界
 - コホート研究、ケースコントロール研究、横断研究
 - それぞれメリット・デメリットあり
- ランダム化できない観察研究では交絡調整を行う
 - 統計モデルを用いた多変量解析という手法が用いられることが多い