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

2023.8.26 (土)

2023年度倫理審査委員会・治験審査委員会委員養成研修

多くのスライドは ICRweb [https://www.icrweb.jp]の資料を改変しています

今日の話題

1 **ランダム化が必要な理由**
研究の対象者を2つ以上のグループにランダムに分けること

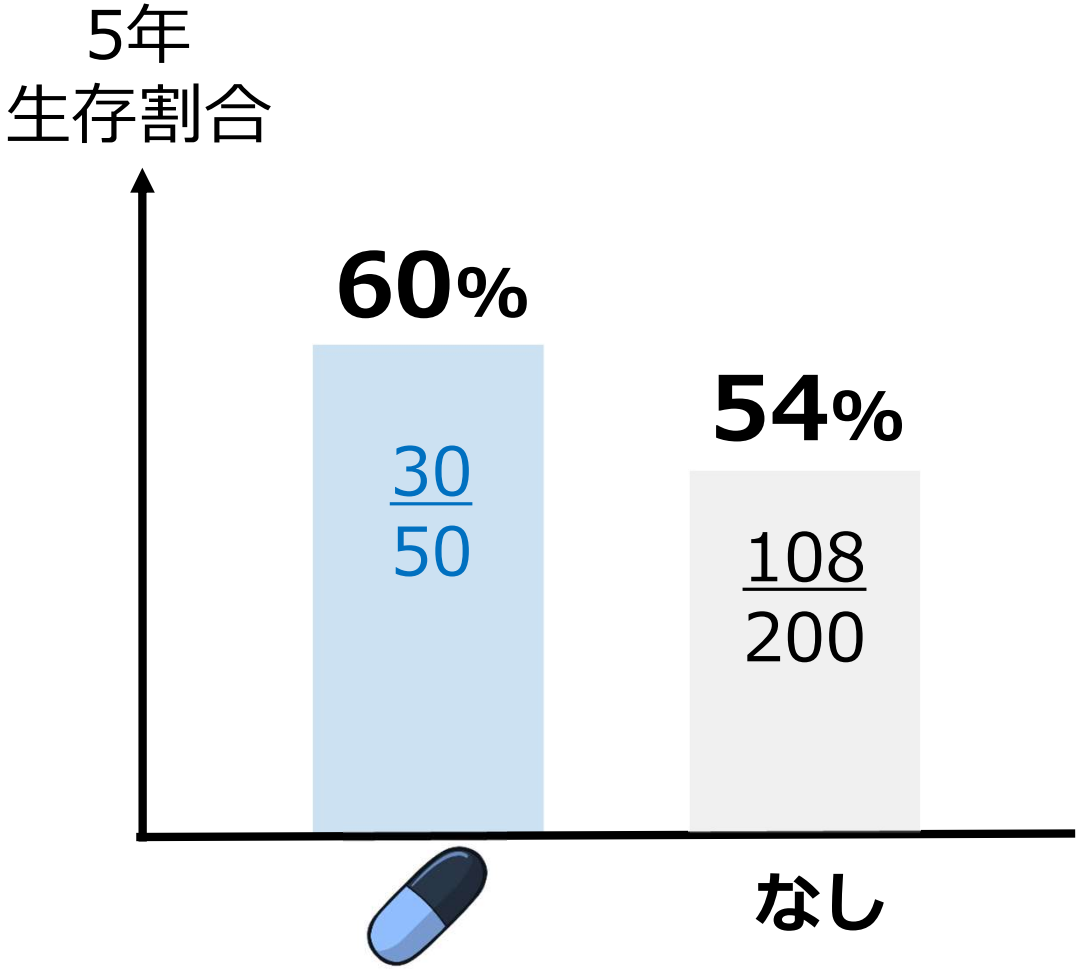
2 **臨床試験の予定登録数はどのように決まっているのか**

3 **勝たなくても良い治療って何？**

注：仮想例

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%

20/**30**

66%

33/**50**

なし

術後の体調**不良**の患者さん

5年生存割合

50%

10/**20**

50%

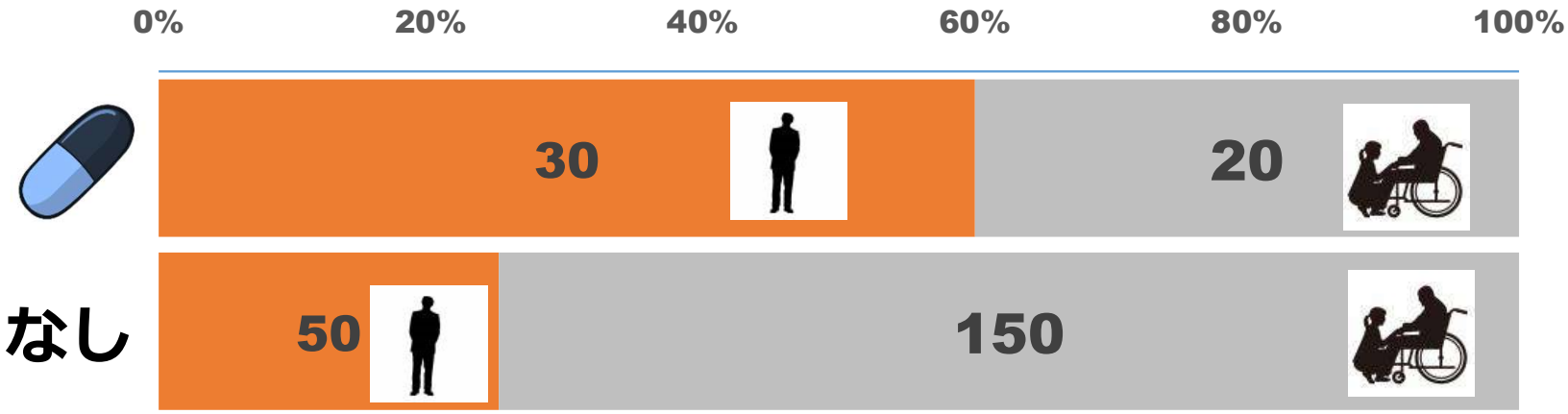
75/**150**

なし

注：仮想例

妥当な比較のために？

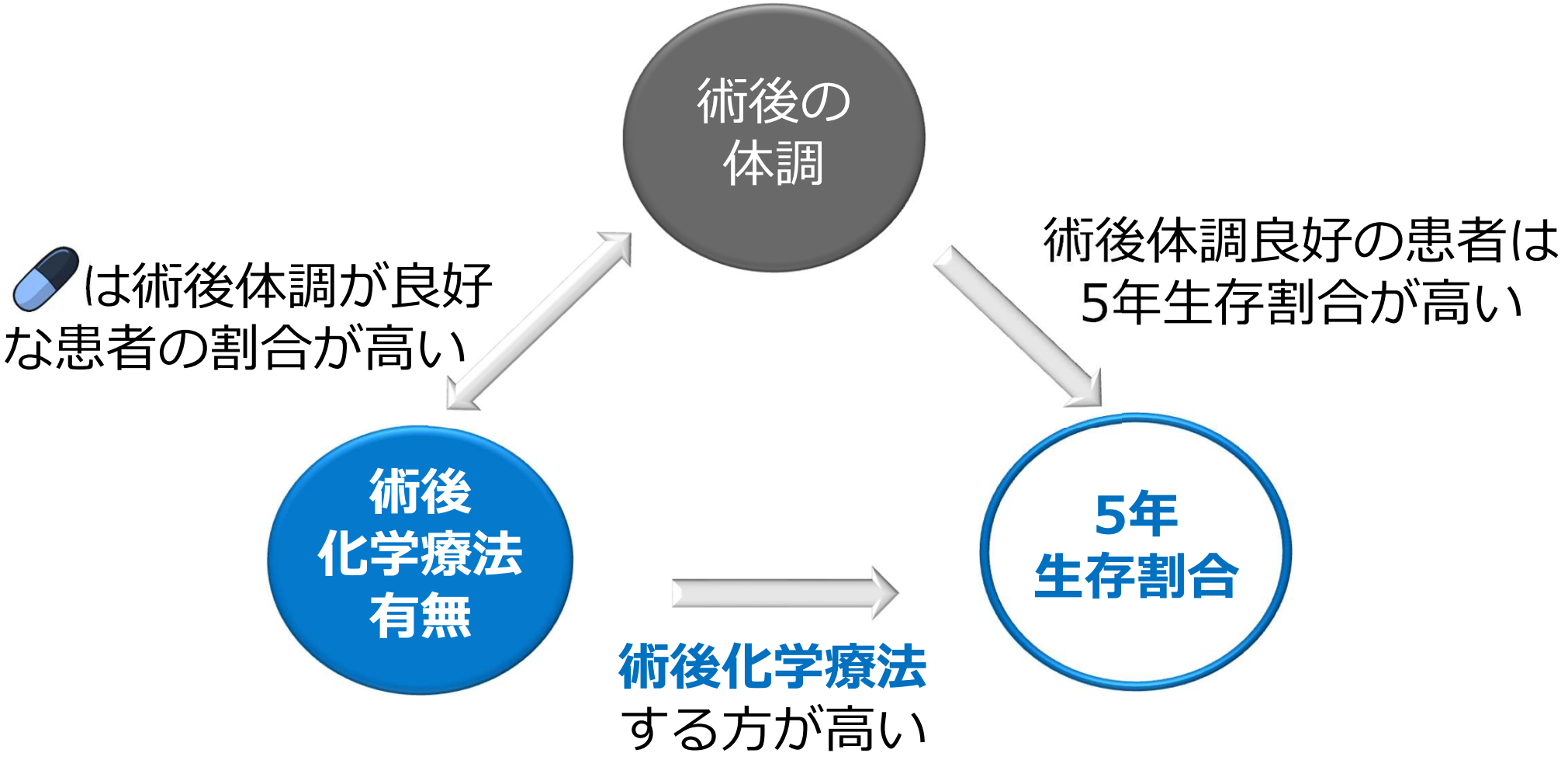
治療法以外の5年生存割合に影響する要因の条件が同じにする



-  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やその他の因子（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/ランダム割付

- 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

統計的観点からの倫理審査のポイント

- 日常臨床に還元する（標準治療を決める）ことを意図する検証的な目的の試験では「ランダム化比較試験」が王道である
- この目的の試験でランダム化が計画されていない場合、なぜランダム化されないのかが計画書に記載されている必要がある
  - 記載がない場合は審査時に確認すべき事項

今日の話題

1 **ランダム化が必要な理由**
研究の対象者を2つ以上のグループにランダムに分けること

2 **臨床試験の予定登録数はどのように決まっているのか**

3 **勝たなくても良い治療って何？**

【質問】どちらかと言えば臨床試験の患者数は、

1.多ければ多い方が良い

- 結果の信頼性↑、10例のデータ < 1000例のデータ

2.少なければ少ない方が良い

- 企業のコスト↓
- 患者さんにとっての倫理性

3. その他

お医者さんがあなたに言いました

あなたは膵臓がんと診断されました。

私が手術します。

危険な手術ですが、私は失敗したことはありません。
つまり、成功率は100%です。

私にお任せください。

100%と聞いてあなたが想像するのは？

- **300回**の経験があり、全て成功した
- **100回**の経験があり、全て成功した
- **10回**の経験があり、全て成功した
- **5回**の経験があり、全て成功した
- **1回**の経験があり、全て成功した

成功率は全部100%、ただし精度が異なる

- 手術件数（データ）が**多い**方がより**信頼**できる
  - 300例やって100%と、1例やって成功割合100%では全く信頼度が違う
- 統計学の用語では、ここでいう”100%”のことを「**点推定値**」と言う
- 「100%」に対する**「信頼度」**を表す指標を添えた方が適切に判断できる
  - この信頼度を**「信頼区間 (confidence interval)」**というもので表す

信頼区間(CI)はデータが多いほど狭くなる

- 300回の経験があり、全て成功した
  - 95% CI[98.8%-100%]
- 100回の経験があり、全て成功した
  - 95% CI[96.4%-100%]
- 10回の経験があり、全て成功した
  - 95% CI[69.2%-100%]
- 5回の経験があり、全て成功した
  - 95% CI[47.8%-100%]
- 1回の経験があり、全て成功した
  - 95% CI[2.5%-100%]

ここまでのまとめ

- 患者数が増えれば得られた結果の信頼度が上がるので、統計的な精度という意味では予定登録数は多い方が良い
  - この「信頼度」を統一的に表す指標が信頼区間
- 一方で、試験実施側の観点でも患者さんに対する倫理的な観点でも多すぎる患者数は好ましくはない

典型的な臨床試験の研究計画書予定登録数の記載

- 背景：○○癌に対するxx療法の有効性を検討する
- デザイン：単群試験
- 研究で最も重要な評価項目：3年生存割合
  - 予定登録数：45例

    片側有意水準（**α**）＝5%

    検出力（**1-β**）＝80%

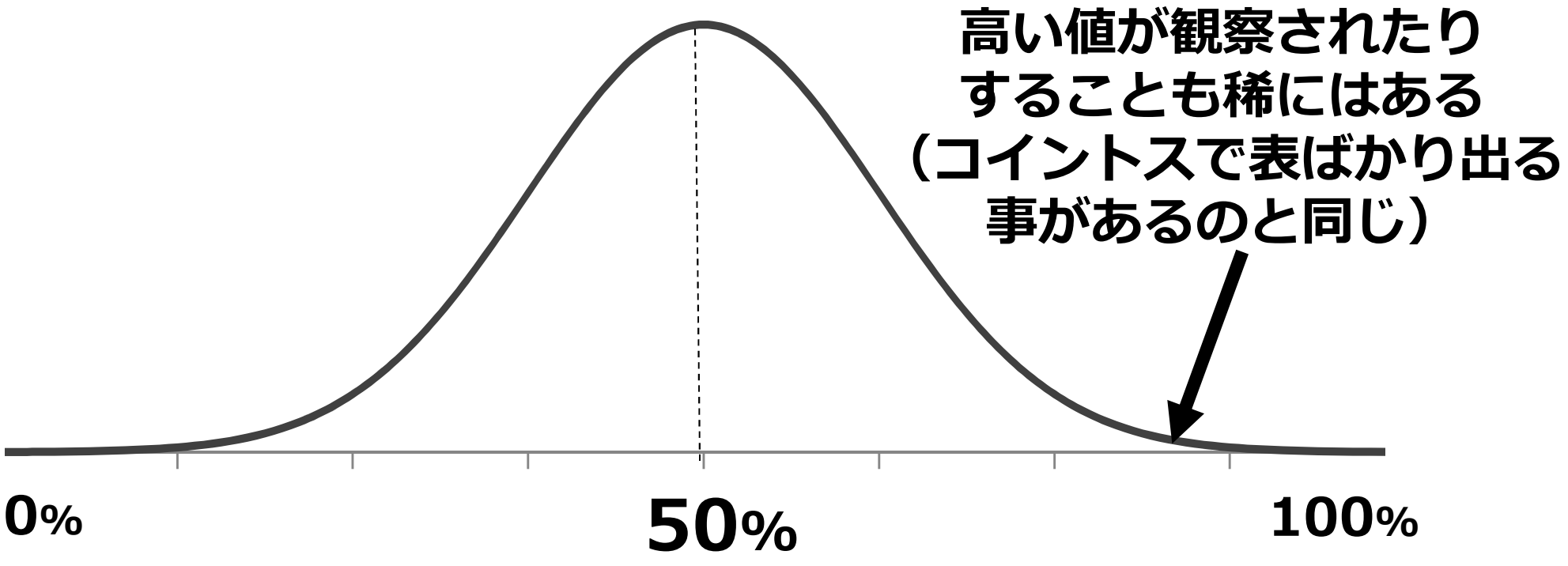
    閾値＝50%、**期待値**80%

- 閾値：試験治療は無効と判断する3年生存割合
- 期待値：試験治療に期待する3年生存割合

- 予定登録数の決定に必要なのはα、β、閾値・期待値、評価項目（エンドポイント）
- そのため登録数決定の考えを理解するにはこれらの意味を理解する必要あり

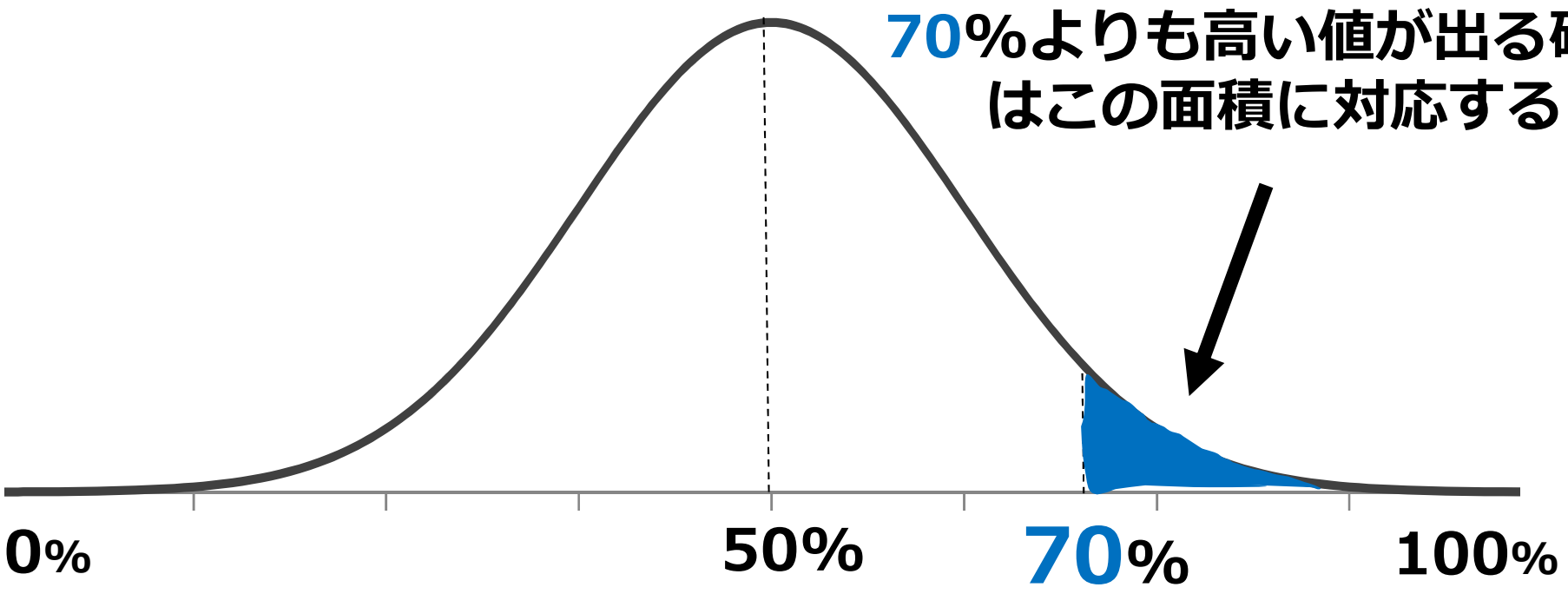
有効な治療法であることをどう判断するか？

- 神のみぞ知る真の3年生存割合は50%（=閾値）と仮定する
- 実際に臨床試験を何度も繰り返したら得られる3年生存割合は？



**真実の3年生存割合が50%なら、
50%が最も多く観察される**

実際に得られた結果が70%だったら？



-  の部分が十分小さいならば、そもそも真の3年生存割合が50%という仮定が間違っていた、と判断するのが妥当だろう
- 仮に  の面積が7%だとしたら、これは十分小さい（≒真の3年生存割合が50%が間違っている）と言えるだろうか？

質問：  は十分小さい？

1.十分小さい

- そもそも真の3年生存割合が50%という仮定がおかしいほど、稀な結果が起きた、と考えるべき

2.十分小さいとは言えない

- そもそも真の3年生存割合が50%という仮定がおかしいほど、稀な結果が起きた、とは言えない

事前に判断規準を定めておきましょう

- 結果を見てから稀かどうかを判断すると後付けになってしまうので、事前に稀かどうかの規準を決めておく
  - この判断規準のことを**有意水準(α level)**という
  -  の確率（=p値という）が有意水準を下回ったら、【有意差あり】と結論する
    - そもそも真の3年生存割合が50%という仮定がおかしい
      - 真の3年生存割合は50%以上であると判断できる

今回の例（α=0.05、p=0.07）はどのように判断できますか？

検定における2種類の誤り

		真実	
		効果なし	効果あり
検定結果	有意差なし	正しい	誤り （**βエラー**）
	有意差あり	誤り （**αエラー**）	正しい （**検出力.1-β**）

- **検定の結果が必ずしも真実を反映しているわけではない**
- **検定の誤りを小さくするには、精度を高くする（＝登録患者数を増やす）必要がある**

通常のαとβの設定値（誤りの大きさの許容範囲）

- 検証的試験（Phase 3）の場合のα
  - 片側α=2.5% / 両側α=5%がデフォルト
  - ICH E9（国際的なガイドライン）で決まっている
    （消費者が負うリスク）
- 検証的試験（Phase 3）の場合の検出力（1-β）
  - 80%以上がデフォルト（企業が負うリスク）
- 検証的試験（Phase 3）の場合、α < β がデフォルト
  - αエラーの方がβエラーよりも社会にとってリスク
  - Nを増やさない限り、αとβを同時に小さくはできない

計画書の記載の意味

- 研究で最も重要な評価項目：3年生存割合
  - 予定登録数：45例

    片側有意水準（α）＝5%

    検出力（1-β）＝80%

    閾値＝50%、期待値80%

④ ①～③の条件を満たす時、45例の患者さんを登録する必要があります

①効果がないのに効果ありと言ってしまう確率は5%以下にします。
言い換えれば、p値が5%以下なら**有意差**ありと判断します

②効果がある時に効果ありと言える確率は80%以上にします

③最低でも真の3年生存割合が50%以上の治療法でないと、良い治療法と宣言したくありません。80%を期待できると思っています

統計的観点からの倫理審査のポイント

- 多くの臨床試験は、統計的仮説検定で判断するために必要な予定登録数が設定されている
  - α、β、閾値・期待値など
- 試験の相や疾患の希少性によって相場はあるが、αやβの値が相場と異なっている場合には、その理由について確認する
- その治療が最低限上回らなければいけない値（＝閾値）については、過去データと治療のリスクなどに基づいて決まっていることが多い。試験計画書の背景記載から閾値が受け入れ可能かどうかを確認する

今日の話題

1 **ランダム化が必要な理由**
研究の対象者を2つ以上のグループにランダムに分けること

2 **臨床試験の予定登録数はどのように決まっているのか**

3 **勝たなくても良い治療って何？**

日常臨床では経過観察している場合

(標準治療は経過観察の場合)

胃がん手術後の
患者さん1,000人

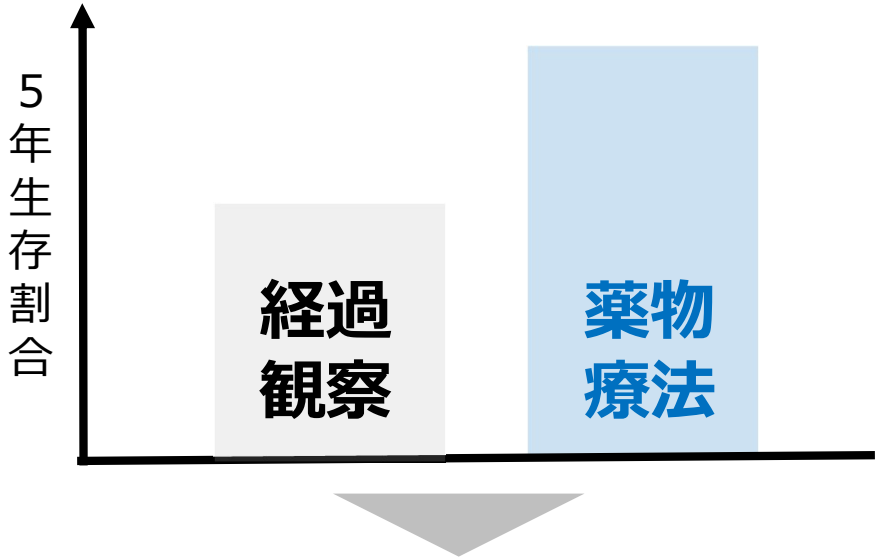
ランダム化

経過観察

薬物療法

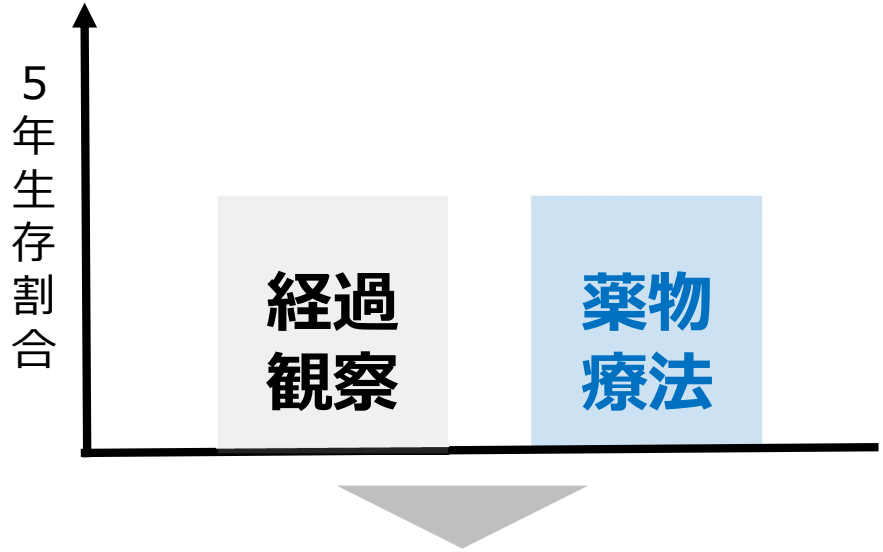
の5年生存割合を比較

薬物療法が経過観察に勝った



薬物療法が新たな標準治療

薬物療法が経過観察に勝らなかった



経過観察が標準治療のまま

日常臨床では薬物療法を施行している場合

(標準治療は薬物療法の場合)

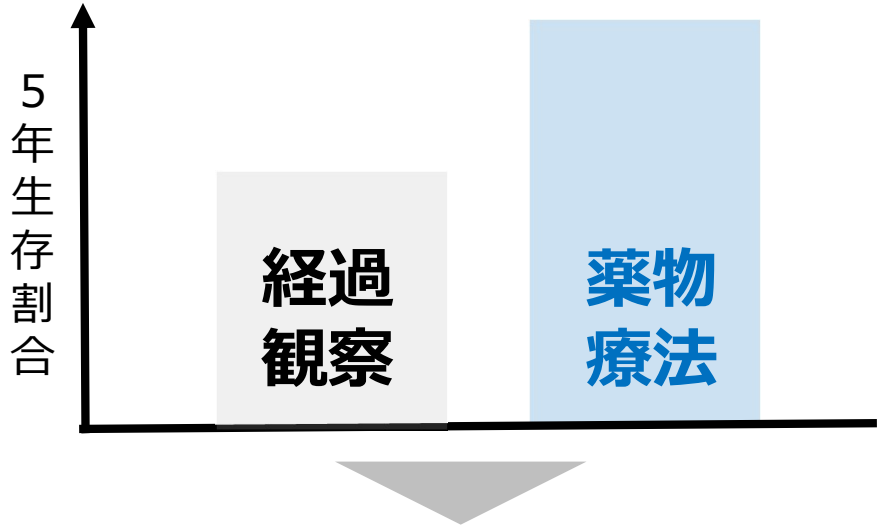
胃がん手術後の患者さん1,000人

ランダム化

- 経過観察
- 薬物療法

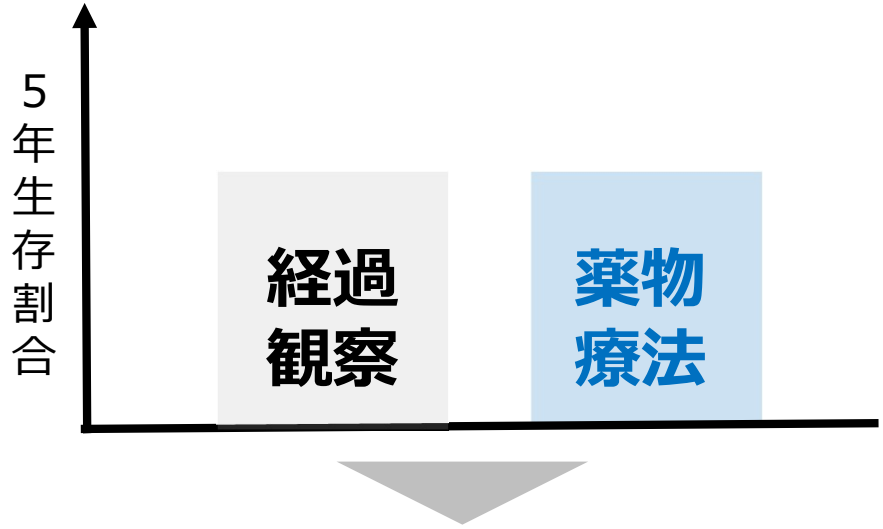
の5年生存割合を比較

経過観察が薬物療法に大きく劣った



薬物療法が標準治療のまま

経過観察が薬物療法に劣らなかった



この時、どう判断したい？

勝たなければいけない試験と
（一定以上）劣らなければいい試験

優越性試験

- 試験治療は標準治療よりも毒性が強かったり、利便性が悪い時に使う臨床試験
- 試験治療が勝つかどうかを確かめる（優越していないといけない）試験

非劣性試験

- 試験治療は標準治療よりも毒性が軽かったり、利便性が良い時に使う臨床試験
- 試験治療が一定以上劣っていない（劣性では非ず）ことを確かめる試験
- “一定以上”の規準は後出しにならないように試験開始前に決めておく

統計的観点からの倫理審査のポイント

- ランダム化試験か否かによらず、その治療が標準治療群に対して優らなければならないかを試験計画書から読み取り、本来、優越性試験なのか非劣性試験なのかを判断する
  - 自分の判断と異なる場合には確認する
- 優越性では「このくらい上回らなければならない」値、非劣性では「これ以上劣ってはいけない」値が予定登録数設定根拠に記載してある。これが試験計画書の背景記載から受け入れ可能か/理解できるかどうかを確認する
  - 非劣性の場合、リスク/ベネフィットバランスによる臨床的判断で決められることも多い

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
- 臨床試験に必要な患者数は多すぎず、少なすぎず
  - 試験の目的に合わせて有意水準、検出力、治療効果の大きさを決め、数式に基づいて計算
- 勝たなくても良い治療法の臨床試験 ＝ 非劣性試験
  - 試験治療の効果が一定以上劣っていないことを確かめる試験

用語についての補足スライド

- 最も重要な評価項目：主要評価項目、primary endpointという
  - これ以外の評価項目のことは副次評価項目、secondary endpointという。
  - 最も重要な評価項目に基づいて予定登録数や治療法が良いかどうかの判断をする
- 有意差あり：p値が有意水準αよりも小さくなること
  - ここで言っているのは統計的な有意差（**意**味の**有**る**差**）の有無のこと
  - リスク（毒性）や利便性などに見合った意味の有る差の有無は別
  - 臨床試験で有意差があると positive、ないとnegativeと言ったりする