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

2021.1.22 (土)

2021年度倫理審査委員会・治験審査委員会委員養成研修

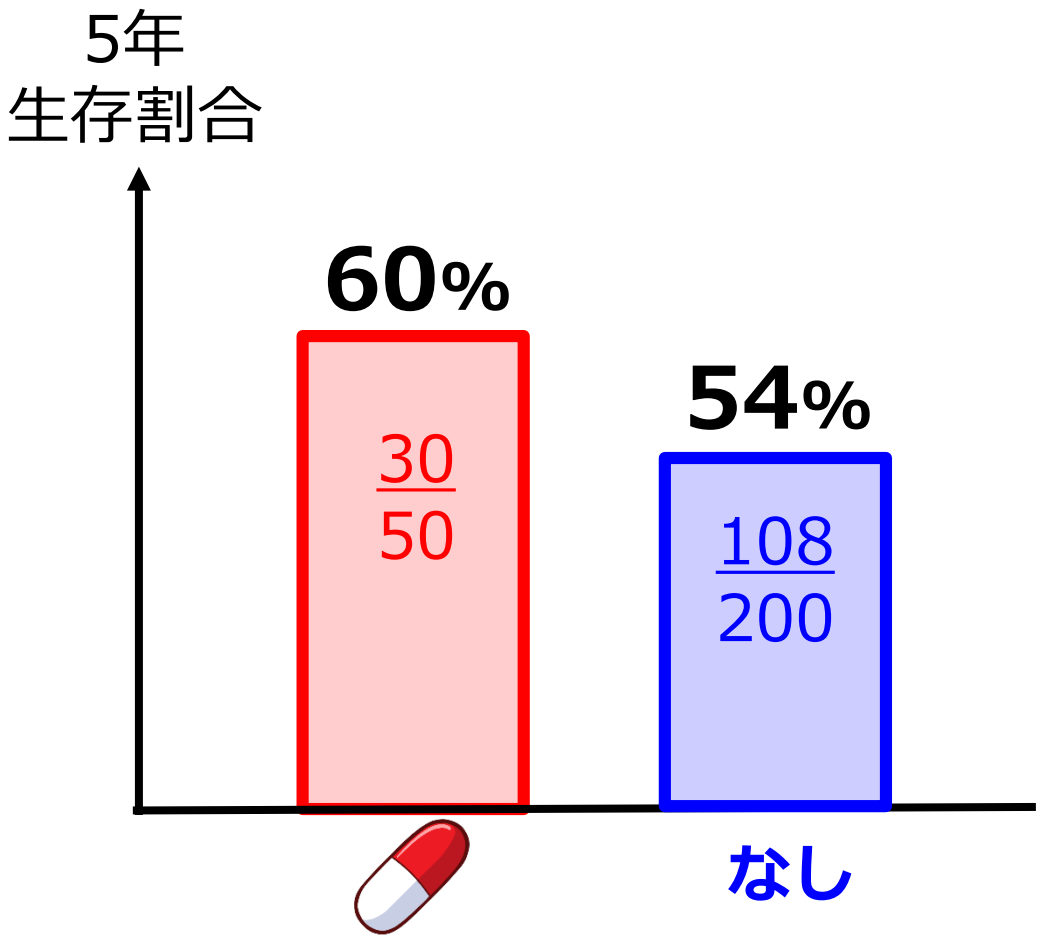
多くのスライドは ICRweb [https://www.icrweb.jp]の資料を改変しています

なぜランダム化は必要なのか

注：仮想例

とある学会で見かける発表

- 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)の予後を検討した

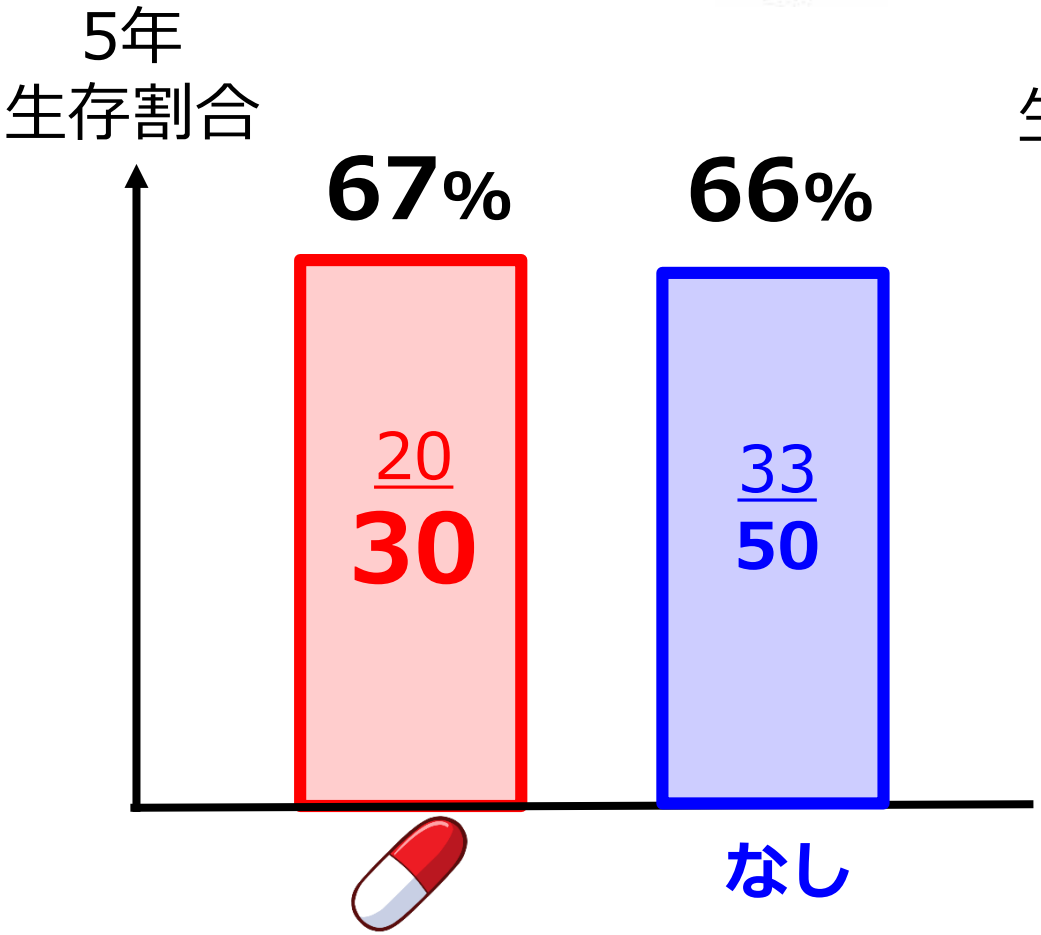


術後化学療法の方が
予後が改善
↓
80歳以上の高齢者にも
術後化学療法を
施行した方が良い

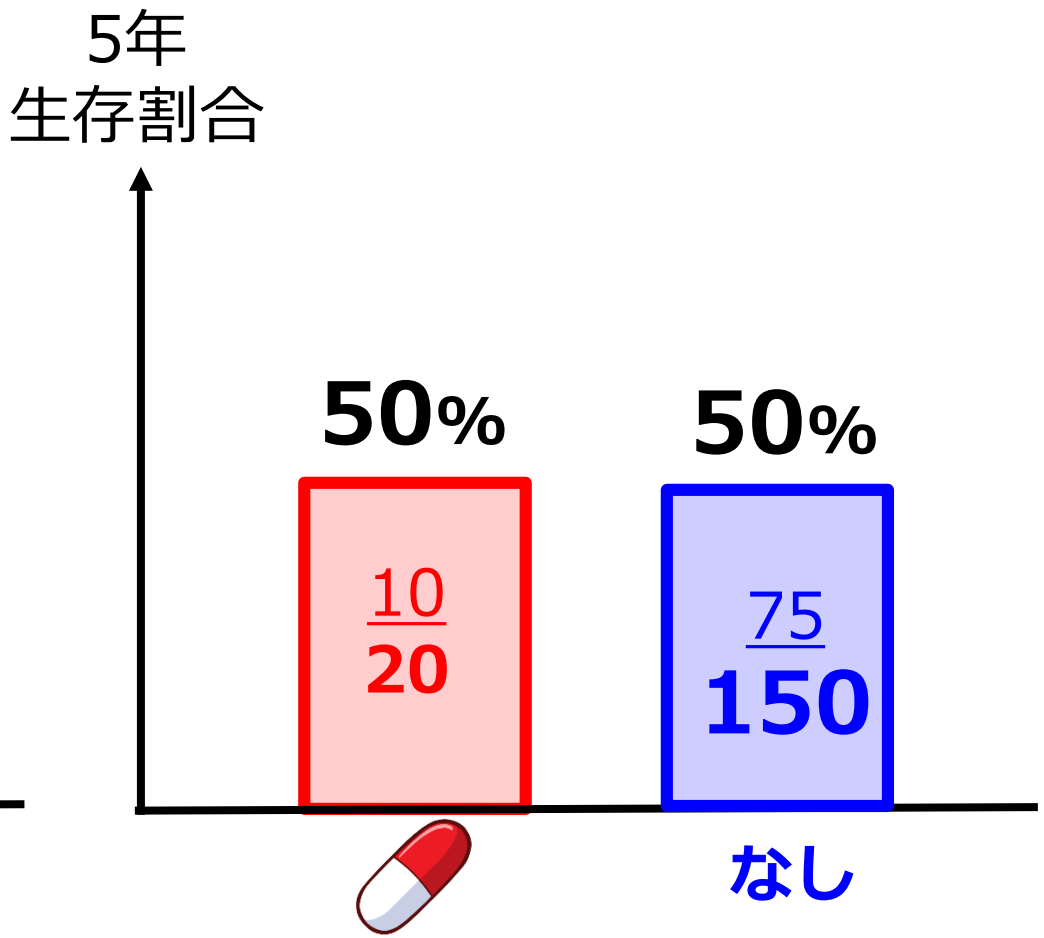
注：仮想例

術後の体調で分けた場合の予後

術後の体調**良好**の患者さん



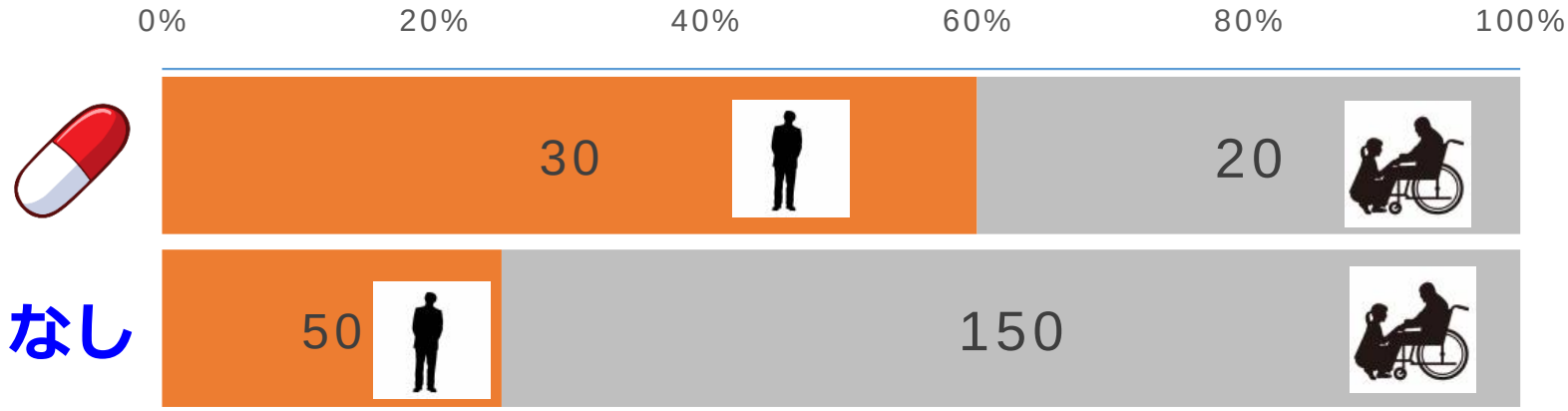
術後の体調**不良**の患者さん



注：仮想例

妥当な比較のためには？

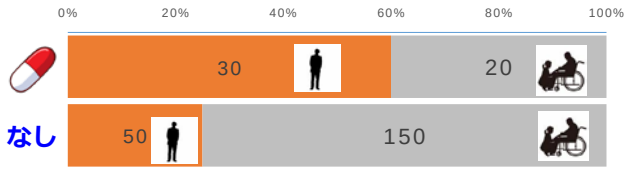
治療法以外の予後に影響する要因の条件が同じにする



- 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**アウトカム**（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



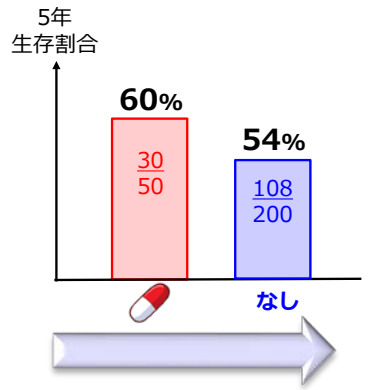
術後の体調

術後化学療法有無

5年生存割合

術後体調は術後化学療法が良好な患者の割合が高い

術後体調良好は5年生存割合が高い



術後化学療法する方が高い

交絡がないことを保証するには

- 比較したいもの（治療）間でアウトカム（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
  - 予後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

↓

<u>ランダム</u>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をすることで
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

必要患者登録数は

どのような考えで

決まっているのか

【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。

中山健夫.健康・医療の情報を読み解く 健康情報学への招待.丸善より改変

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100%、ただし精度は異なる

- 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

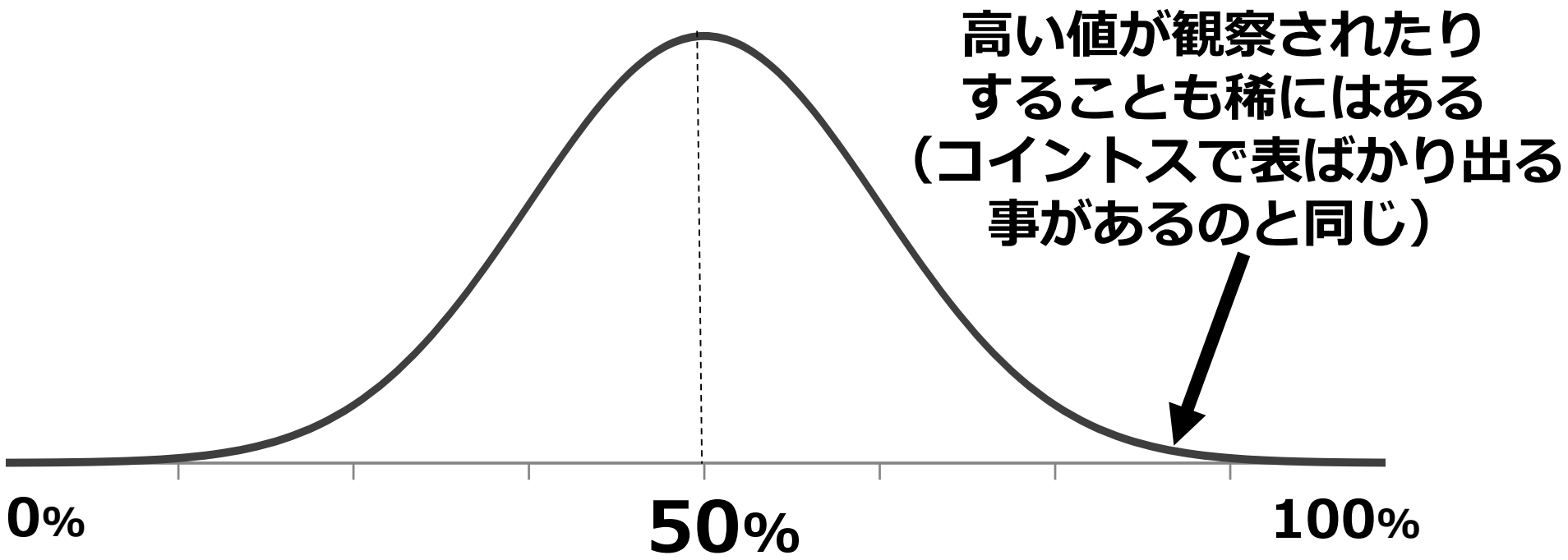
典型的な臨床試験の研究計画書予定登録数の記載

- 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
- Primary endpoint：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：この値では治療法は無効と判断する3年生存割合
- 期待値：まさに今回の治療法で期待する3年生存割合

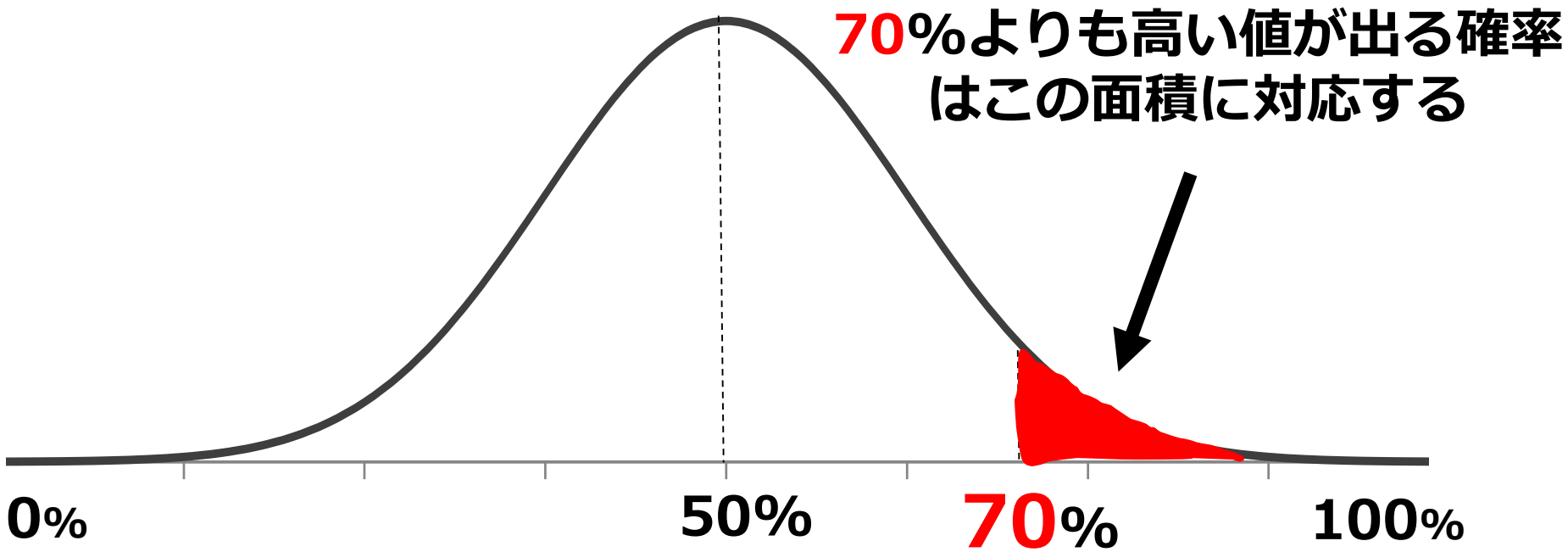
有効な治療法であることをどう判断するか？

- 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<u>判断規準</u>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**<u>有意水準(α level)</u>**という
  - の確率（=**<u>p値</u>**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？

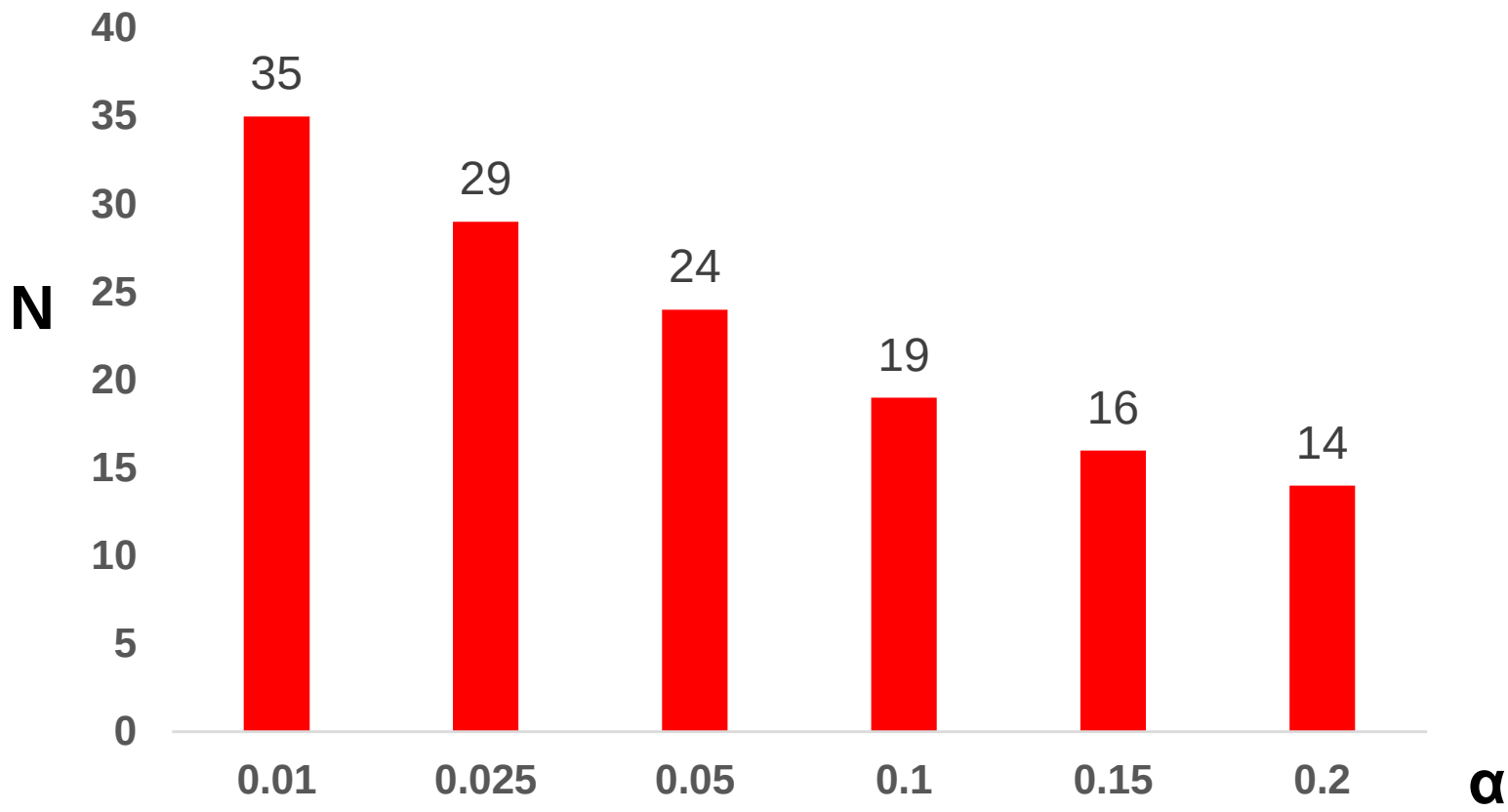
検定における2種類の誤り

		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？

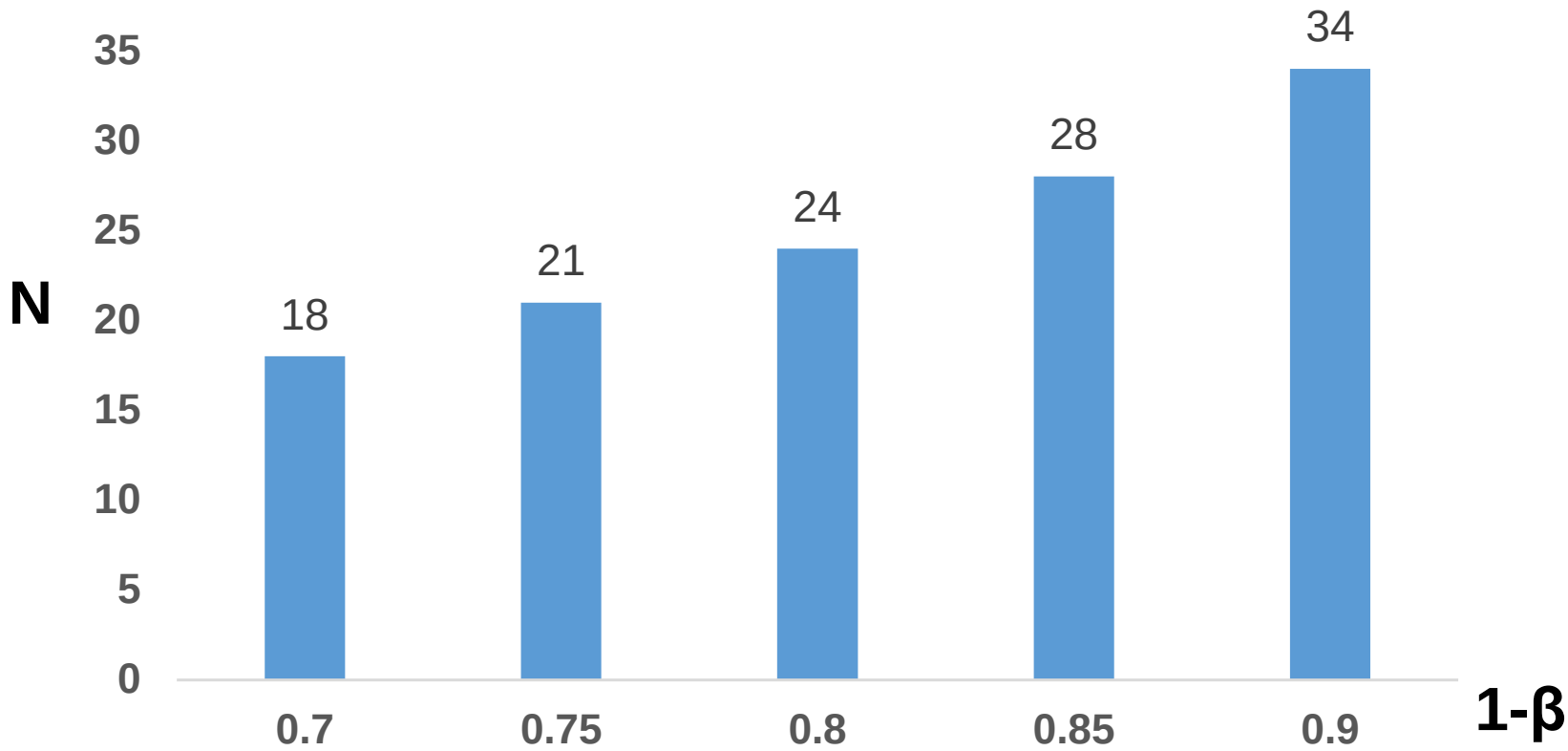
有意水準（α）とNの関係

- αエラー（効果のない薬を効果ありと判断）を小さくするにはNを増やす
- 例：閾値=10%・期待値=30%（治療効果=20%）、検出力=80%



検出力（1-β）とNの関係

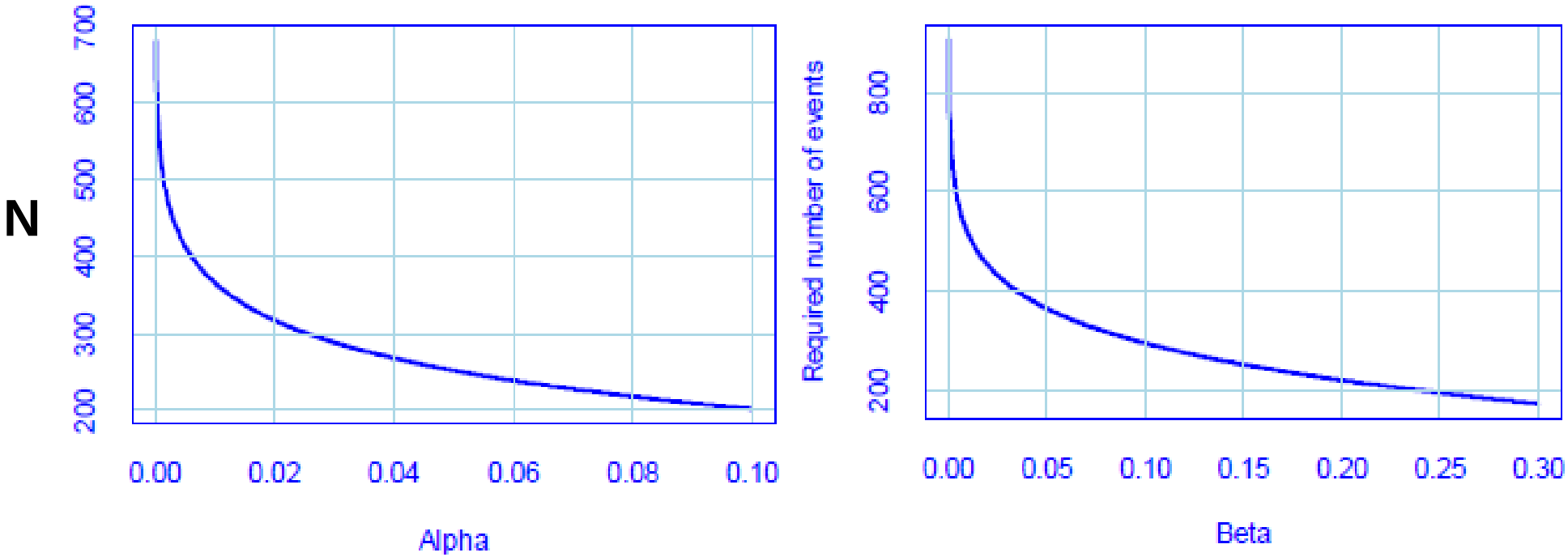
- βエラー（効果のある薬を誤って効果なしと判断）を小さくするにはNを増やす
  - 検出力（効果のある薬を正しく効果ありと判断）を高くするにはN↑
- 例：閾値=10%, 期待値=30%、α=5%



通常のαとβの設定値

- 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

なぜエラーを受け入れないといけないのか？

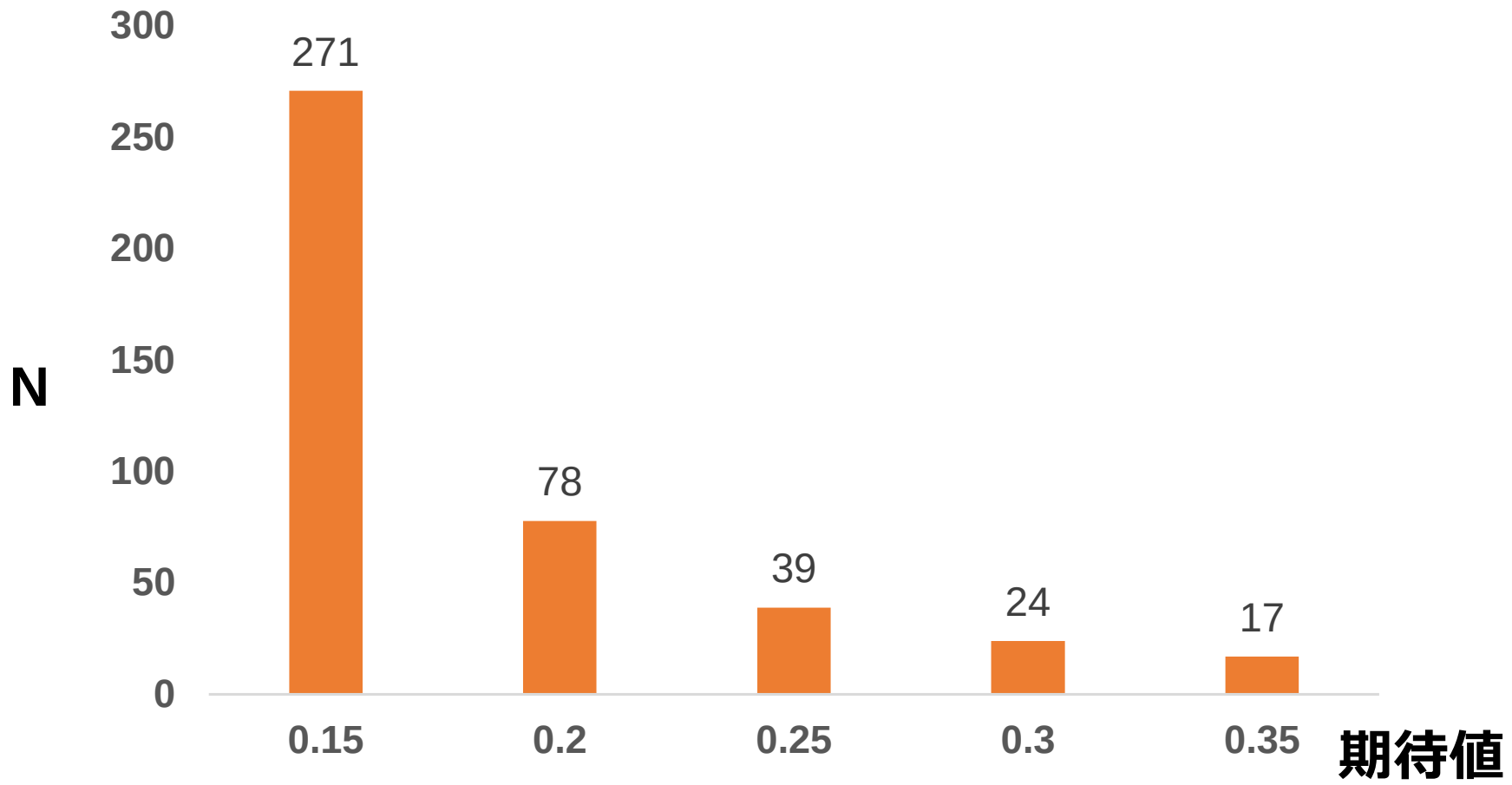


- αとβを厳しくするとNは増えていく
- α=β=0を実現しようとすると、無限のNが必要
  - 現実世界では0リスクは不可能

閾値・期待値とNの関係

- 閾値に比べて大きな治療効果を期待するのであれば、少ないNでOK

  例：α=5%、検出力=80%、閾値=10%の場合



問題：予定患者数が増える条件を全て選んでください

1. 有意水準を10%→5%にした ○

2. 検出力を80%→90%にした ○

3. 見込みの治療効果（割合）の差を30%→20%にした ○

まとめ

- ランダム化は交絡を排除する最強の方法
  - 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すること
  - α↓なら必要N↑
  - β↓（=検出力↑）なら必要N↑
  - Δ↓なら必要N↑