

# プロトコールを深く読んでみる ～統計解析を中心に考える～

国立がん研究センターがん対策情報センター  
山本精一郎

本セミナーの内容は、JCOGデータセンター長福田治彦先生のスライドおよび、平成22年度JCOG臨床試験セミナーの内容をもとに作成しました。

## 臨床研究のプロトコールとは

- 日本語では、研究実施計画書、試験実施計画書、治験実施計画書（省令**GCP**）と呼ばれる。
- どのように臨床試験が実施されるかについて記述した公式文書

「異なる施設/専門/役割の人々が参加して研究を行うにあたり必要な事項が書かれている公式文書」

ープロトコールの中にどこまで記載するか。

施設で共通・個別、時期により変化するか

（例：生化学検査の具体的な検査方法など）

標準業務手順書（**SOP**）：

試験によらない施設単位、役割単位での詳細手順

# 臨床研究には様々な人々が関わる



施設の担当医  
 -患者の治療・評価  
 -患者に対する  
 医学的責任  
 -データ信頼性



臨床検査  
 技師



CRC  
 -試験の登録  
 -調整業務  
 -CRF記入  
 -データ信頼性



看護師

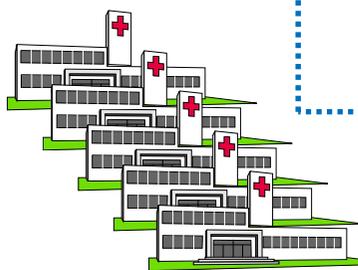


薬剤師

研究代表者/事務局  
 -試験に対する医学的責任  
 -調整業務  
 -進捗管理  
 -結果の解釈



3



製薬企業



各種委員会委員  
 -第三者的監視



データセンター

生物統計家  
 -試験デザイン  
 -試験結果の解析・解釈



事務担当者  
 -書類の管理・保管  
 -庶務



データマネージャー  
 -データ管理  
 -CRF作成  
 -データ信頼性



メディカルライター  
 -試験計画書作成支援



システムエンジニア  
 -データベース構築・管理



監査担当者  
 -データの品質保証  
 -研究プロセスの評価



## プロトコールは誰が読む・使うか？

1. 臨床医・看護師・薬剤師・検査技師等(Local users)
  - ・ 疾患の専門医
  - ・ 非専門医：他分野の専門家、レジデント、研修医
  - ・ **CRC**
  - ・ 病棟、外来の看護師
  - ・ 薬剤部の薬剤師
  - ・ 臨床検査部の検査技師
2. 生物統計家・データマネージャー・モニター等(Central users)
3. 第三者(審査に係わる): IRB、委員会、規制当局等(reviewers)
  - ・ 臨床医：非専門家
  - ・ 生物統計家
  - ・ 倫理の専門家
  - ・ 患者の代表
  - ・ 監査担当者

4

→試験に係わる全ての全ての人が理解出来る必要あり

# 何が記載されるべきか

1. 試験の**意義を示す**記述：声明文的側面（主に背景）
  - ・なぜ、今、この試験を行う価値があるのか（**rationale**）を主張
  - ・研究者相互の認識の共通化
  - ・試験実施機構の意志決定：その試験を実施するかどうか
  - ・医療機関（施設）が受け入れるかどうか（施設**IRB**の審査）
  - ・患者さんへの説明：希望があれば患者さんに開示（説明文書との整合性）
2. 試験の**実施手順**：マニュアル的側面（主に本文）
  - ・対象患者の特定、登録手順
  - ・治療の内容、評価方法
  - ・データ収集と管理の方法、解析方法
  - ・責任体制、連絡先

# プロトコールに必要な記載事項

## 臨床研究に関する倫理指針（2008年7月改訂）〈細則〉

- イ 被験者の選定方針
- ロ 当該臨床研究の意義、目的、方法及び期間、当該臨床研究に参加することにより期待される利益及び起こり得る危険並びに必然的に伴う心身に対する不快な状態、当該臨床研究終了後の対応、当該臨床研究に係る個人情報  
の保護の方法
- ハ 共同臨床研究機関の名称
- ニ 研究者等の氏名
- ホ インフォームド・コンセントのための手続
- ヘ インフォームド・コンセントを受けるための説明事項及び同意文書
- ト 当該臨床研究に係る資金源、起こり得る利害の衝突及び研究者等の関連組織との関わり
- チ 当該臨床研究に伴い被験者に生じた健康被害の補償のための保険等必要な措置
- リ 試料等の保存及び使用方法並びに保存期間
- ヌ 代諾者を選定する場合はその考え方
- 【被験者からインフォームド・コンセントを受けることが困難な場合】
- ル 当該臨床研究の重要性、被験者の当該臨床研究への参加が当該臨床研究を実施するに当たり必要不可欠な理由

# 臨床研究の7つの倫理要件 とプロトコール



7

# 臨床研究の7つの倫理要件

倫理要件

詳細要件

|  |  |
|--|--|
| <b>Social or scientific value</b><br>社会的/科学的価値           | 診断・予防・治療の向上に貢献、疾患・健康に有用な知識を得る、既にある知識や無駄な重複ではない   |
| <b>Scientific validity</b><br>科学的妥当性                     | 一般的に認められた科学的方法論<br>適切な統計手法・正しいデータ  |
| <b>Fair subject selection</b><br>適正な被験者選択                | 社会的弱者の保護・過大なリスクのある被験者の除外<br>利益を受ける集団とリスクを受ける集団が解離しない   |
| <b>Favorable risk-benefit ratio</b><br>適切なリスク/ベネフィットバランス | リスクの最小化・ベネフィットの最大化<br>被験者のリスクに見合う被験者/社会のベネフィット   |
| <b>Independent review</b><br>第三者審査                       | 研究と利害関係を持たない独立した第三者による<br>デザイン・対象・リスク/ベネフィットの評価  |
| <b>Informed consent</b><br>インフォームドコンセント                  | 研究目的・方法・リスク・ベネフィット・代替治療の<br>十分な説明 ( <b>information</b> )、理解 ( <b>comprehension</b> )、自発同意 ( <b>voluntariness</b> ) |
| <b>Respect for potential and enrolled subjects</b>       | 同意撤回の自由、プライバシー保護<br>開始後の新知見や研究結果の説明  |

8

Emmanuel EJ, et al. JAMA 285:2701-11, 2000

継続的に被験者同意を監視

# 臨床研究における7つの倫理的要件

## 社会的・科学的価値 (Social or Scientific Value)

将来の診断法、治療法や公衆衛生の進歩・発展に貢献する結果/結論を導くことができる。

なぜ、この研究が必要か、何がわかるか、かわるか？

→疾患、対象、治療に関する網羅的かつ最新の知見の記述

# 臨床研究における7つの倫理的要件

## 科学的妥当性 (Scientific Validity)

一般的に正しいと認められた科学的原則に基づいて研究を実施

なぜ、このデザインか、エンドポイントか、サンプルサイズか？

→ 妥当な解析方法、モニタリング、監査手順の記述

# 臨床研究における7つの倫理的要件

## 適正な被験者選択 (Fair Subject Selection)

適切な適格規準・除外規準の設定

不適格例を登録しない仕組み・適格性の事後  
チェック

なぜ、この患者は含め、この患者は含めないか？

→適切かつ明確な適格規準・除外規準、不適格例の  
取り扱い、その最小化の方法を記述

# 臨床研究における7つの倫理的要件

適切なリスク・ベネフィットバランス

**(Favorable Risk-Benefit Ratio)**

被験者のリスクとベネフィットの明確化

被験者のリスクの最小化・被験者の利益の最大化

→ 予期される有害事象

適切な安全性評価とモニタリング

適切な治療変更規準

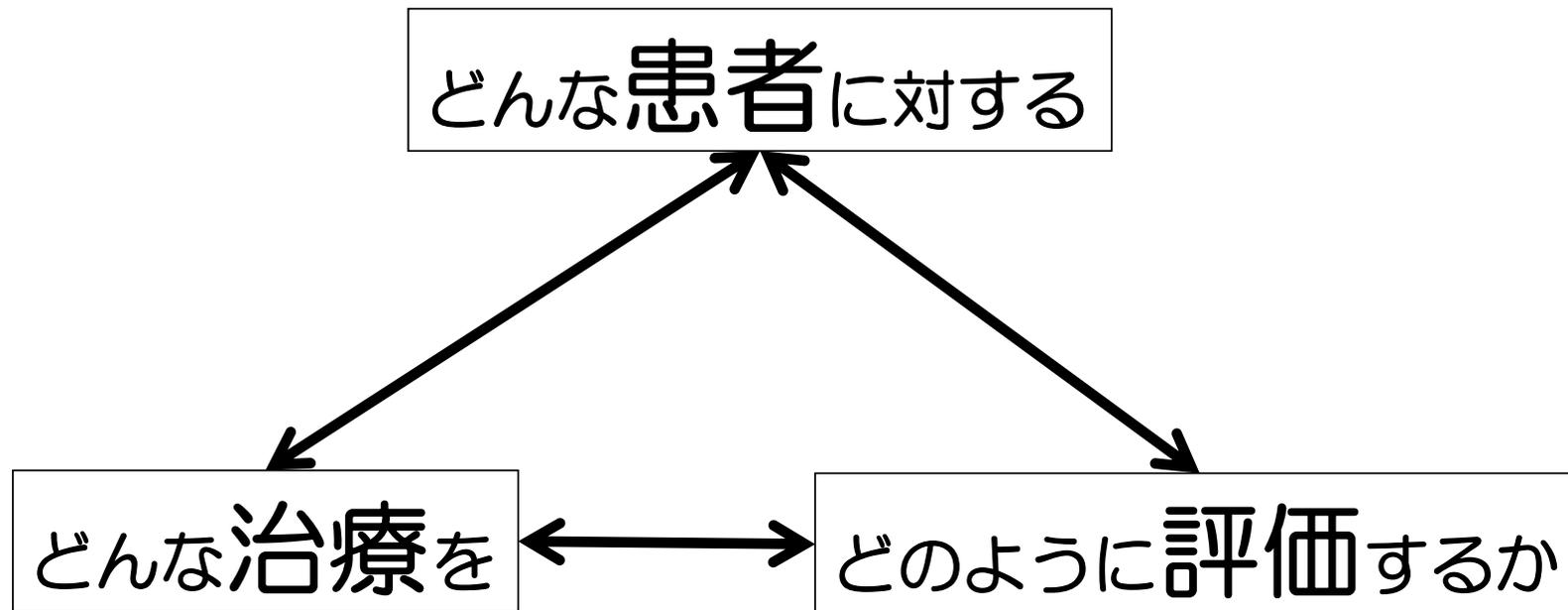
逸脱・違反の扱いと最小化の方法

重篤な有害事象の扱いと最小化の方法を記述

# 臨床研究のプロトコールの作成



## プロトコール：3つの基本要素



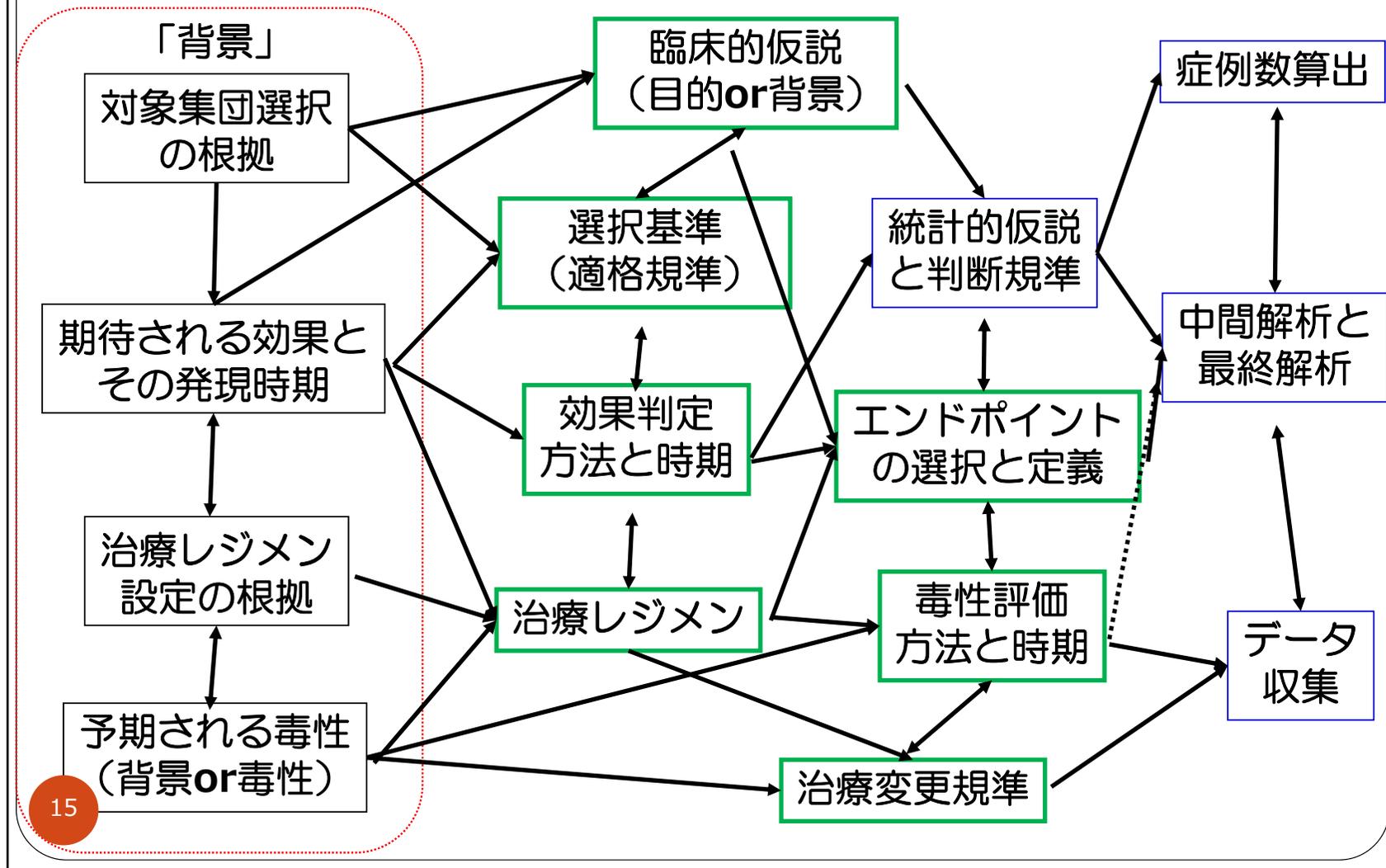
- ・ 明確であること
- ・ たがいに整合性があること

14

cock SJ: Clinical Trials: A Practical Approach, John Wiley, Chichester, 1983)

# プロトコルの記述項目の相互関係

(佐藤、福田：臨床試験の進め方(大橋、荒川編)、南江堂,2006)



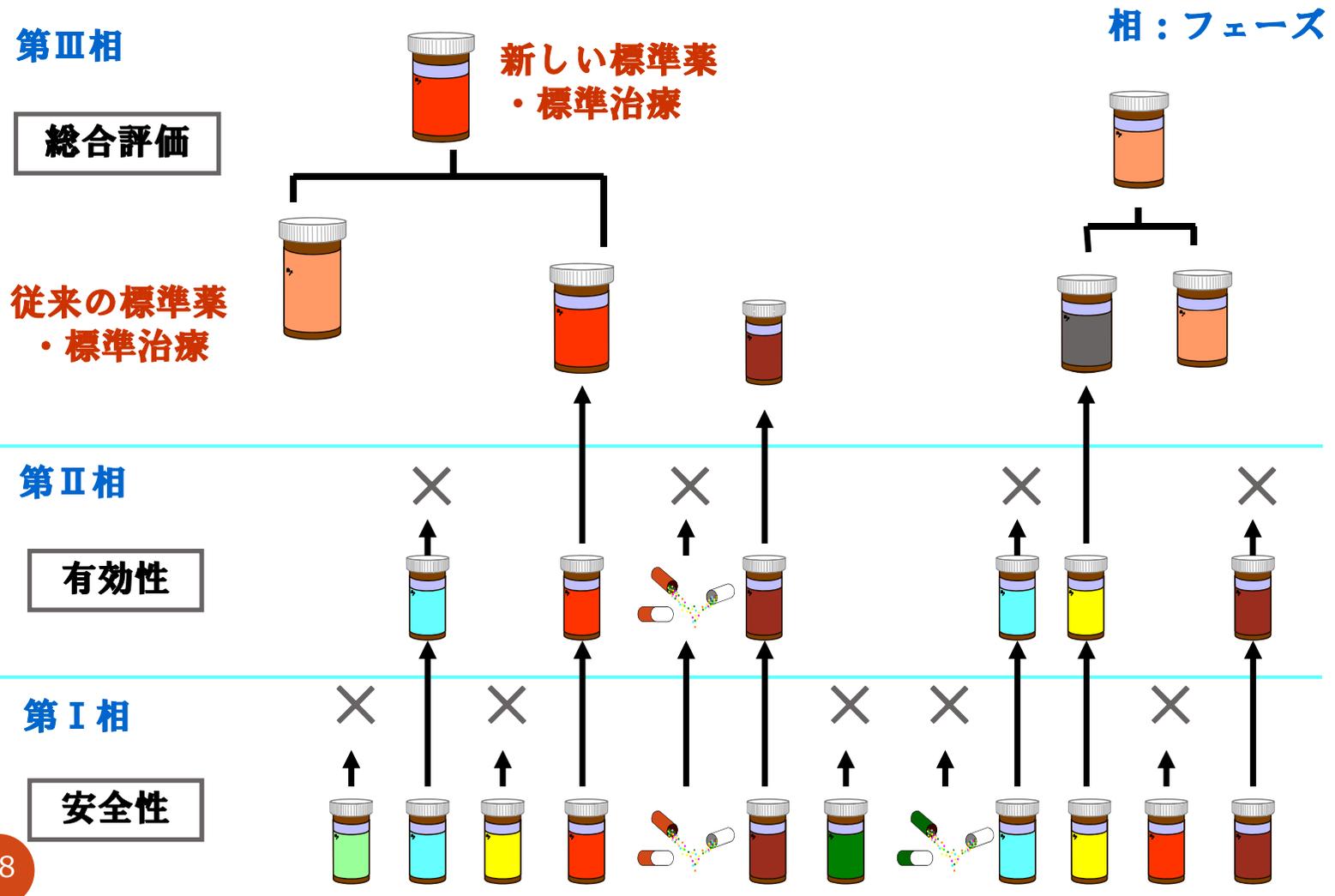
15

# JCOGプロトコルマニュアル (ver2.1)における章構成

0. 概要
1. 目的
2. 背景と試験計画の根拠
3. 本試験で用いる規準・定義
4. 患者選択規準
5. 登録・割付
6. 治療計画と治療変更規準
7. 薬剤情報と予期される有害反応
8. 評価項目・臨床検査・  
評価スケジュール
9. データ収集  
など
10. 有害事象の報告
11. 効果判定とエンドポイントの定義
12. 統計的事項
13. 倫理的事項
14. モニタリングと監査
15. 特記事項
16. 研究組織
17. 研究結果の発表
18. 参考文献
19. 付表  
説明・同意文書、CRF

# ランダム化比較試験(phase3)の研究計画書における統計記述

# 治療開発は臨床試験によるスクリーニング



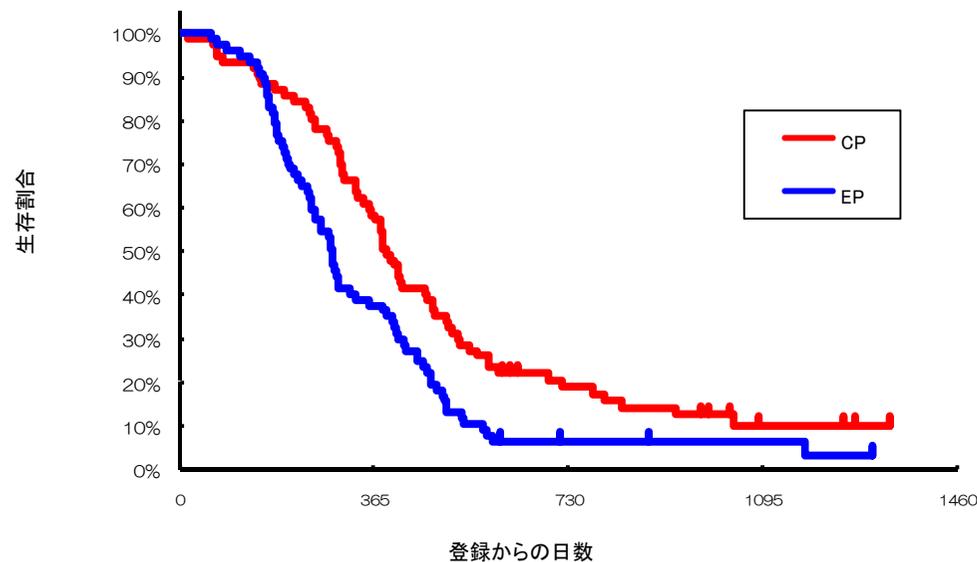
18

## 第III相臨床試験 (Phase III)

### ● 目的

- 現在の最も有効な治療法（標準治療）とランダム化による同時比較を行うことで、比較可能性を最大限に高め、「新しい治療法が標準的治療法となりうるか」を検証する
  - 標準治療を対照群にしてRCTを行う
  - 治療法開発の最後のステージ
  - 検証的な結果を出すことが必要
  - 検証的＝何が標準治療がはっきりさせる

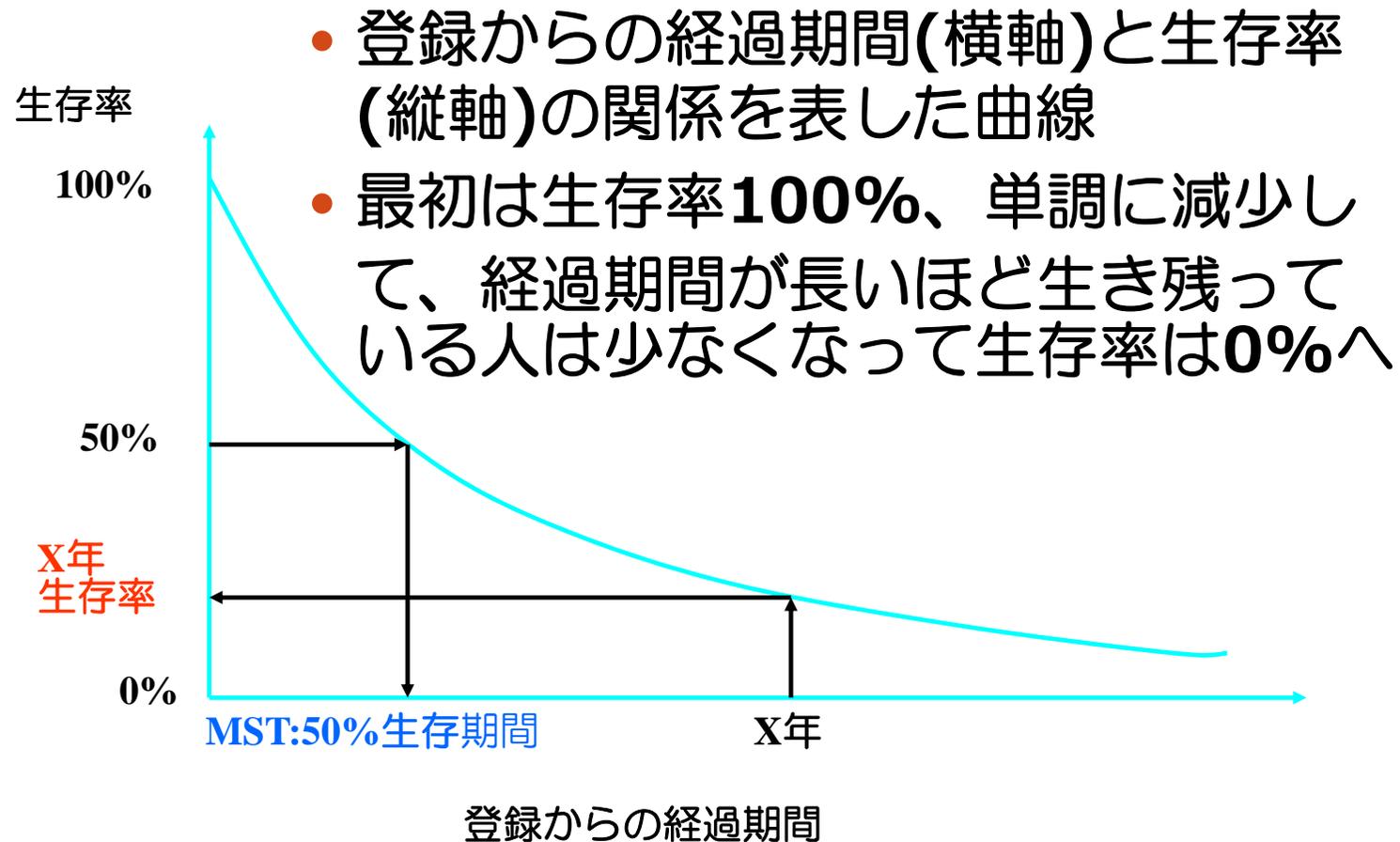
# がん第Ⅲ相試験の結果とは？



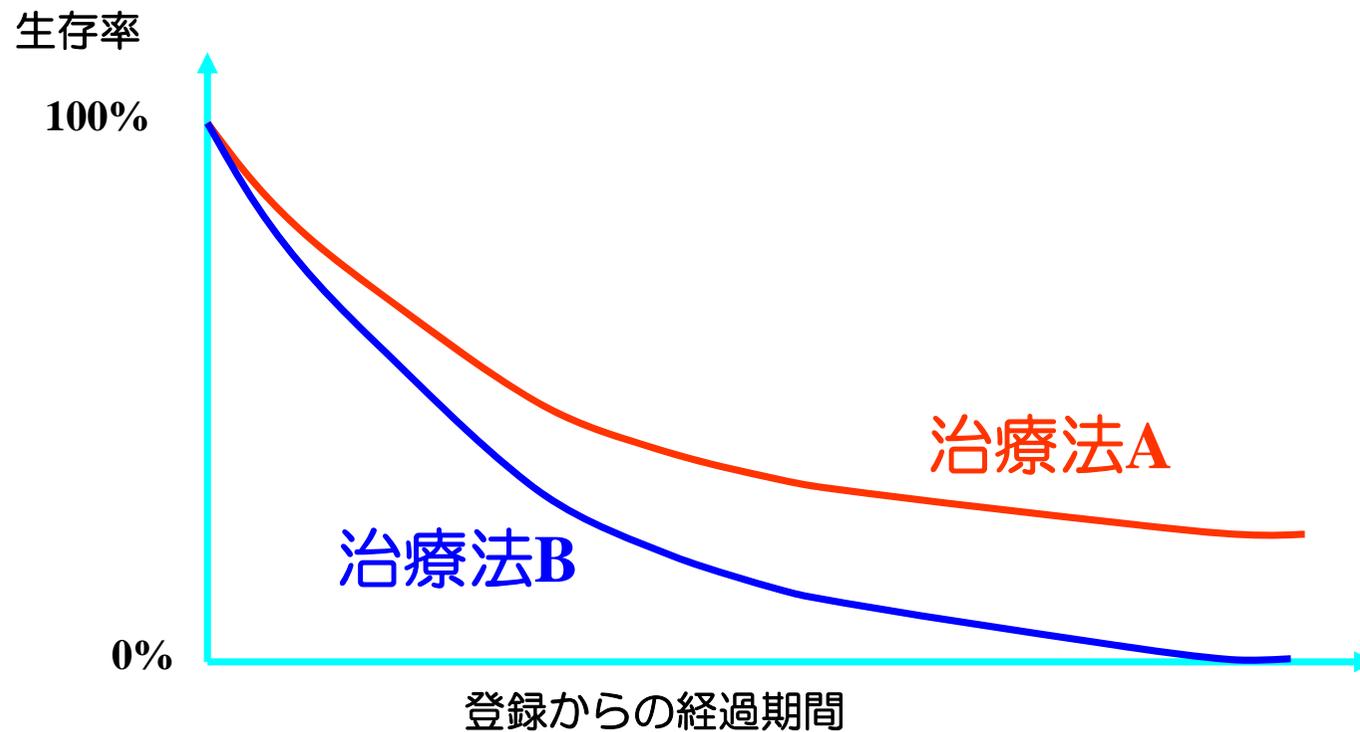
|                              | A 群 (CP)           | B 群 (EP)           |
|------------------------------|--------------------|--------------------|
| 解析対象例                        | 77 例               | 77 例               |
| うち、解析時生存例                    | 10 例               | 4 例                |
| 打ち切り例の最長追跡期間 (日)             | 1333 日             | 1301 日             |
| 1 年生存割合 (%) (95%CI)          | 58.4 (47.4 - 69.4) | 37.7 (26.8 - 48.5) |
| 2 年生存割合 (%) (95%CI)          | 18.9 (10.0 - 27.8) | 6.5 (1.0 - 12.0)   |
| 50% 生存 期 間 (MST: 日 ) (95%CI) | 390 (356-464.0)    | 287 (247-330)      |

調整しない logrank 検定の両側 P 値=0.0021

# 生存曲線はこれだ！



# がん第Ⅲ相試験は生存曲線の 違いを確かめる



## 検証的な研究と探索的な研究

- **検証的とは必ず結論が出る研究**
  - よくコントロールされたランダム化比較試験(RCT)
  - phase III trial
- **探索的とは結果が説得力に欠ける研究**
  - カルテに基づく観察的研究など
  - いわゆる「**レトロスペクティブな研究**」やphase II trial

**必ず結論が出るとはどういうことか？**

## 研究の結果は2種類

- 有意差があった( $p < 0.05$ )
    - ~~治療法以外による差~~ → バイアス・交絡
    - ~~たまたま差が出た~~ →  $\alpha$ エラー
    - 治療法による差が本当にあった
  - 有意差が見られなかった( $p > 0.05$ )
    - ~~治療法以外の要因で差が見えなくなった~~
    - 本当は差があった →  $\beta$ エラー
    - 本当に差がなかった
- Randomized Controlled Trial  
比較可能性の確保
- 偶然の確率を小さく
-

## 有意→偶然か、治療法の差か

- **偶然？**

- コインで同じ面が5回 (6%)
- サイコロで同じ数が3回 (3%)
- 奏効率が65%の時、3例続けて奏効なし(4%)

- **偶然以下なら治療法による差があるとする**

- 5%以下だったら偶然、というのが慣習

→ $\alpha$ エラーを5%とする、という

有意でない

→差がないか、差が出なかったか

- **イカサマ？**

- コインで2回連続表
- サイコロで2回連続1
- 症例数が少ないと、偶然かどうかよくわからない = 差があっても有意になりにくい
- 症例数が十分あると偶然かどうかわかりやすい = 差があるときにはちゃんと有意になるはず

→検出力を上げる (βエラーを小さくする) という

## Phase III trialが検証的な研究 であるためには

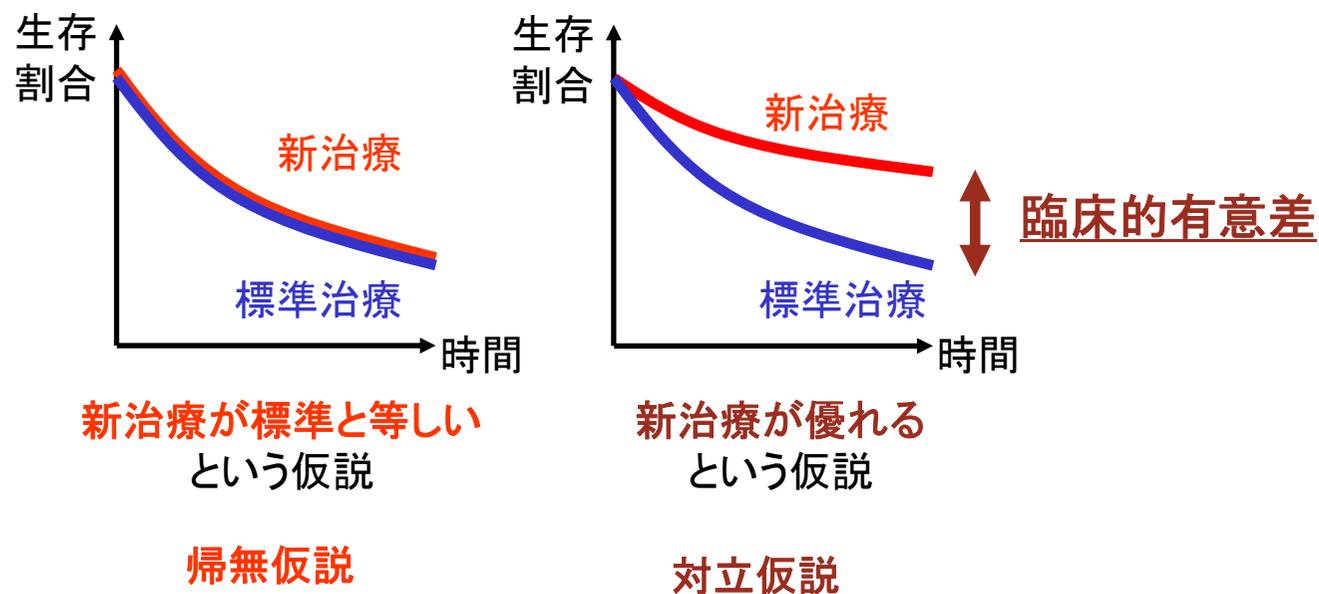
- ランダム化し、よくコントロールすることによって群間の**比較可能性を確保**する
- 差がないときに、偶然差が出たといわないように **$\alpha$ エラーを厳守**する
- 差があるときに、ちゃんと差がでるように **検出力（登録数）を確保**する

研究の途中で変なことしちゃいけません。

# ① 仮説検定の基本的な考え方

## がんのランダム化試験で検証したい仮説

- 新治療が標準治療に対して**生存時間**で優るか否か？
  - 以下の2つの仮説のどちらが正しいか



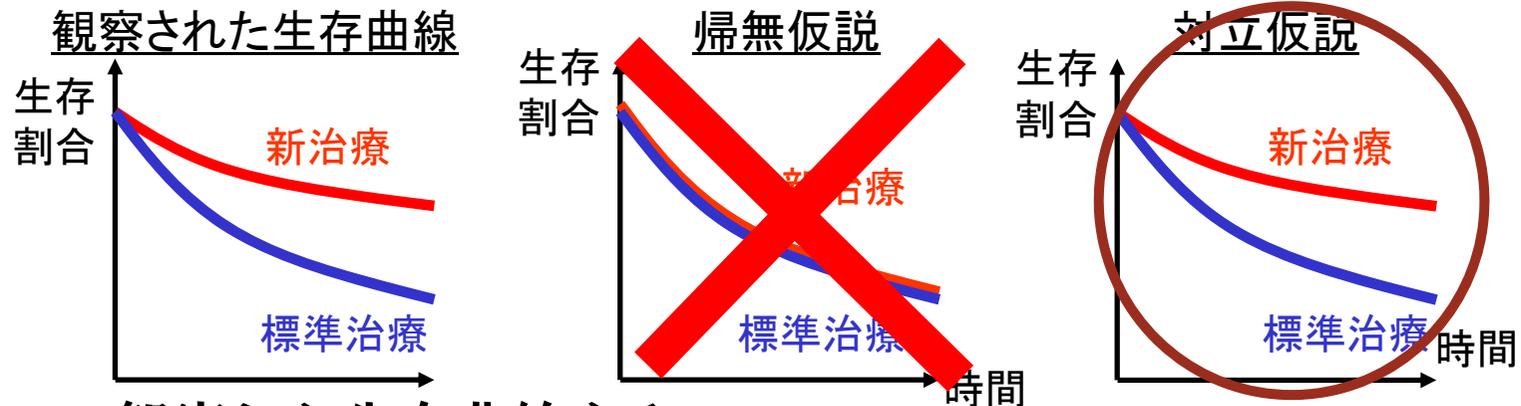
## 対立仮説を証明するには？

{ 帰無仮説： 新治療が標準治療と等しい  
対立仮説： 新治療が標準治療よりも優れる

- 一般に仮説が“正しいこと”は検証できない
  - どれだけデータを得ても、正しいかは分からない
    - 「ある治療の5年生存割合=50%」が正しいことは検証困難。  
もし得られた結果が45%だった場合、仮説は正しいと言い切れますか？
- 反対に、仮説が誤っていることを証明するのは簡単
  - データを集めていけば、誤りであることの検証は簡単
    - ある治療の5年生存割合が50%で“ない”ことは検証可能
      - 多くのデータを収集した結果45%・・・**50%でない可能性が非常に高い**
- 同様に、**対立仮説が正しいこと**を証明するのは困難

反対に、**帰無仮説が誤っていること**を証明するのは簡単

つまり、



1. 観察した生存曲線をみて、  
**帰無仮説「等しいこと」**が誤っていることを示す  
 →**治療効果は等しくない**と結論
2. 等しくないことと生存曲線が上にあること  
 によって、**優れている（対立仮説）**と判断

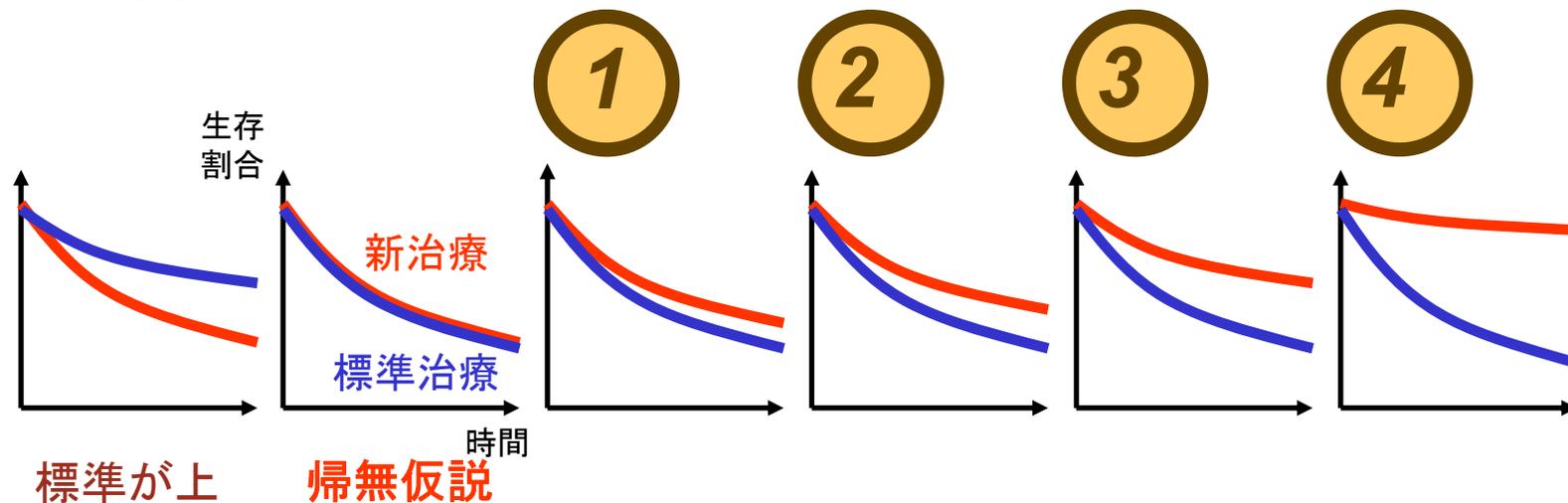
これこそが**仮説検定**の考え方

## ② 仮説検定の具体的なプロセス

**P値とは何か？**

## 試験で得られうる結果を予想してみる！

- 仮説検定：帰無仮説を否定するか／否定しないか判断する

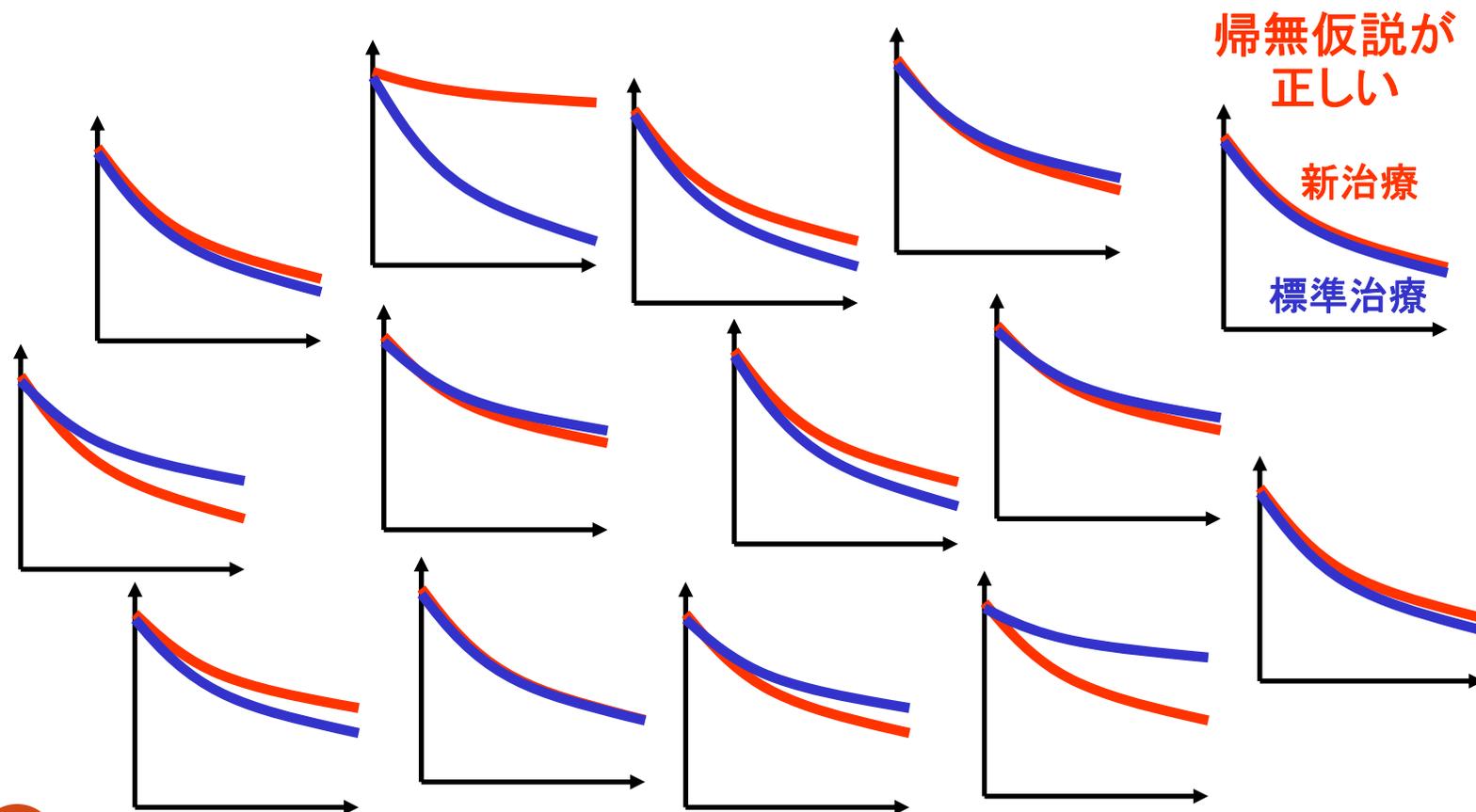


帰無仮説から離れるほど否定しやすいのは明らか  
さて、どうやって判断しようか、..

帰無仮説を否定するのだから帰無仮説が正しい場合を考えよう！

# 帰無仮説が正しくても結果はバラつく

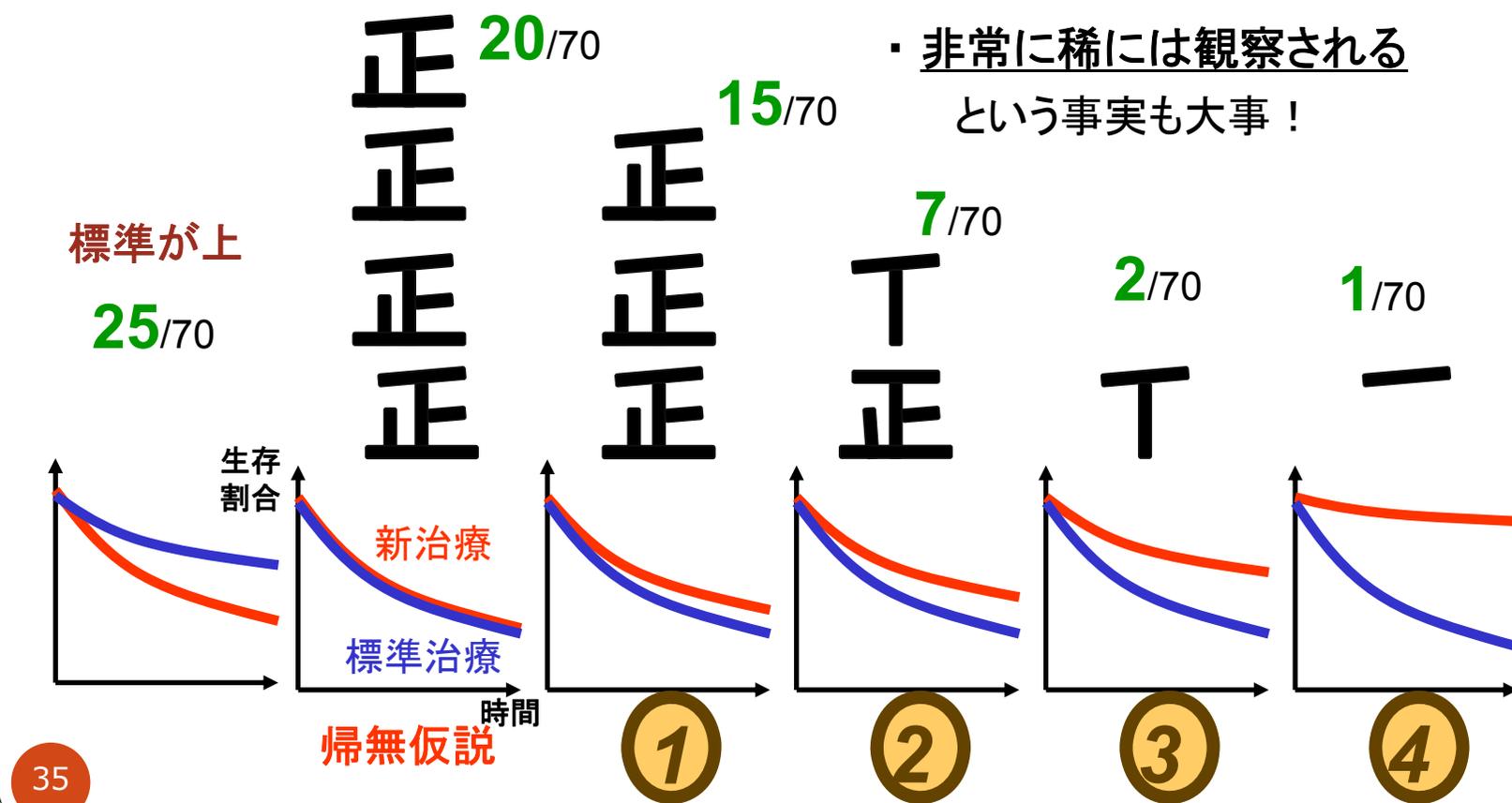
- 同じ試験を何度もやったとしたら結果は当然バラつく



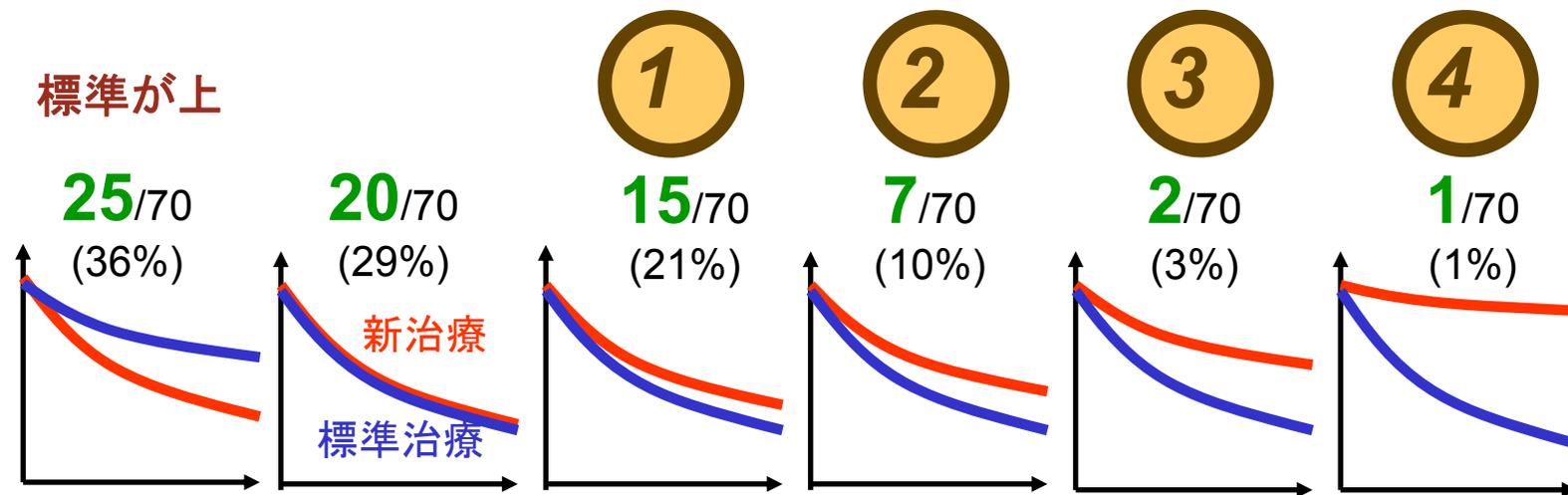
34

# 70回分のバラつきの程度を具体的にみる！

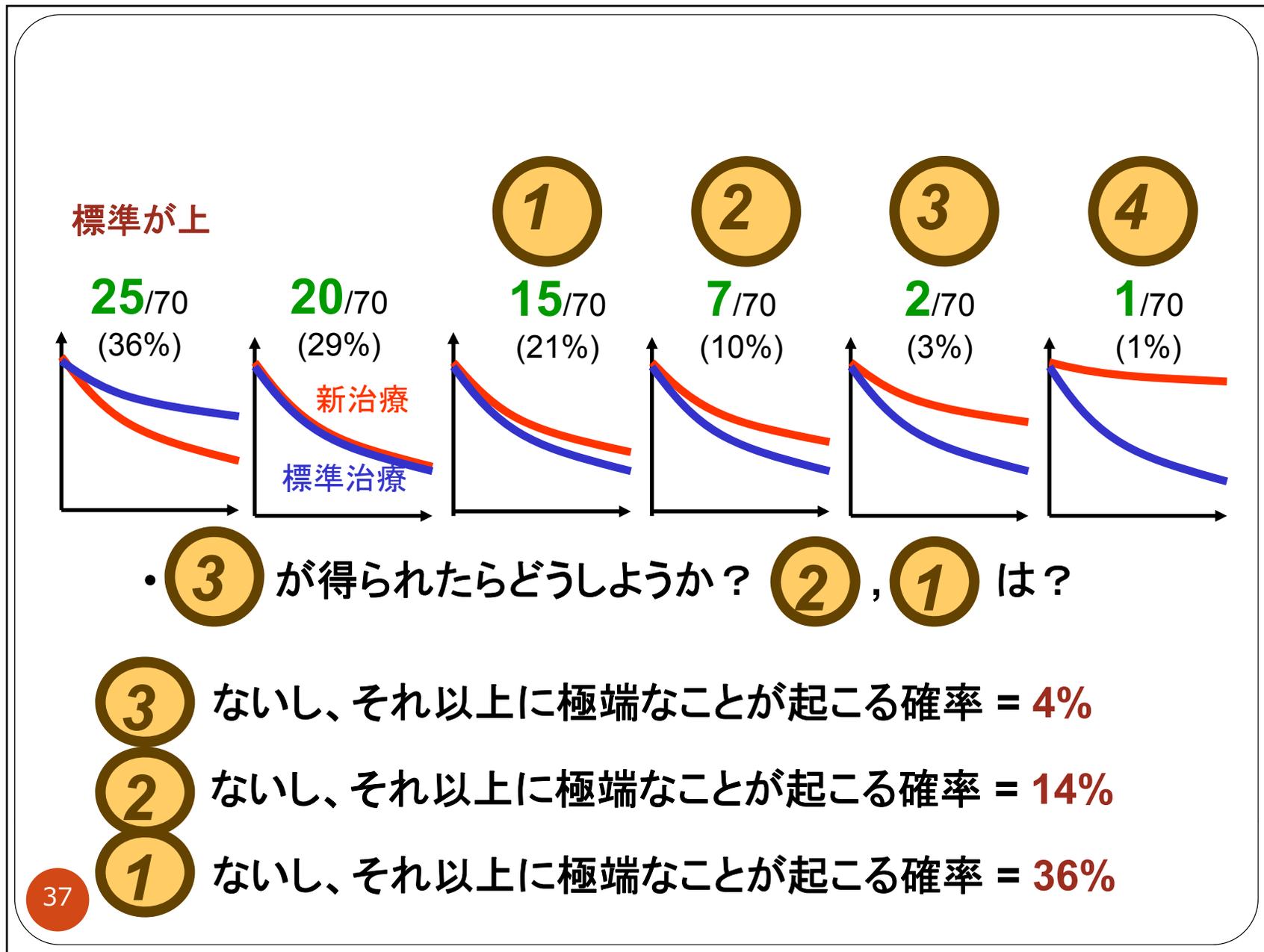
- 真実に近い結果ほど頻繁に観察されやすい
- 真実から離れた結果ほど稀にしか観察されない



帰無仮説が誤りだと言えそうなのは . . .



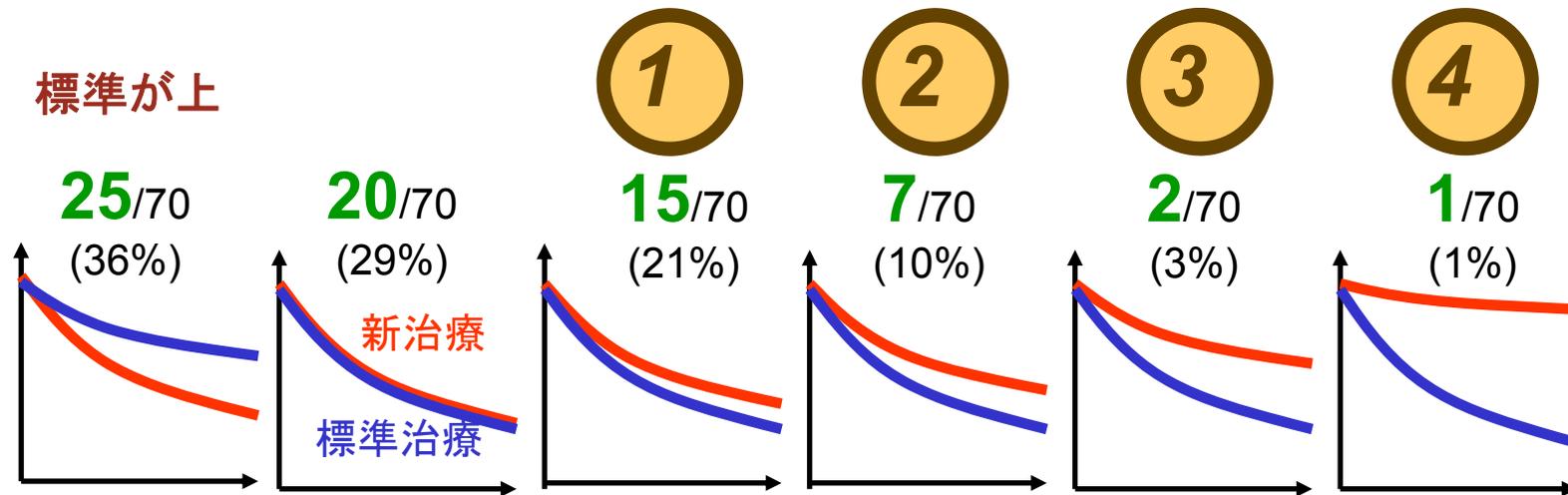
- 帰無仮説が正しいときに, **4** はほとんど起きそうにない
- 試験の結果, **4** が得られたら帰無仮説が正しいというのは無理がある. → 対立仮説が正しいと考えるのが合理的
- **4** が得られたら, 帰無仮説を否定しよう!



## 帰無仮説に対する 結果の不自然さの指標

- **●** ないし、それ以上に極端なことが起こる確率
- 得られた結果が、帰無仮説が正しい下で起こりえる確率を表している
- 帰無仮説の下で  
**得られた結果** ないし **それ以上に極端なこと**  
が起こる確率のことを **p値** とよぶ
- P値がいくつ以下だったら帰無仮説を否定するかは、事前に決めておく

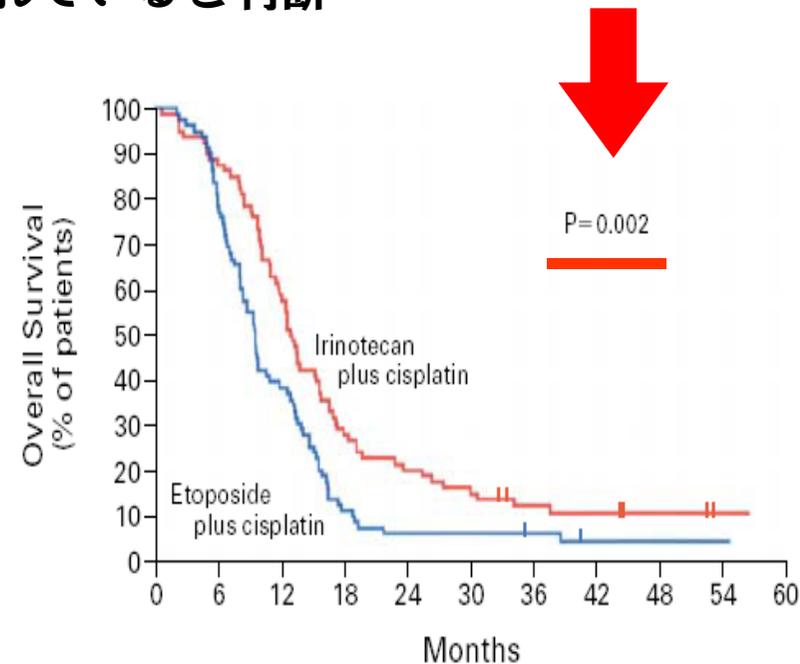
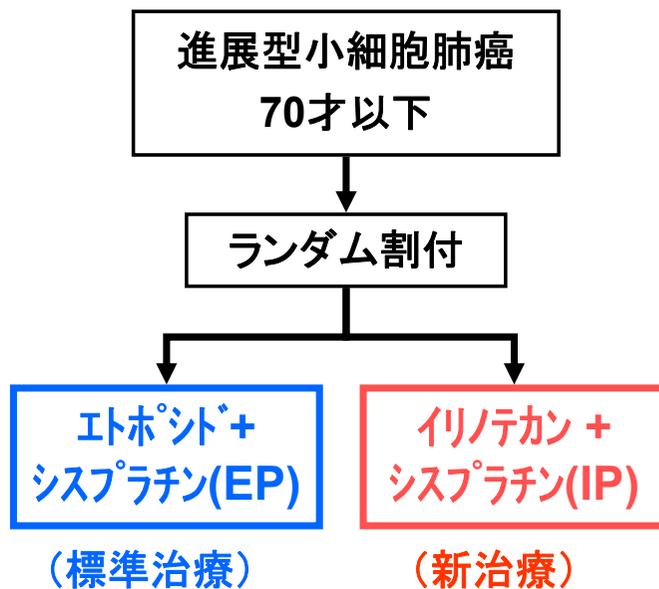
# 帰無仮説を否定するのは . . .



- $P \leq 5\%$ と決めていたら、③ or ④ のとき否定
- $P \leq 2.5\%$ と決めていたら、④ のとき否定
- $P \leq 15\%$ と決めていたら、④ or ③ or ② のとき否定

# 実際に見てみよう！ JCOG9511

- **p=0.2%**, 両群が等しいとしたら稀にしか得られない結果
  - 事前に決めた規準: 「p値 $\leq$ 5%」 規準も満たす
  - **IP群**が**EP群**に比べて優れていると判断



③ 仮説検定にともなう“判断の誤り”

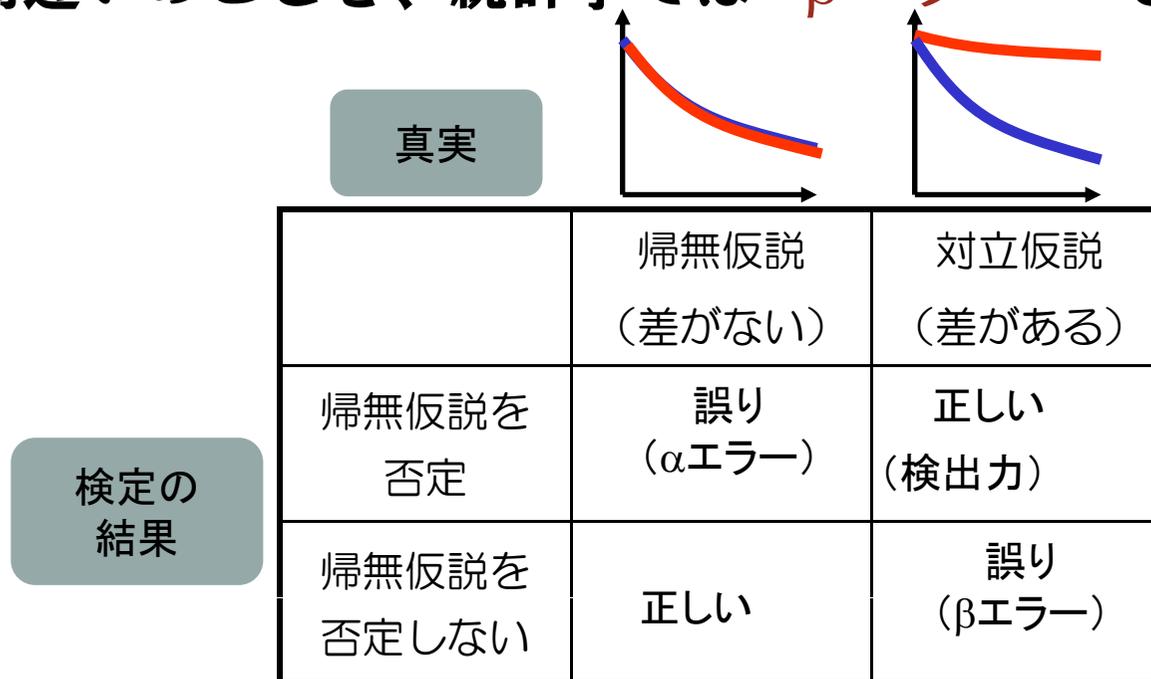
$\alpha$ エラーと $\beta$ エラー

## $P \leq \alpha\%$ のとき . . .

- 帰無仮説の下では、稀な結果が得られたことになる。そんな結果が“たまたま”今起こったと考えるのは不自然なので、そもそも前提（帰無仮説が正しい）を否定しよう  
→ 仮説検定の考え方
- でも、“たまたま”起こることだって有り得るはず。その場合上のルールは間違っただ判定を下すことになる。
- この間違いのことを、統計学では“ **$\alpha$ エラー**”と読んでいる
  - $P$ 値  $\leq 5\%$  の場合に帰無仮説を否定すると事前に決めれば、 $\alpha$ エラーの確率 = 5%以下
  - $P$ 値  $\leq 15\%$  の場合に帰無仮説を否定すると事前に決めれば、 $\alpha$ エラーの確率 = 15%以下

反対に、

- 「帰無仮説（差がない）が間違っているのに、対立仮説（差がある）を主張しなかった」という間違いもある
- この間違いのことを、統計学では " **$\beta$ エラー**" と呼んでいる



両方小さくできればいいのだけれど、

- $\alpha$ エラーと $\beta$ エラーを同時に小さくすることはできない
  - どちらの誤りを小さくすることを優先するべきか？
    - $\alpha$ エラー：効果のない治療が世の中に出してしまう(消費者危険)
      - 後で効果のないことがわかったら、生産者側のリスクにもなる
    - $\beta$ エラー：効果のある治療を世の中に出せない(生産者危険)
- 消費者に対するリスクを抑えるため  $\alpha$ エラーを小さくすることを優先

## サンプルサイズ設計

sample size determination

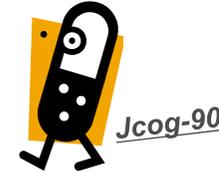
プロトコールに記載されている  
「サンプルサイズ」はどうやって決まるのか？

45

## サンプルサイズ???

- サンプルサイズはどうやって設定しているか？
  - プロトコールを見ると  
「…各群150人を目標登録患者数として…」
- よく言われるのは、
  - 適当に“えいや”と決めているのでしょ
  - これ位だったら2年の登録期間で集まるからそうしたのでしょ
  - 契約病院ごとの契約症例数を合計して、これになったのでしょ
- でも、プロトコールには“検出力を90%として”??とか  
もっともらしいことが書いてあるんだよね……

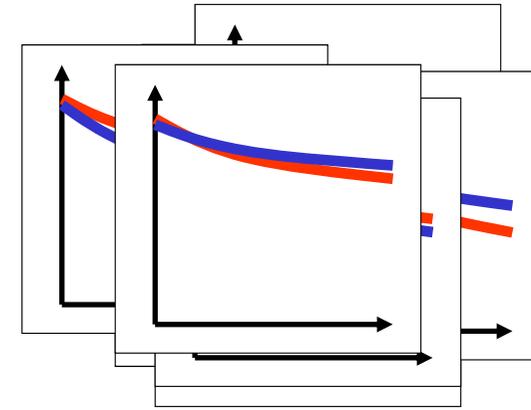
# ここでは簡単に5年生存割合での比較を考える



- エンドポイント：5年生存割合
- 新規分子標的薬(J)の臨床試験
  - 殺細胞薬(S)+J併用療法 vs. S単剤療法のランダム化
  - ランダム化比較試験の結果
    - S+J併用群の5年生存割合： **80%**
    - S単剤群の ~ : **40%**
  - S+J併用群がS単剤群に比べて優るか？
    - 帰無仮説が正しいのに偶然生じた可能性も十分にある
    - 統計学を使う
      - 仮説検定でp値を求め、帰無仮説が否定できるか判断
      - もちろん、p値がいくつ以下の場合に否定するかも必ず事前に決めなければならない

じゃあ、統計学で

- 割合の群間差に含まれるバラつきを統計学を用いて実際に見積もる
  - 帰無仮説が正しいとした場合の得られたデータの不自然さをp値で表して **仮説検定!**
    - **帰無仮説** 「割合の差=0」
- 統計学では以下が分かっている
  - データを得た症例数が **“多”** ければ、バラつきは **“小”** さくなる
  - 反対に、データを得た症例数が **“少”** なければ、バラつきは **“大”** さくなる



同じ割合でもねえ

- **小規模試験 (各群5人)**
  - **S+J併用群** 5人中 **4人生存 (80%)**
  - **S単剤群** 5人中 **2人生存 (40%)**
  
- **大規模試験 (各群1000人)**
  - **S+J併用群** 1000人中 **800人生存 (80%)**
  - **S単剤群** 1000人中 **400人生存 (40%)**

偶然によって、2人ぐらい・・・

- **小規模試験 (各群5人)**

- S+J併用群 5人中 **2人**生存 (40%)

- S単剤群 5人中 **4人**生存 (80%)

2人↓減少  
2人↑増加

- **大規模試験 (各群1000人)**

- S+J併用群 1000人中 **798人**生存 (80%) 2人↓減少

- S単剤群 1000人中 **402人**生存 (40%) 2人↑増加

つまり、先ほどの

- **小規模試験 (各群5人)**

- S+J併用群 **5**人中 **4**人生存 (80%)

- S単剤群 **5**人中 **2**人生存 (40%)

- **大規模試験 (各群1000人)**

- S+J併用群 **1000**人中 **800**人生存 (80%)

- S単剤群 **1000**人中 **400**人生存 (40%)

- 2つの結果は全く異なる解釈をすべき!

つまり、

Scientist  
科学者



- サンプルサイズの **“多い”** 試験の結果であれば治療法の効果について**確かなこと**が言える
  - 各群1000人の結果であれば、明らかにS+J併用群がS単剤群に対して生存割合で優ることを検証できそう
- じゃあ、とにかく大きな試験をやれば良いのか？
- **科学（実験）的には Yes ○**
  - 結果が信頼できる
  - どんなに僅かの差の勝負であっても白黒つく

大規模試験にすれば、、、



- **大規模試験の結果**

- S+J併用群 1000人中 **800**人生存 (**80%**)
- S単剤群 1000人中 **750**人生存 (**75%**)

- **偶然性によって2例ぐらゐの結果が**

- S+J併用群 1000人中 **798**人生存 (**80%**)
- S単剤群 1000人中 **752**人生存 (**75%**)

- **もっと小さな5%の差であっても十分検証可能**

- 統計的に厳密に考察しても各群1000人の試験ならば5%の差に対する**検出力**は80%以上 ( $\alpha$ エラー=5%)

## 倫理的には . . . 絶対に NO !

- もっと少なくとも科学的な検証が十分にできていたのにも関わらず、余分な患者さんを使って人体実験！
- そもそも、非常に多数の患者さんに参加してもらって僅かな差を“科学的（統計的）”に検証できたとしても“臨床的”には無意味な差であつたら意味がない！
  - 生命を脅かす毒性が高頻度であるのに生存割合を僅かにしか改善しない治療ならば不要

統計的に意味のある差 = 臨床的に意味のある差  
では無い！

## そこで

- 差を証明するのに十分で多すぎない症例数
  - ここでの差は、臨床的に意味のある差（臨床的有意差）  
→ 証明するに値する最小の差
  - この差が本当にあったときに、この差を証明できる確率（検出力）が十分な大きさとなるように設計
    - 80%～90%程度がよく用いられる
    - 検出力は研究の倫理的担保・保証
      - 臨床的有意差があっても高い確率で証明できないような小規模な試験は実施するだけ無駄で、倫理的に問題あり
- 通常、統計家が臨床家の持つ臨床的仮説と相談しながら設計する！

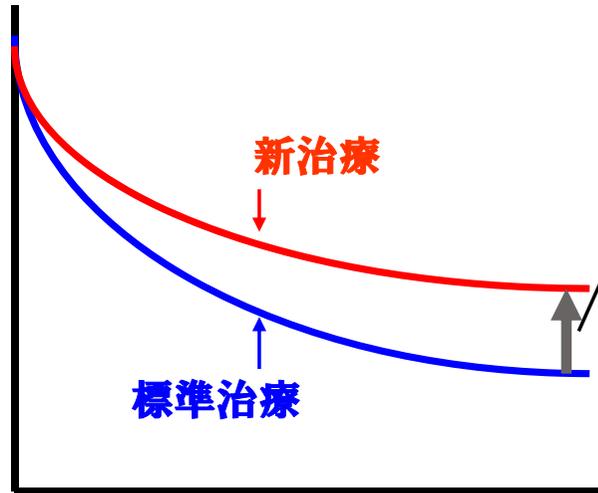


## 第Ⅲ相試験には2つのタイプがある

- Standard vs. Toxic New
  - **毒性の強い新治療**
- Standard vs. Less Toxic New
  - **毒性が軽い新治療**

# Standard vs. Toxic New

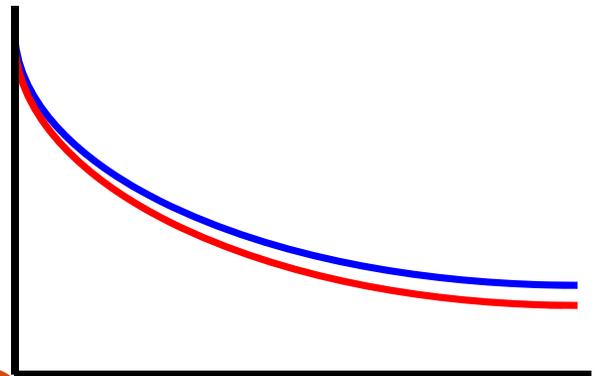
毒性のデメリットに見合う分だけ  
新治療は有効性で勝たなくてはならない



臨床的に意味のある差

これくらい上回っていれば臨床医が新治療を選ぶ  
これ以下だと臨床医が標準治療を選ぶ

「臨床的に意味のある差」があるかどうか  
を統計学的に検証する **優越性試験**  
通常ログランク検定



毒性の強い新治療が下回っている場合は

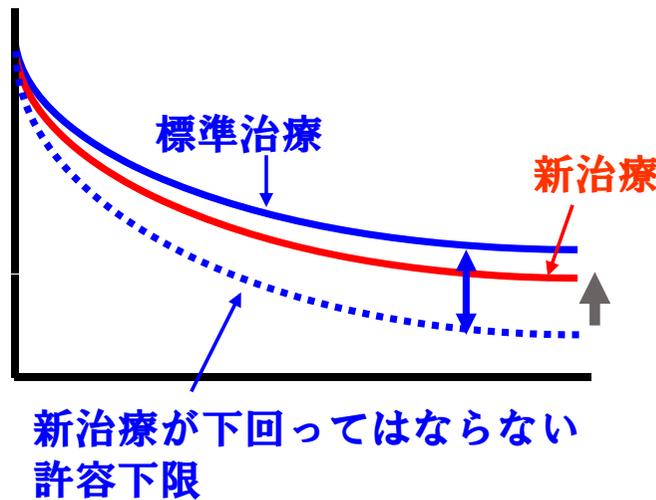
「有意に」劣っているかどうかは関心事項ではない

→ 統計を使わない判断

→ 統計的には「片側」仮説の優越性試験

# Standard vs. Less Toxic New

毒性が少ないのだから新治療の有効性は  
標準治療と同じくらいでよい

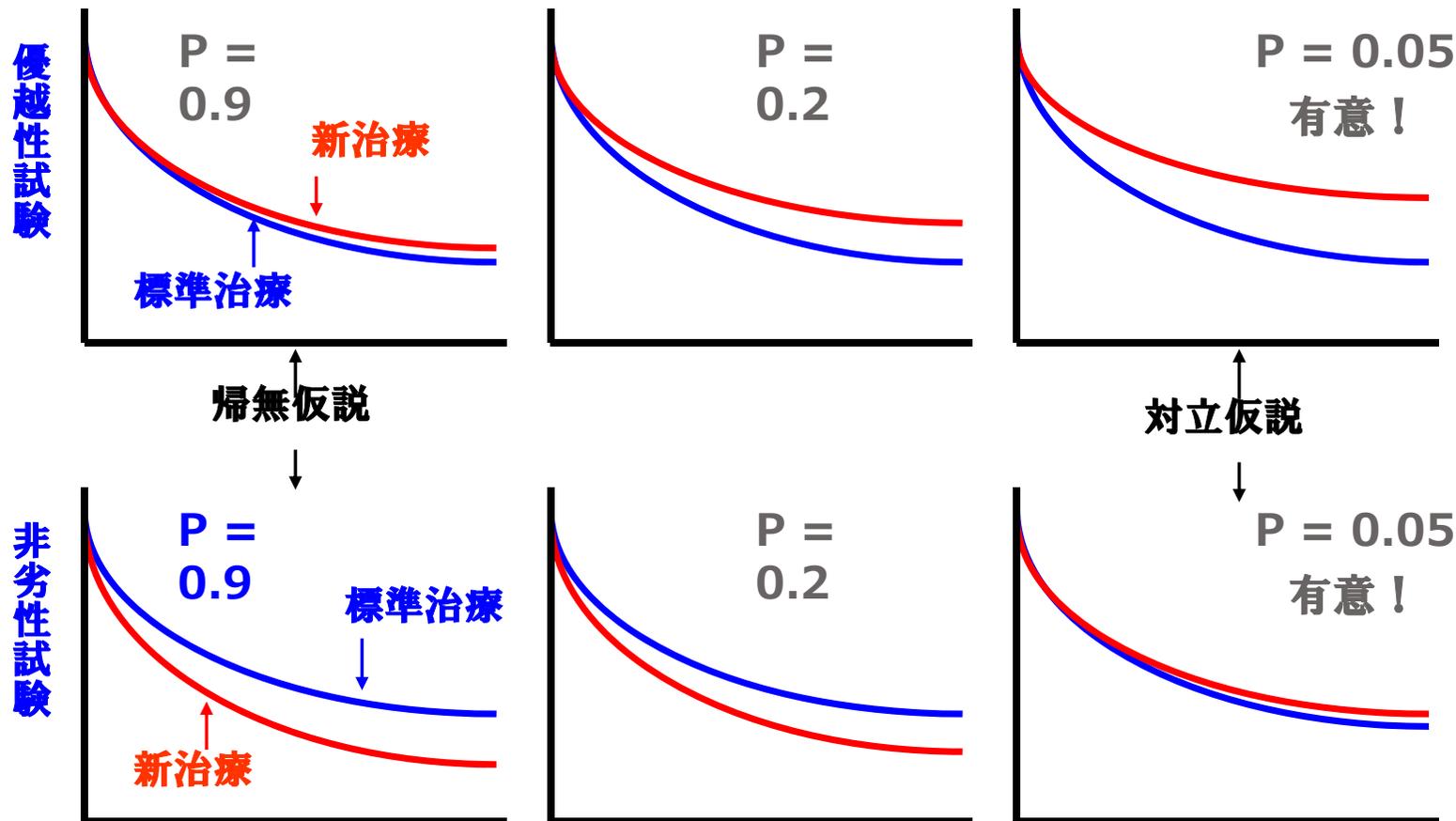


同じ位とは言いにくいので、新治療が  
許容下限を下回らないと言い換える  
下回らないことが言えれば  
「劣っていない」＝「同じ」  
→ 非劣性試験と呼ぶ

許容下限に勝っているかどうかだけを見る  
＝片側仮説

ログランク検定はできないので  
比例ハザードモデルによるハザード比が  
許容限界を超えているかどうかを検定

# 優越性試験と非劣性試験の違い



59

対立仮説に近い状況が観察されたときに有意になる

## 非劣性試験、よくある誤解？

- 対立仮説に近い状況が観察されたときに有意になる
- 対立仮説はふつう“同等”である
  - 「有効性は少し劣っているけど他のエンドポイントで勝っている」ではない。あくまで同等。
  - しかし、必ずしも対立仮説を”同等”にする必要はない
    - 勝っていてもいい
      - 成績もいいし、他のエンドポイントでも勝っている
    - 少し負けていてもいい？
      - できれば同等以上でありたい

## 非劣性デザインの注意点

- 優越性ならランダム化された全例を解析対象にした結果から結論(Intention To Treat:ITT解析)
- いい加減に実施した試験でITTで解析すると、群間で差がなくなる → 非劣性が言い易い？
- そうでないことを何を示すべきか？
  - プロトコールが遵守されていること
  - プロトコールが遵守された症例だけを解析対象にした結果でも結論が変わらないこと(Protocol Compliant解析)
  - 他のエンドポイントで優越性を示すこと

実際にプロトコールにどう書かれているか見てみましょう

# JCOGプロトコルマニュアル (ver2.1)における章構成

0. 概要
1. 目的
2. 背景と試験計画の根拠
3. 本試験で用いる規準・定義
4. 患者選択規準
5. 登録・割付
6. 治療計画と治療変更規準
7. 薬剤情報と予期される有害反応
8. 評価項目・臨床検査・  
評価スケジュール
9. データ収集  
など
10. 有害事象の報告
11. 効果判定とエンドポイントの定義
12. 統計的事項
13. 倫理的事項
14. モニタリングと監査
15. 特記事項
16. 研究組織
17. 研究結果の発表
18. 参考文献
19. 付表  
説明・同意文書、CRF

# JCOGプロトコールマニュアルの記述

## 12.統計的事項 12.1 主たる解析と判断基準

- 試験の主要な目的を統計的な表現で説明し、統計解析によって検討する仮説を判断基準 (decision criteria) とともに記載する。
- **優越性試験の例**
- 本試験の主たる解析の目的は、標準治療群であるA群 (〇〇療法) に対し、試験治療群であるB群 (XX療法) が、primary endpoint である全生存期間において有意に上回るかどうかを検証することである。主たる解析における両群の全生存期間が等しいという帰無仮説の検定は、グループでの検討を経て決定した全適格例を対象に、割付調整因子を用いた層別ログランク検定により行う。参考として全登録例を対象とした解析も行う。
- 試験治療群が標準治療群に劣っている時は統計的に有意かどうかは関心事ではないため検定は片側検定を行う。試験全体の有意水準は片側2.5%とする。
- 試験治療群が標準治療群を統計学的に有意に上回った場合、試験治療である〇〇療法がより有用な治療法であると結論する。有意に上回らなかった場合は、標準治療である××療法が引き続き有用な治療法であると結論する。
- 累積生存曲線、生存期間中央値、年次生存割合などの推定はKaplan-Meier法を用いて行い、Greenwoodの公式を用いて95%信頼区間を求める。群間比較には施設以外の割付調整因子を層とした層別ログランク検定を用いる。治療効果の推定値として、Coxの比例ハザードモデルを用いて群間の治療効果のハザード比とその95%信頼区間を求める。必要に応じて割付調整因子に加え、偏りが見られた背景因子で調整したCox回帰を行う。

## 12.統計的事項 12.1 主たる解析と判断基準

- **非劣性試験の例**
- 本試験の主たる解析の目的は、標準治療群であるA群（〇〇療法）に対し、試験治療群であるB群（XX療法）が、primary endpointである全生存期間において許容範囲を超えて下回ることが否定できること(非劣性)を検証することである。
- 試験治療群が標準治療群に対し非劣性であることが証明でき、他のエンドポイントで優越性が示された場合には、試験治療である〇〇療法がより有用な治療法であると結論する。非劣性が証明できなかった場合、非劣性であっても他のエンドポイントで優越性が示されなかった場合には、標準治療である××療法が引き続き有用な治療法であると結論する。
- 主たる解析は、グループでの検討を経て決定した全適格例を対象に、施設以外の割付調整因子と治療法を共変量として含めたCox比例ハザードモデルをあてはめ、試験治療群の標準治療群に対するハザード比の信頼区間を用いて検証する。信頼区間の構成はWald法を用いる。非劣性の比較における許容域は、脾温存群の脾摘群に対するハザード比 $<1.16$ とする。参考として全登録例を対象とした解析も行う。
- 非劣性試験であるため、片側検定を行う。試験全体の有意水準はICH-E9に従い、片側2.5%（あるいは単に片側5.0%）とする。累積生存曲線、生存期間中央値、年次生存割合などの推定はKaplan-Meier法を用いて行い、Greenwoodの公式を用いて95%信頼区間を求める。治療効果の推定値として、Coxの比例ハザードモデルを用いて群間の治療効果のハザード比とその95%信頼区間を求める。必要に応じて偏りが見られた背景因子で調整したCox回帰を行う。

# 述

## 12.1 予定登録数・登録期間・追跡期間

- 「2.4.3.臨床的仮説と登録数設定根拠」で示されたパラメータと判断規準に基づき、必要登録数の算出について記述する。いくつかの仮定の下に計算されたサンプルサイズをあげることが望ましい。
- 「2.4.3.臨床的仮説と登録数設定根拠」で示した背景に基づき、A群の○年生存割合を○○%と仮定し、B群のそれが△%上回るかどうかを検出する優越性試験デザインとした場合、登録X年、追跡X年、 $\alpha=5\%$ （片側）、検出力80%として、Schoenfeld & Richterの方法（→引用：Schoenfeld DA, Richter JR. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. Biometrics 1982;38(1):163-170.）を用いて必要登録数を求めると、1群XXX例、両群計XXX例が必要となる。若干の不適合例を見込んで、下記のように設定した。
  - 予定登録数：各群XXX例、両群計XXX例
  - 登録期間：X年、追跡期間：登録終了後X年
- ただし6か月以内の登録期間の延長は、プロトコル改訂手続き不要とする。

## 述

## 2.4.3 臨床的仮説と登録数設定根拠

- 臨床的仮説と判断規準を明記する。第III相試験の場合、優越性試験か非劣性試験かの区別を明記し、非劣性試験の場合は非劣性での判断を行うことの妥当性（例：毒性が軽い、外来治療可能などの有効性以外の試験治療群のメリット）を述べる。

本試験の主たる研究仮説は「試験治療（〇〇療法）群の全生存期間が標準治療（××療法）群に対して有意に上回った場合、〇〇療法をより有用な治療法と判断する」とする。

- 登録数の設定に必要なパラメータ（例：MST、5年生存割合、奏効割合などの過去のデータおよび期待できる上乗せ効果、臨床的に意味があると判断される上乗せ効果など）を根拠と共に示す。登録数の設定は主たる試験の結論を導くために必要な数字であるため、ここで必要なパラメータは、primary endpointに関するものである。ここでは登録数設定の根拠となったパラメータと設定した登録数を記述するのみでよく、統計的考察を含む詳細は「12.2. 予定登録数・登録期間・追跡期間」で述べる。

したがって、本試験では5年生存割合として〇%の上乗せ効果を期待することとし、両群の全生存期間の真の差が5年生存割合で〇%に相当する差より小さければ、臨床的に意義なしと判断することとした。

以上のパラメータを用いて後述（「12.2. 予定登録数・登録期間・追跡期間」参照）する考察に基づいて必要登録数を計算し、登録期間〇年、追跡期間△年として、両群計XXX例を登録予定登録数とした。

## JCOG9912プロトコールの記述

### 10.5.1. 予定登録症例数算出根拠

- JCOG9205 における5-FUci 療法、CPT-11+CDDP 併用療法及びS-1 単独療法の第II 相試験の生存率を下記に示す（再掲）。

| 治療方法             | 6ヶ月 | 12ヶ月 |
|------------------|-----|------|
| 5-Fuci 療法        | 50% | 28%  |
| CPT-11+CDDP 併用療法 | 65% | 40%  |
| S-1 単独療法         | 60% | 35%  |

- 本研究の結果が臨床的に高い価値を持つためには、5年以内に結果を出すことが必要である。
- そこで、登録4年、追跡1年としてそれぞれの対比較を行うために必要なサンプルサイズを計算すると次のようになる。

## CPT-11+CDDP 併用療法の5-FUci 療法に対する優越性の検証

- CPT-11+CDDP 併用療法群で有害事象の発生頻度がより多くなることが予想されるため、CPT-11+CDDP 併用療法が5-FUci 療法に代わって標準的治療となるためには、6ヶ月生存率、12ヶ月生存率でともに10%以上上回ることが必要である。この場合、CPT-11+CDDP 併用療法の5-FUci 療法に対する優越性の検証に必要な1群あたりの症例数は以下のようになる。

| 期待される CPT-11+CDDP 併用療法群<br>6ヶ月生存率 | 5-FUci 療法 6ヶ月生存率 |     |     |     |     |     |
|-----------------------------------|------------------|-----|-----|-----|-----|-----|
|                                   | 50%              |     |     | 55% |     |     |
|                                   | 検出力              |     |     |     |     |     |
|                                   | 70%              | 75% | 80% | 70% | 75% | 80% |
| 60%                               | 140              | 157 | 178 | 534 | 601 | 680 |
| 65%                               | 58               | 65  | 74  | 124 | 140 | 158 |

| 期待される CPT-11+CDDP 併用療法群<br>1年生存率 | 5-FUci 療法 1年生存率 |     |     |     |     |     |
|----------------------------------|-----------------|-----|-----|-----|-----|-----|
|                                  | 25%             |     |     | 30% |     |     |
|                                  | 検出力             |     |     |     |     |     |
|                                  | 70%             | 75% | 80% | 70% | 75% | 80% |
| 35%                              | 168             | 190 | 214 | 703 | 791 | 895 |
| 40%                              | 76              | 86  | 97  | 178 | 200 | 227 |

## S-1 単独療法の5-FUci 療法に対する 非劣性の検証

- S-1 単独療法で有害事象の発生頻度が5-FUci 療法群と比べてそれほど多くなく、非入院生存期間が延長されるという条件の下で、S-1 単独療法が5-FUci 療法に代わって標準的治療なるためには、6ヶ月生存率、12ヶ月生存率とともに5%以上下回ってはならない。この場合、S-1 単独療法の5-FUci 療法に対する非劣性の検証に必要な1群あたりの症例数は以下のようになる。

| 期待される<br>S-1 単独療法群<br>6ヶ月生存率 | 5-FUci 療法 6ヶ月生存率 |     |     |     |     |     |
|------------------------------|------------------|-----|-----|-----|-----|-----|
|                              | 50%              |     |     | 55% |     |     |
|                              | 検出力              |     |     |     |     |     |
|                              | 70%              | 75% | 80% | 70% | 75% | 80% |
| 55%                          | 154              | 174 | 196 | 598 | 673 | 761 |
| 60%                          | 65               | 73  | 83  | 141 | 159 | 180 |

| 期待される<br>S-1 単独療法群<br>1年生存率 | 5-FUci 療法 1年生存率 |     |     |     |     |     |
|-----------------------------|-----------------|-----|-----|-----|-----|-----|
|                             | 25%             |     |     | 30% |     |     |
|                             | 検出力             |     |     |     |     |     |
|                             | 70%             | 75% | 80% | 70% | 75% | 80% |
| 30%                         | 153             | 172 | 195 | 658 | 740 | 837 |
| 35%                         | 76              | 86  | 97  | 170 | 191 | 216 |

- 5-FUci 療法群6ヶ月生存率が50%、55%の場合、6ヶ月生存率で5%下回るということは、ハザード比でそれぞれ、1.15、1.16上回ることに対応し、1年生存率が25%、30%の場合、1年生存率で5%下回るということは、ハザード比でそれぞれ、1.16、1.15上回ることに対応する。そこで、許容域をハザード比 $<1.16$ とする。

70

| 期待される CPT-11+<br>CDDP 併用療法群<br>6ヶ月生存率 | 5-FUci 療法 6ヶ月生存率 |     |     |     |     |     |
|---------------------------------------|------------------|-----|-----|-----|-----|-----|
|                                       | 50%              |     |     | 55% |     |     |
|                                       | 検出力              |     |     |     |     |     |
|                                       | 70%              | 75% | 80% | 70% | 75% | 80% |
| 60%                                   | 140              | 157 | 178 | 534 | 601 | 680 |
| 65%                                   | 58               | 65  | 74  | 124 | 140 | 158 |

| 期待される<br>S-1 単独療法群<br>6ヶ月生存率 | 5-FUci 療法 6ヶ月生存率 |     |     |     |     |     |
|------------------------------|------------------|-----|-----|-----|-----|-----|
|                              | 50%              |     |     | 55% |     |     |
|                              | 検出力              |     |     |     |     |     |
|                              | 70%              | 75% | 80% | 70% | 75% | 80% |
| 55%                          | 154              | 174 | 196 | 598 | 673 | 761 |
| 60%                          | 65               | 73  | 83  | 141 | 159 | 180 |

| 期待される CPT-11+<br>CDDP 併用療法群<br>1年生存率 | 5-FUci 療法 1年生存率 |     |     |     |     |     |
|--------------------------------------|-----------------|-----|-----|-----|-----|-----|
|                                      | 25%             |     |     | 30% |     |     |
|                                      | 検出力             |     |     |     |     |     |
|                                      | 70%             | 75% | 80% | 70% | 75% | 80% |
| 35%                                  | 168             | 190 | 214 | 703 | 791 | 895 |
| 40%                                  | 76              | 86  | 97  | 178 | 200 | 227 |

| 期待される<br>S-1 単独療法群<br>1年生存率 | 5-FUci 療法 1年生存率 |     |     |     |     |     |
|-----------------------------|-----------------|-----|-----|-----|-----|-----|
|                             | 25%             |     |     | 30% |     |     |
|                             | 検出力             |     |     |     |     |     |
|                             | 70%             | 75% | 80% | 70% | 75% | 80% |
| 30%                         | 153             | 172 | 195 | 658 | 740 | 837 |
| 35%                         | 76              | 86  | 97  | 170 | 191 | 216 |

- 表の網掛けの部分を見ると、試験治療群の成績が5-FUci 療法群の成績に対し、優越性で10%、非劣性で5%上回っている場合には、1群あたり150例でほぼ70%以上の検出力を確保できる。PhaseIIの結果により、この成績は十分期待できると考える。そこで、予定登録症例数を3群合わせて450例とする。

## Phase III trialが検証的な研究 であるためには

- ランダム化し、よくコントロールすることによって群間の**比較可能性を確保**する
- 差がないときに、偶然差が出たといわないように **$\alpha$ エラーを厳守**する
- 差があるときに、ちゃんと差がでるように **検出力（症例数）を確保**する

研究の途中で変なことしちゃいけません。

なぜ途中でデータを見る必要があるか

- 全ての安全性の問題を早急に把握

→安全性モニタリング

- 試験進捗に関する問題点を把握

- 登録ペースが遅い
- 背景因子の分布が偏っている
- 脱落やプロトコール逸脱が多い

- 研究目的に対する**結論**がでているか確認

→中間解析

## どうやって中間解析をするか —直感的な方法—

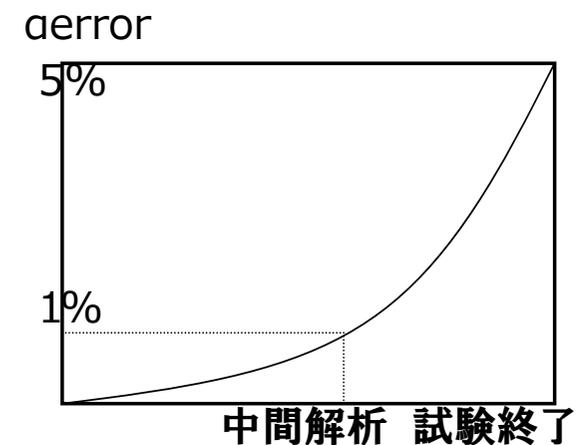
- 両群とも5年生存率0.33, 5年追跡
  - (試験治療と標準治療に差がない)
- 有意かどうか毎回5%で検定する

| 試験全体の $\alpha$ エラー(simulation) |        |          |       |      |      |
|--------------------------------|--------|----------|-------|------|------|
|                                | 中間解析なし | 2.5年めに1回 | 1,3年め | 毎年   | 半年毎  |
| $\alpha$ (%)                   | 5.4    | 7.0      | 7.7   | 13.0 | 15.0 |

- $\alpha$ エラーを守れない!

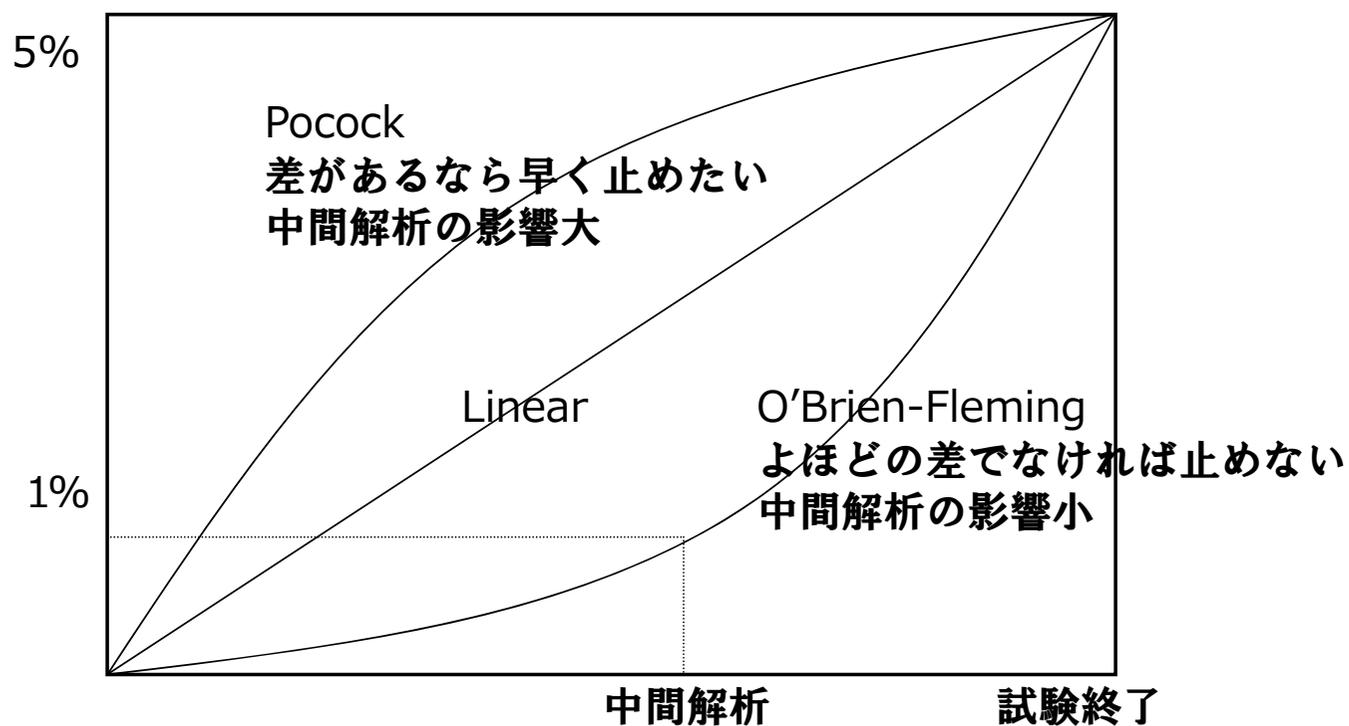
## $\alpha$ 消費関数の利用

- Lan-DeMetsの $\alpha$ 消費関数
  - 少しずつ $\alpha$ を消費することによって試験全体として決められた $\alpha$ を**厳守**
  - 例えば中間解析で1%か、最終解析で4%を下回った場合に「**5%有意**」という



# α消費関数の種類

αerror



## 何回見るかで結論が異なる？

- 最終解析で $p=0.045$ のとき、1回も見なければ有意なのに、中間解析をしたために有意でないという結論になる
- 何度も見ることによって**検出力が低下**
- 救済法：中間解析する分を考慮して、必要症例数を上乘せする

## 検証的な試験における中間解析

- 中間解析を行って
  - 試験治療が有効 or 標準治療が有効という結論が出た場合に試験を中止
- さらに、倫理的な配慮から
  - このまま続けても試験治療が勝てそうにない場合にも試験を中止すべき
    - *consensus*の取れた方法なし
    - 統計的な問題とすべきでないかも

## どんな試験で中間解析が必要か

- 試験を継続するのが倫理的でない場合とは
  - 登録中に結論がでて、これ以上登録を続けるのが倫理的でない
  - 登録終了から最終追跡までに結果がでて、結果を公表しないことが倫理的でない
- つまり、
  - 登録が長期にわたる試験
  - 登録終了から最終追跡までが長い試験

→通常、中間解析で止まることはほとんどない  
元々見込んだよりずっとよいかずっと悪いかの場合だけ

## いつ、誰が、どうやって？

- いつ→中止勧告を出して意味のある時期
  - 登録中、登録後
- 誰が→**比較可能性**を確保するために
  - 研究と独立した機関
  - 効果・安全性評価委員会（グループ代表者）
  - データセンターが実際の解析
- どうやって→ **$\alpha$ エラー厳守、症例数確保**
  - *protocol* で規定

## 独立でないと

- 判断にバイアスが入る可能性
- その後の比較にバイアスが入る可能性
  - *informed consent*が取りにくい
  - 被験者の特徴が変化
  - スタッフの態度の変化
  - エンドポイントの評価に偏り

実際にプロトコールにどう書かれているか見てみましょう

## 述

## 12.3 中間解析と試験の早期中止

- 試験期間の途中において、試験の主たる目的が達成されたかどうかを判断するために主として有効性のエンドポイントの解析を行うことを中間解析と呼ぶ。
- ここでは中間解析の目的、時期、解析方法について記述する。定期モニタリングにおいて安全性の点から試験を中止する場合の規準については、「14.1.定期モニタリング」に記述する。
- 中間解析・最終解析の詳細な手順については、解析を行う前に別途「解析計画書」を作成してもよい。中間解析を行わない場合には理由とともにその旨明記する。
- 非劣性試験(phaseIII)の場合、中間解析の場合の有効中止(非劣性中止)、無効中止の規準について十分に検討する。例えば以下のようなものが考えられる。
  - 有効中止：優越性 or 非劣性 + secondary endpoint の優越性
  - 無効中止：試験治療群の標準治療群に対するハザード比の点推定値が許容ハザード域 (ハザード比 < 1.16) を超えて上回った場合
- 中間解析で登録中止となった場合のその後の追跡期間は、予め登録完了後の追跡期間としてプロトコルで予定していた期間を標準とする。予定していた期間よりも追跡期間を短縮もしくは延長する場合は、効果・安全性評価委員会に改訂申請が必要である。標準以外の設定を用いる場合、プロトコルに明記する

# 述

## 12.3.1 中間解析の目的と時期

- 試験の途中で本試験の主たる目的が達成されたかどうかを判断する目的で2回の中間解析を行う。1回目の中間解析は、登録中に登録を続けることが妥当かどうかを判断する目的で、2回目の中間解析は登録終了後早期に、予定した期間の追跡を続けるかどうかを判断する目的で行う。いずれの場合も試験の主たる目的が達成されていると判断された場合は試験を中止し、すみやかに試験結果を学会および論文にて公表する。
- 1回目の中間解析は、予定登録数の半数の登録が得られた時点以降に問い合わせを行う最初の定期モニタリングのデータを用いて行い、2回目の中間解析は、登録が終了し、すべての登録患者のプロトコール治療が終了する時期を目途に、データセンターと研究事務局で相談した上で適切と思われる時期の定期モニタリングに合わせて行う。
- 原則として中間解析中も登録は停止しない。

## 述

## 12.3.2 中間解析の方法

- 中間解析はデータセンターが行う。試験全体の $\alpha$ エラーを5%に保つために、中間解析と最終解析における検定の多重性をLan & DeMetsの $\alpha$ 消費関数を用いて調整し、群間の生存期間の差について統計学的有意性を調べる。 $\alpha$ 消費関数として、O'Brien & Fleming タイプを用いる (→引用: Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70(3):659-663.)。中間解析の詳細について、データセンターの当該グループ担当統計スタッフは、中間解析の時点までに解析計画書を作成する。実際の中間解析は、当該グループ担当ではない統計スタッフが行い、中間解析レポートを作成する。中間解析において、B群の生存期間がA群のそれを上回り、層別ログランク検定のp値が上記方法により規定された水準を下回った場合、統計的に有意と判断し、原則として試験を中止する。B群の生存曲線がA群のそれを下回っている場合には、検定による判断を行わず、総合的に試験中止の可否を検討することとする。

## 述

## 12.3.3 中間解析結果の報告と審査

- 中間解析結果は中間解析レポートとしてデータセンターより効果・安全性評価委員会に提出され、試験継続の可否および結果公表の可否について審査を受ける。効果・安全性評価委員会は、会議により試験継続の可否を検討し、審査結果に基づいて研究代表者またはグループ代表者に試験継続の可否および結果公表の可否を勧告する。
- ただし、効果・安全性評価委員会委員のうち、当該グループのメンバーは審査には加わらない。また、中間解析の結果により効果・安全性評価委員会から本試験の中止の勧告が出されない限り、最終追跡が終了するまで、本試験の研究代表者、研究事務局、参加施設の研究者、グループ代表者、グループ事務局は中間解析結果を知ることができない。
- 中間解析レポートの審査により、効果・安全性評価委員会より試験の全部または一部について中止または変更の勧告がなされた場合、研究代表者およびグループ代表者は勧告内容を検討し、試験の中止または一部の変更を行うか否かを決定する。
- 試験の中止または試験の一部変更を行う場合には、研究代表者およびグループ代表者は効果・安全性評価委員会に連名で「試験中止許可願い」または「プロトコル改訂願い」を文書で提出する。効果・安全性評価委員会の承認を経て研究代表者は試験を中止または試験の一部を変更することができる。
- 研究代表者およびグループ代表者は効果・安全性評価委員会の勧告内容に異議申し立てができるが、効果・安全性評価委員会との間で意見の調整ができなかった場合、最終的にはJCOG代表者の指示に従う。
- 試験中止となった場合、その後の追跡期間は最終登録から5年間とする。
- 中間解析により試験中止となった場合、その中間解析が本試験の主たる解析となる。データセンターは、研究代表者、研究事務局と協力して、当該中間解析結果を中心に、不完全データの補完や結果の公表に必要な解析を行い、すみやかに「主たる解析レポート」を作成してグループと効果・安全性評価委員会に提出する。

## まとめ

- **プロトコールは情報の宝庫**
  - 手順に関する情報のみでなく、研究の背景やrationale、研究の論理など
- **統計的記述は研究のrationaleそのもの**
  - 統計的記述は統計家だけでは書けない
  - 統計的記述の論理を追うことで研究のrationaleや科学性を確認できる
- **研究計画と異なることは研究の科学性を貶める**
  - 具体的に書いてあるのがいい研究計画書
  - 計画書の中身を理解することが、なぜそれが必要かを理解することにつながる

# 臨床研究入門(略称：ICRweb)

## http://www.icrweb.jp

Welcome to ICRweb  
ICR臨床研究入門へようこそ

ようこそ！  
ユーザー情報 | ログアウト

ICRwebは、臨床研究に関する倫理指針、ガイドライン集、ICRwebのeラーニング教材など、臨床研究に携わるすべての人を対象としています。ICRwebのeラーニング教材を利用するには、ICRwebのeラーニング教材に登録が必要です。(ユーザー登録へ)

**指針の教育義務  
に対応、修了証  
も発行している**

E-LEARNING COURSE

- 臨床研究入門 初級編
- 臨床研究入門 中級編
- 被験者保護
- CRC関連
- 医師主導治験関連
- その他の教育プログラム

WHAT'S NEW !

- 「臨床研究入門初級編」に、「総合テスト」を追加しました 2009年01月29日
- 「臨床研究入門初級編」に、「総合テスト」を追加しました。(続きを読む...)
- 指針・ガイドライン集のコーナーを更新しました 2009年01月08日
- 指針・ガイドライン集の研究倫理指針のコーナーに、「臨床研究に関する倫理指針質疑応答集(Q&A)の周知について」(平成20年12月26日付医政研究第1226001号厚生労働省医政局研究開発振興課長通知)を追加掲載しました。(続きを読む...)
- 2008年10月に修正されたヘルシンキ宣言へのリンクを規制・ガイドライン集に追加しました 2008年12月25日
- ヘルシンキ宣言は、2008年10月に開催されたWMAソウル総会で大幅に修正されました。(続きを読む...)

サイト内検索

検索

検索オプション

アクセス情報

|     |       |
|-----|-------|
| 今日: | 129   |
| 昨日: | 189   |
| 総計: | 33339 |

登録ユーザー数:  
1579名

「臨床研究に関する倫理指針」(2008年改正)では、研究者等に対する教育が義務化された。

同時に発出された医政局長通知で本サイトが紹介されており、既に10000名以上が登録、5000名以上に修了証発行