

がん臨床試験の計画に必要な統計の基礎知識 ～ランダム化とサンプルサイズ～

研究支援センター 生物統計部
/ 東病院臨床研究支援部門

若林 将史

2021. 12.5 (日)

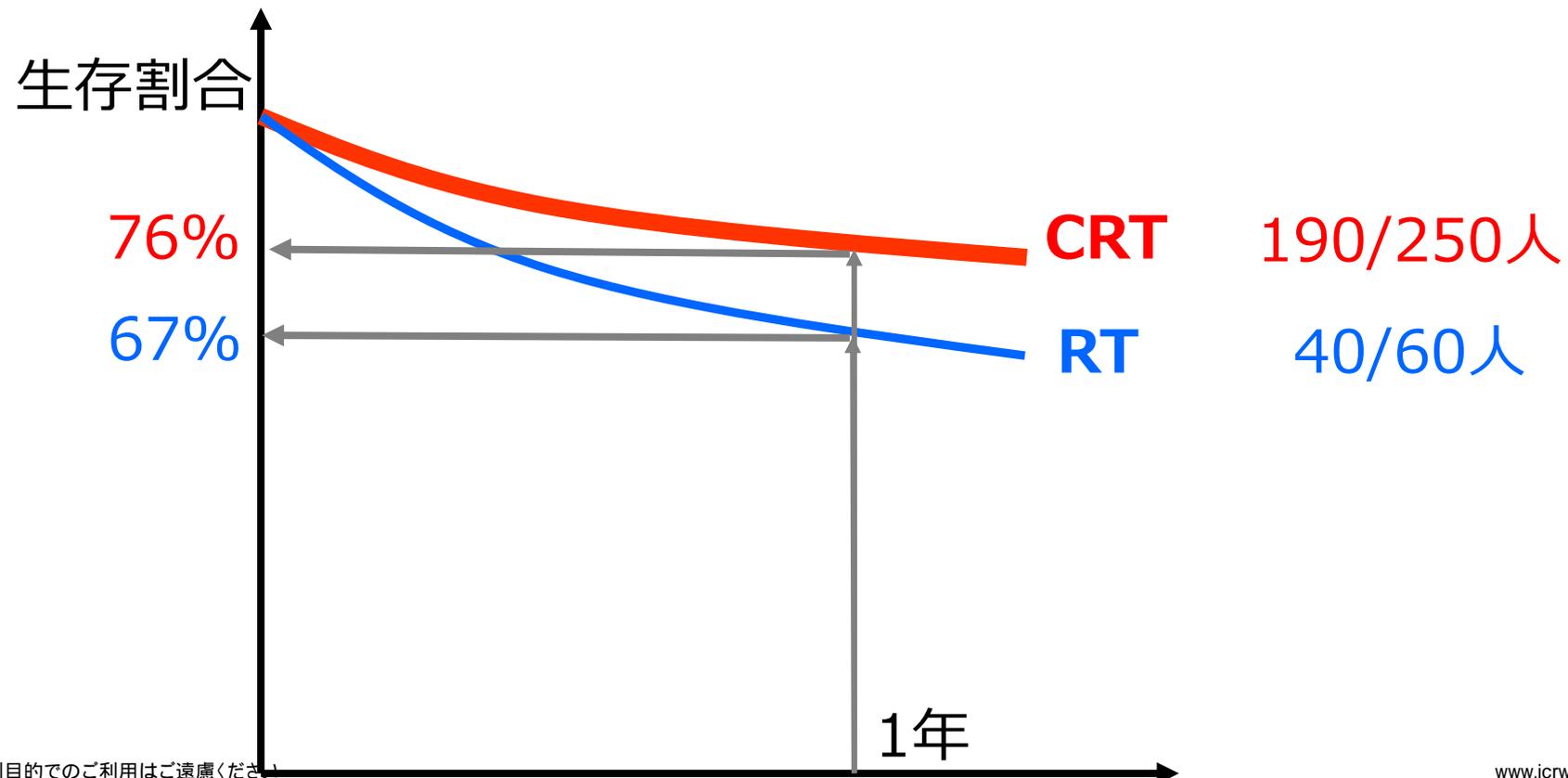
講義の目標

- **ランダム化**はなぜ必要か？
 - 交絡とランダム化
- 臨床家としてサンプルサイズ計算に必要なこと何か？
 - 統計家に相談する際に何を考えておくべきか
 - 統計家が何を考えているのか
- サンプルサイズ計算を理解することで、試験デザインをより深く理解する
 - 優越性試験と非劣性試験
 - リスク・ベネフィットバランス

交絡とランダム化

国内の学会で見かける発表

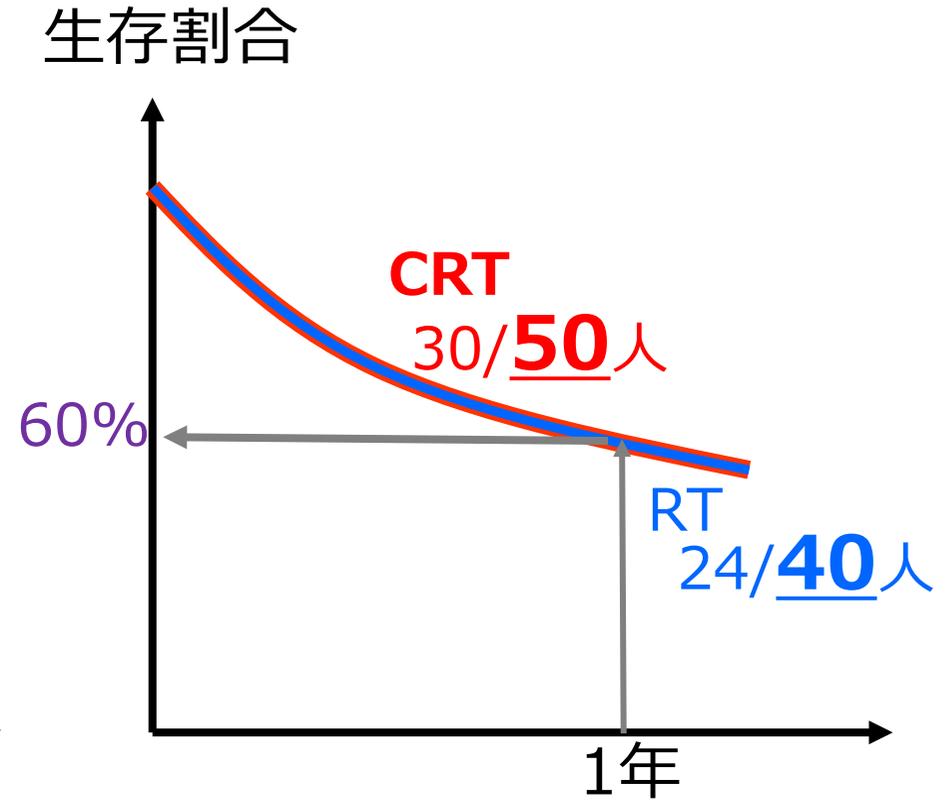
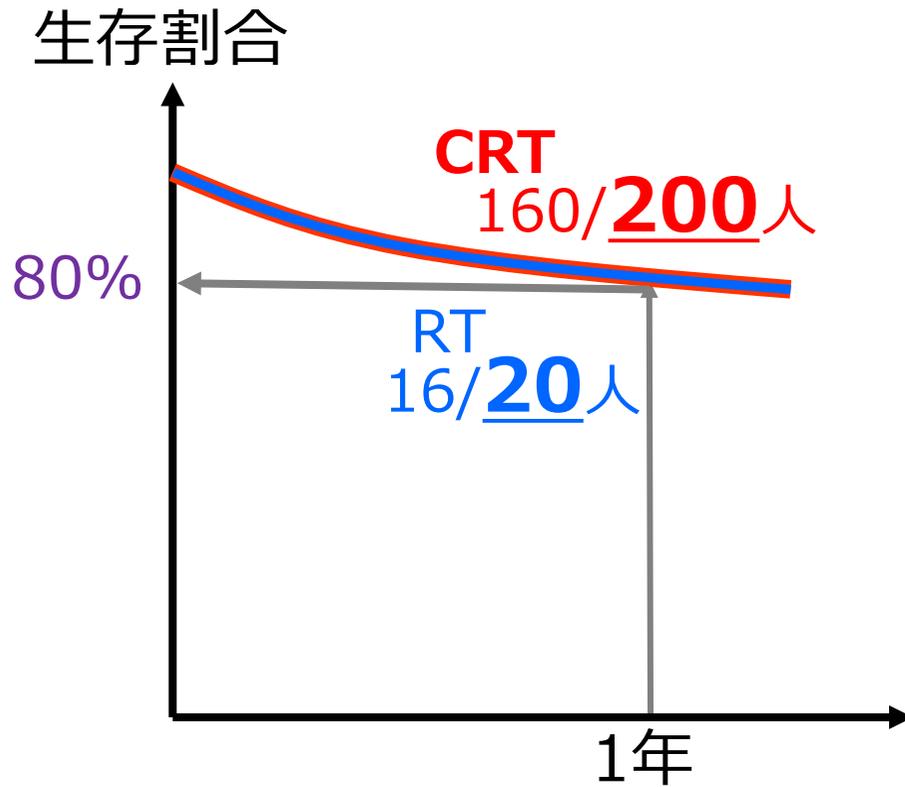
- ある病院の適格規準●●を満たす患者を**CRT(化学放射線療法)群**(250例)と**RT(放射線療法)群**(60例)に分けて後方視的(いわゆるレトロ)に予後に関する検討を行った。
- **CRT群**は**RT群**と比較し予後良好であった。
- この対象に**CRT**をすることが推奨される。



年齢で分けた場合の予後

74歳以下

75歳以上



年齢によらずCRTとRTの予後は変わらない

比較したいのはCRTとRTの違いだから

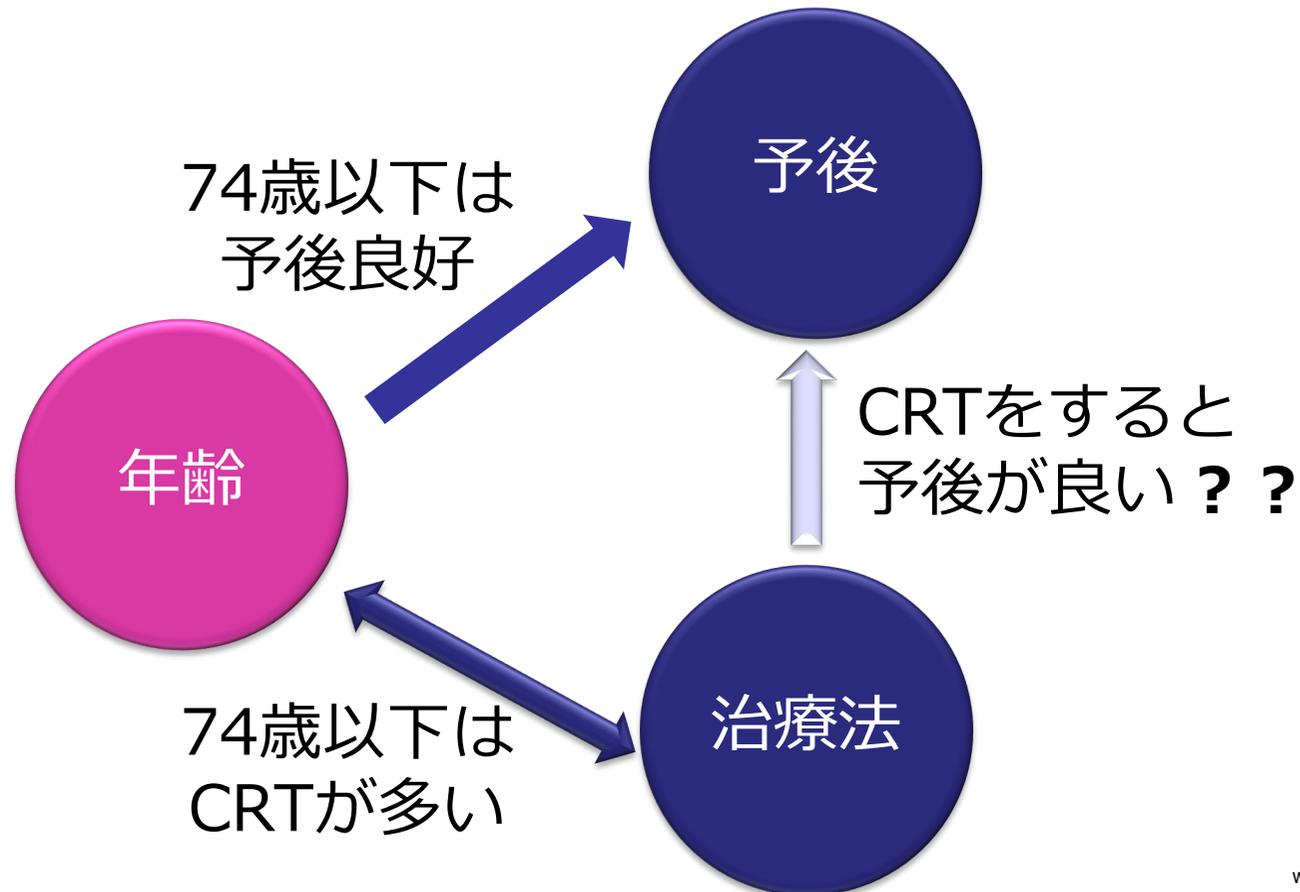
- 治療法以外の予後に影響する要因の条件が同じでなければ、“比較”にならない！！

治療法	74歳以下		75歳以上	合計
CRT	200人 (80%)	>>	50人	250人
RT	20人 (33.3%)	<<	40人	60人

- CRTはRTと比べ「74歳以下」の割合が高い
- 年齢によって予後が異なる(74歳以下は予後良)

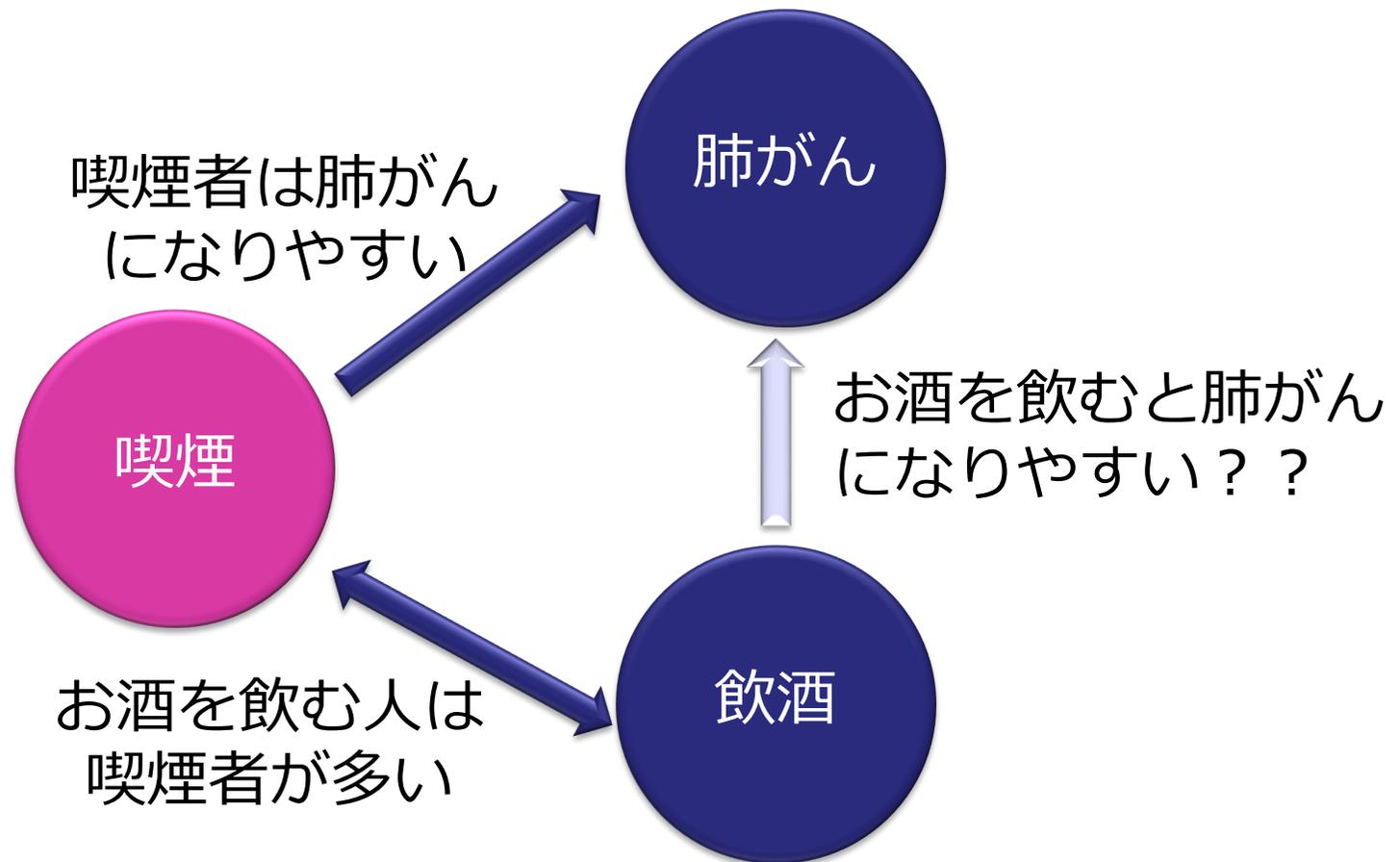
交絡についてのまとめ

- 治療法と予後に関連する第3の因子（年齢）によって見かけ上の関連が生じてしまう現象のこと
 - 交絡を引き起こす因子（=年齢）のことを交絡因子という



その他の交絡の例

- 飲酒が肺がんを起こすわけではないのに、そう見えてしまう
 - この場合、喫煙が**交絡因子**



交絡がないことを保証するには

- 治療群間で予後に関係する背景因子を揃える
 - 年齢
 - stage (がんの進行度)
 - Performance Status (全身状態)
 - その他...

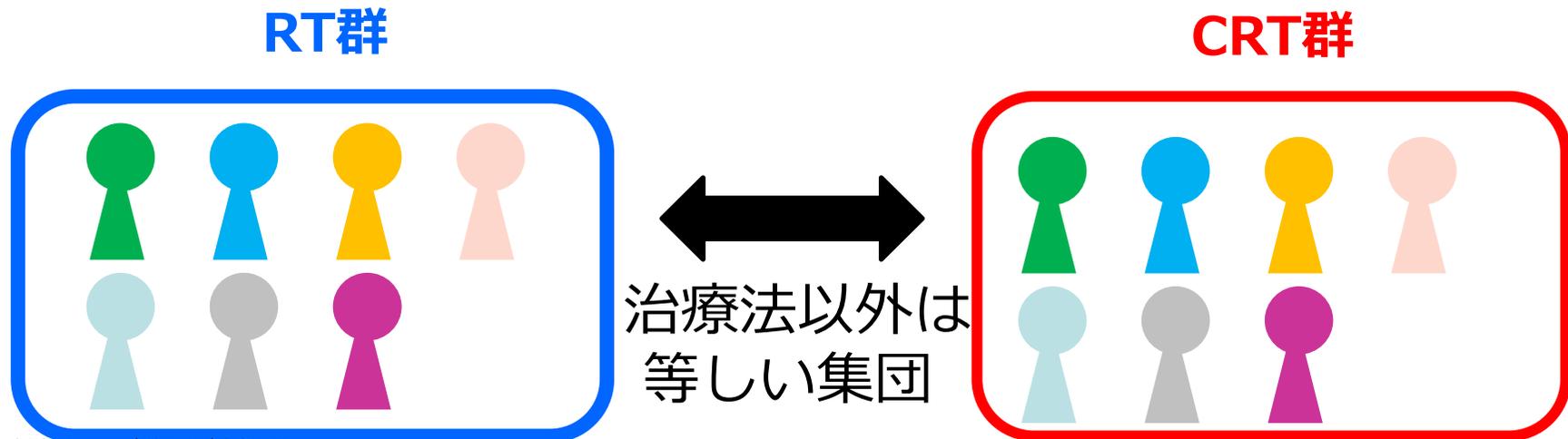
未知のものも含めて因子がたくさんあるために
全てを考慮できない



患者が受ける治療法を**ランダム**に決める

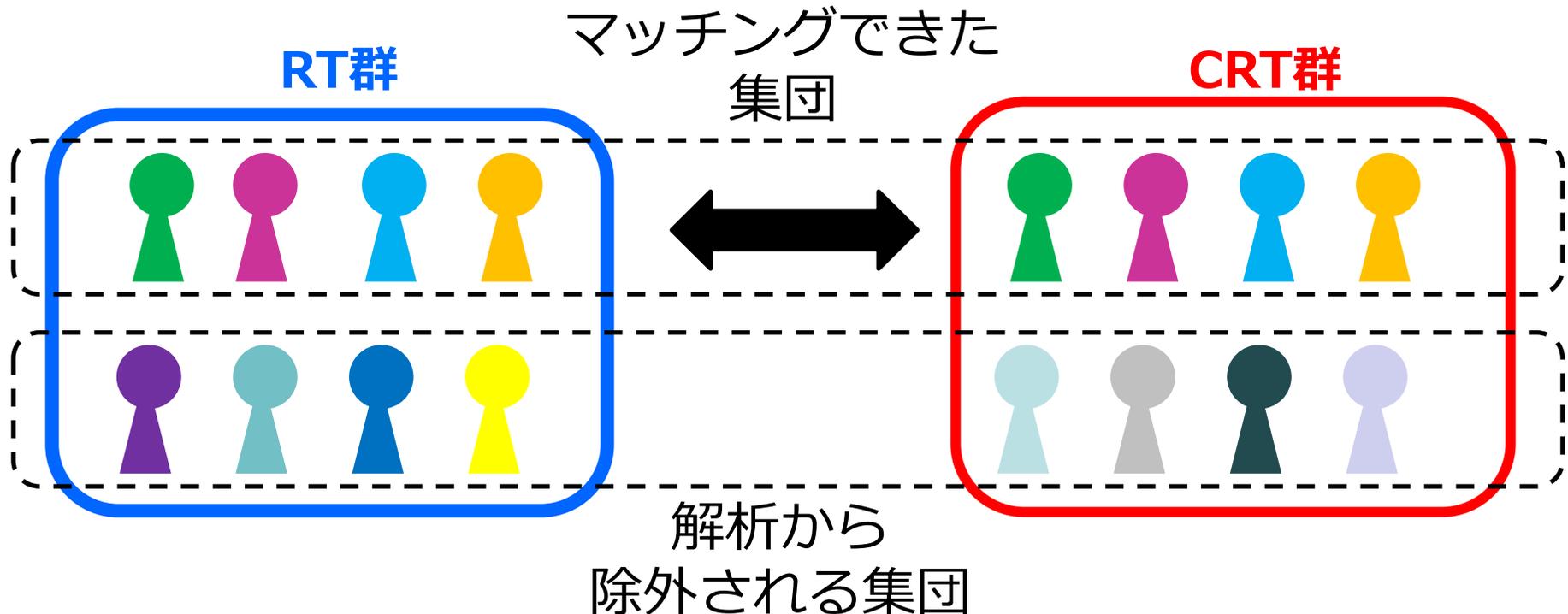
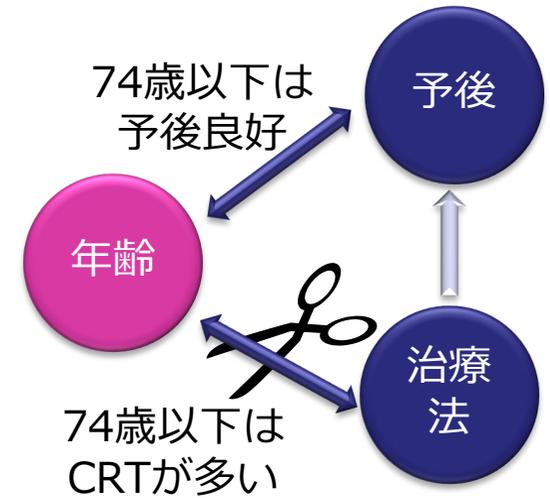
ランダム化 randomization

- 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群に患者を割り付ける
- 予見による患者選択の偏りの防止
 - 状態の良い患者は**試験治療(CRT群)**に割り付けられやすくなる、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



試験計画段階で対処する方法

- マッチング
 - アウトカムに強い影響を与えると考えられる背景因子が一致する患者どうしを合わせる



交絡への対処法

- 試験計画段階（ or データ集積）

- ランダム化(randomization)

- マッチング(matching)

喫煙有無等、倫理的な側面からランダム化を伴う介入研究の実施が難しい場合がある

背景がマッチしない患者集団が解析対象から除外されてしまう（選択バイアス？）

- 解析段階（本講義では割愛）

- サブグループ解析(subgroup analysis)

- サブグループ毎に治療効果を見る

- 層別解析(stratified analysis)

- サブグループ毎の結果を統合(重み付き平均)して、1つのp値、1つの治療効果を求める

- モデルを用いた多変量解析(multivariate analysis)

- ロジスティック回帰やCox回帰など

- 傾向スコア

- ある患者がある治療に割り付けられる確率を求め調整する

サンプルサイズ計算法の考え方

ICRweb 2021年11月上旬時点で会員数約150,000人の臨床研究 e-ラーニング サイト [icrweb.jp]



臨床研究に携わる人のeラーニングサイト

ICR臨床研究入門

よくあるご質問 修了証発行について お問い合わせ サイトポリシー サイトマップ

講座一覧

規制ガイドライン

トリセツ

求人情報

セミナー案内

ICRwebのご紹介

ログイン

ユーザー登録

ICRwebのスマートフォンアプリをご利用ください

動画コンテンツをオフラインでも学習可能なように、ダウンロード型にしてあります。
ダウンロードした動画は通信環境が無くてもご覧頂けます。
ご自身のスタイルに合わせて、自由にご利用ください。

詳細はこちら



ユーザーID

パスワード

ログイン

営利目的でのご利用はご遠慮ください

【重要】Android版アプリをお使いのユーザー様へ

現在、Android版アプリで講義情報が取得できず、講義がご視聴いただけない現象が発生しております。
ご利用のユーザー様にはご不便をおかけして申し訳ございませんが、PC環境にてご視聴いただけますよう、お願い申し上げます。

www.icrweb.jp



• App Store or Google playで“ICRweb”と検索

• インターネット環境でなくとも、

- 講義動画視聴
- サンプルサイズ計算
- 指針等の閲覧

など可能！

ICR臨床研究入門



講座一覧

ICRからのお知らせ

履修状況をPC版へ反映

[PC版サイトはこちら](#)

「ツール」を クリック！



HOME



講座一覧



DL済一覧



ツール



アプリについて



HOME



講座一覧



DL済一覧



ツール



アプリについて

ツール

研究倫理指針

サンプルサイズ計算

p値計算

検定・推定

2群の生存期間の比較（優越性、生存割合に基づく）

2群の生存期間の比較（非劣性、生存割合に基づく）

2群の生存期間の比較（優越性、MSTに基づく）

2群の生存期間の比較（非劣性、MSTに基づく）

2群の生存期間の比較（優越性、ハザードに基づく）

2群の生存期間の比較（非劣性、ハザードに基づく）

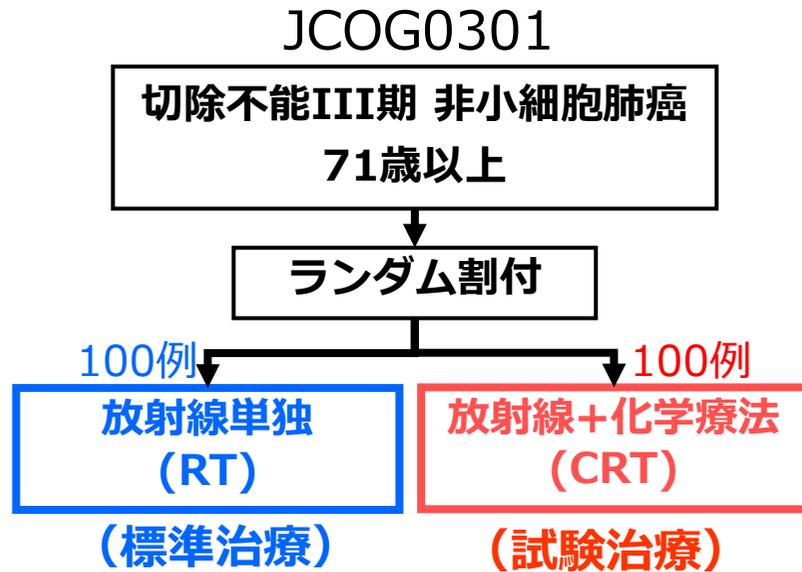
割合の閾値・期待値

2群の平均値の差の比較（優越性、等分散を仮定したt検定に基づく）

2群の割合の差の比較（優越性、カイ2乗検定に基づく）

SimonのSelection Design

論文のサンプルサイズ記載例：優越性試験



Statistical analysis

The trial was designed to have 80% power to detect a 5-month difference in median overall survival (15 months in the chemoradiotherapy group and 10 months in the radiotherapy alone group) via a log-rank test with one-sided alpha of 0.05.^{7,10-12} A sample size of 200 patients was planned by the Schoenfeld and Richter method,¹³ with 1.5 years of follow-up after 4 years of patient accrual.

Thoracic radiotherapy with or without daily low-dose carboplatin in elderly patients with non-small-cell lung cancer: a randomised, controlled, phase 3 trial by the Japan Clinical Oncology Group (JCOG0301)

Shinji Atagi, Masaaki Kawahara, Akira Yokoyama, Hiroaki Okamoto, Nobuyuki Yamamoto, Yuichiro Ohe, Toshiyuki Sawa, Satoshi Ishikura, Taro Shibata, Haruhiko Fukuda, Nagahiro Saijo, Tomohide Tamura, on behalf of the Japan Clinical Oncology Group Lung Cancer Study Group

Summary

Background It is unknown whether combined chemoradiotherapy improves overall survival in elderly patients with locally advanced non-small-cell lung cancer (NSCLC). The aim of this study was to assess whether radiotherapy plus carboplatin results in longer survival than radiotherapy alone in elderly patients with NSCLC.

Methods This was a randomised, controlled, phase 3 trial by the Japan Clinical Oncology Group (JCOG0301). Patients older than 70 years with unresectable stage III NSCLC were randomly assigned to chemoradiotherapy (60 Gy plus concurrent low-dose carboplatin [30 mg/m² per day, 5 days a week for 20 days]) or radiotherapy alone, using a minimisation method with biased-coin assignment balancing on Eastern Cooperative Oncology Group (ECOG) performance status (0 vs 1 vs 2), stage (IIIA vs IIIB), and institution. The primary endpoint was overall survival, which was analysed for the eligible population and stratified by ECOG performance status, stage, and institution. The trial was stopped early as a result of the second planned interim analysis. This study is registered with UMIN Clinical Trials Registry, number C000000060, and ClinicalTrials.gov, number NCT00132665.

Findings 200 patients were enrolled from Sept 1, 2003 to May 27, 2010: 100 in the chemoradiotherapy group and 100 in the radiotherapy group. The second planned interim analysis was done 10 months after completion of patient accrual. At this time, median follow-up for censored cases was 19.4 months (IQR 10.3-33.5). In accordance with the prespecified stopping rule, the JCOG data and safety monitoring committee recommended early publication of this trial because the difference in overall survival favoured the chemoradiotherapy group. Median overall survival for the

nd 16.9 months
1 p value=0.0179).
1 the radiotherapy
thrombocytopenia
ents [12.5%] than
icity were similar
hemoradiotherapy

hemoradiotherapy
population.

group—the Japan
JCOG9812)—was
v-dose carboplatin
: on survival than
with unresectable
was stopped early
rred in four of
ere in the chemo-
ly assessed radio-
8 of 45 evaluable
hat were protocol-
e or normal lung
: treatment-related

Correspondence to:
Dr Shinji Atagi, National Hospital
Organization Kinki-chuo Chest
Medical Center,
1180 Nagasone-cho, Kita-ku,
Sakai, Osaka, 591-8555, Japan
s-atagi@kch.hosp.go.jp

サンプルサイズ計算の基本的な考え方

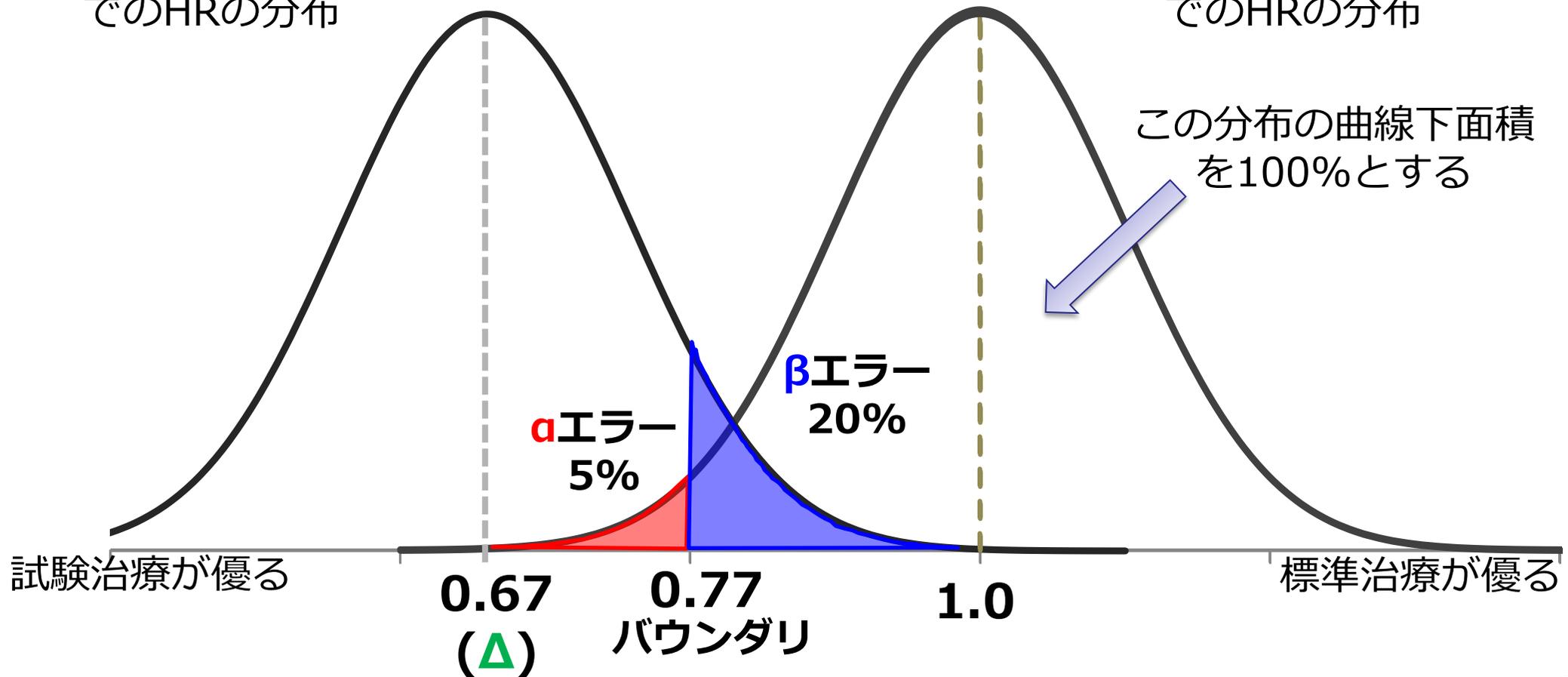
サンプルサイズ計算で求める必要解析対象数は、
以下の条件を満たすように定める

1. 帰無仮説(群間で治療効果に差が無い)が正しい時に誤って新治療が有効だと判断してしまう確率 (**α**) を一定以下に抑える
2. 新治療の真の効果が **Δ** の時(対立仮説が正しい時)に、新治療が有効だと正しく判断する確率である**検出力** (**power [1- β]**) を一定以上にする

α 、 β 、 Δ と得られるHRの分布

対立仮説(HR=0.67)の下
でのHRの分布

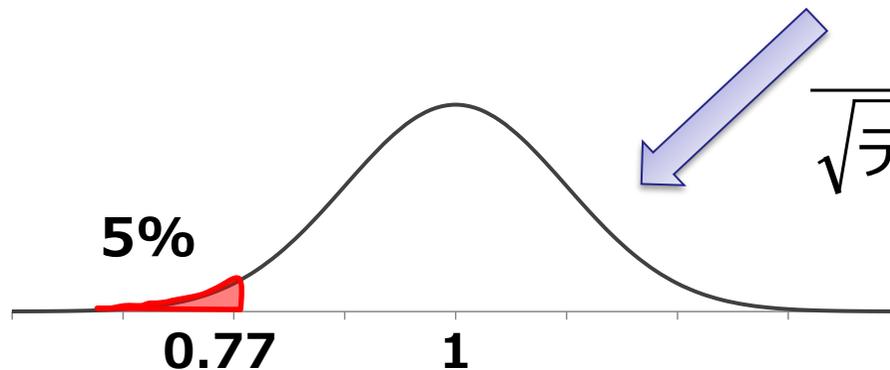
帰無仮説(HR=1)の下
でのHRの分布



サンプルサイズと検定統計量の分布の関係

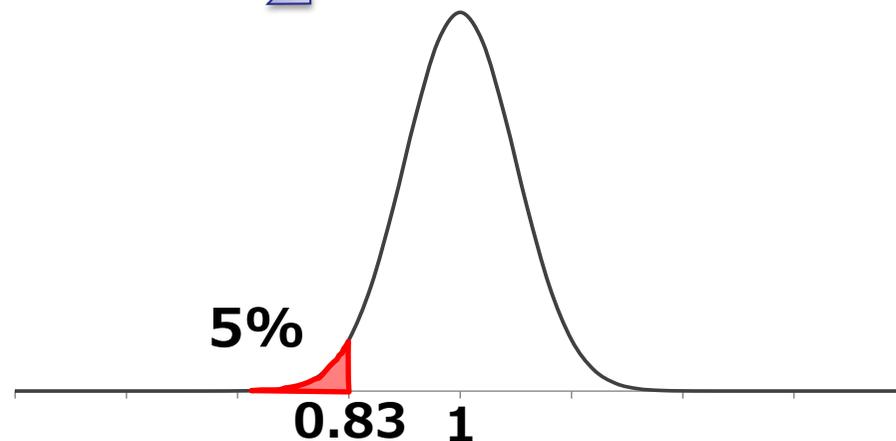
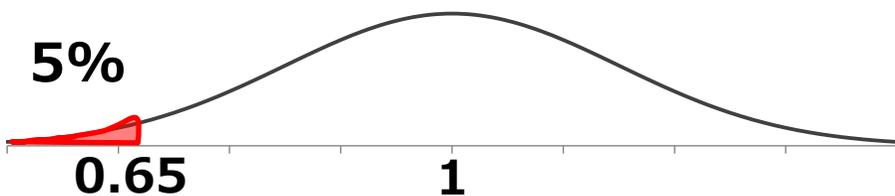
検定統計量の分布を定義する
数式の中にNが含まれている

$$\frac{\Delta}{\sqrt{\text{データのバラツキ}^2/N}}$$



Nが減れば、分布は
ゆるやかになる

Nが増えれば、分布は
きつくなる



サンプルサイズ計算式の骨格

有意水準 α に対応する値を代入
(片側 $\alpha=2.5\%$ ・両側 $\alpha=5\%$ なら1.96、
片側 $\alpha=5\%$ なら1.64)

α 、 β 、 Δ を満たす検定
統計量の分布に
なるようにNを定める

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \text{バラツキ}^2}{\Delta^2}$$

powerに対応する値を代入
(power=80%なら0.84、 power=90%なら1.28)

どのような形式で代入するかは、アウトカムの型
(連続量/2値データ/time-to event) や検定法に依存

各検定法のサンプルサイズ計算式の骨格部分

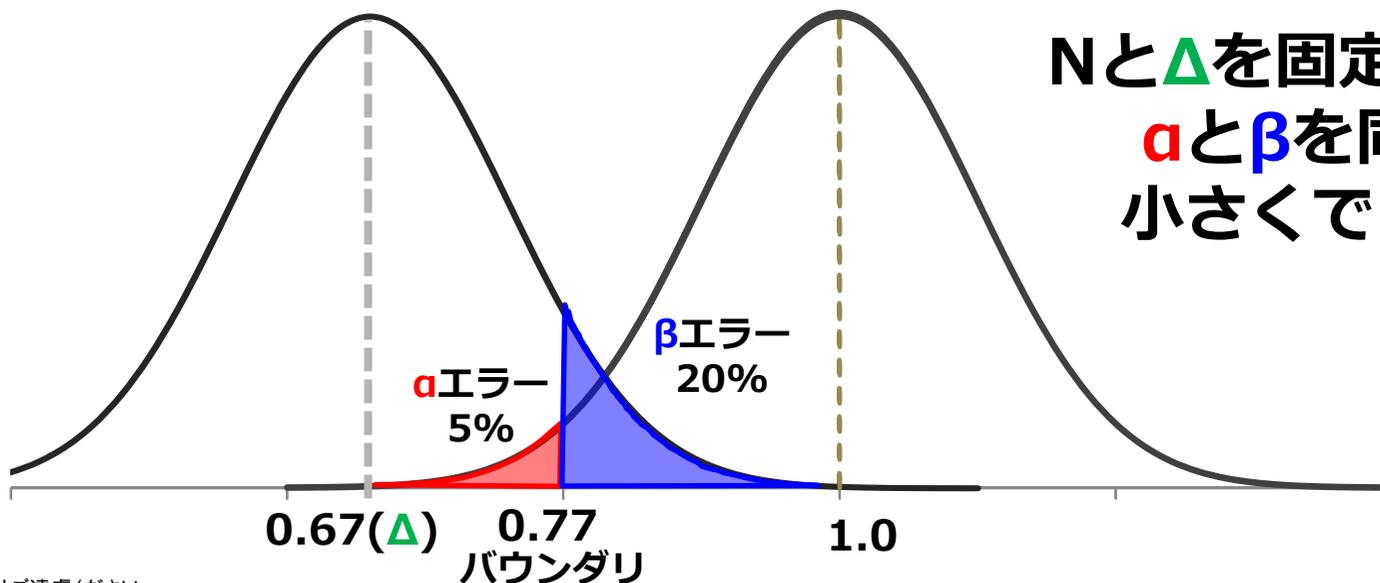
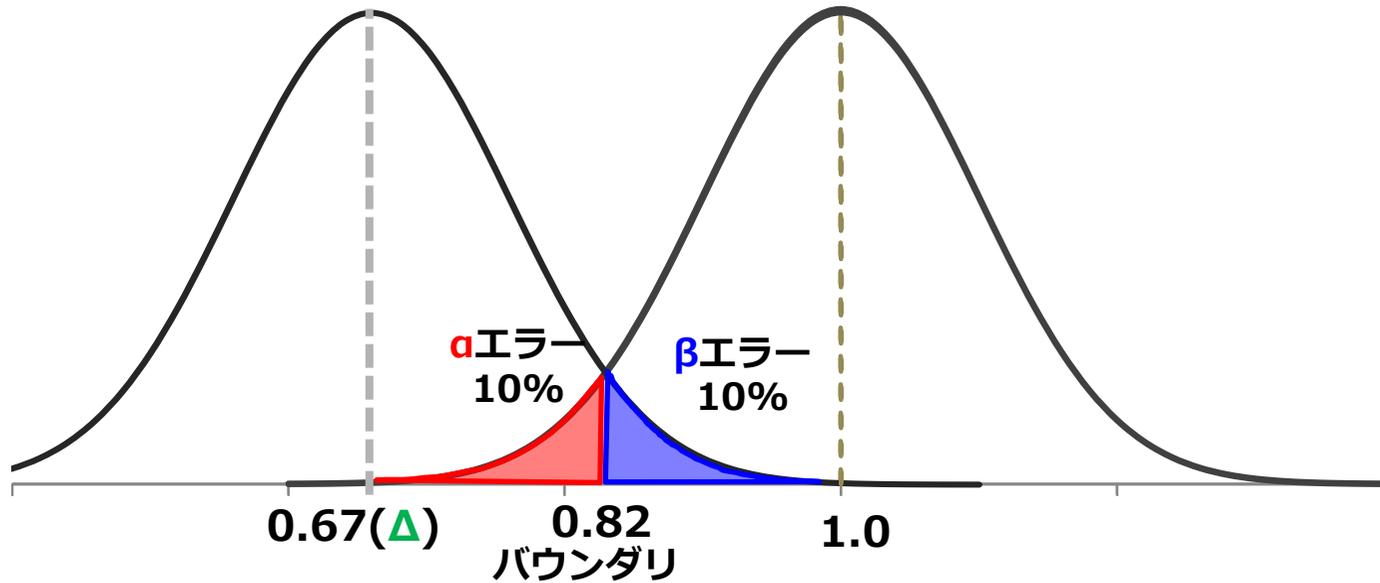
$$t\text{検定} : N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \text{データのバラツキ}^2}{\text{平均値の差}^2}$$

$$\chi^2\text{検定} : N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \left[\begin{array}{l} \text{標準群の割合}(1 - \text{標準群の割合}) \\ + \text{試験群の割合}(1 - \text{試験群の割合}) \end{array} \right]^2}{\text{割合の差}^2}$$

$$\text{log rank検定} : N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \left(\begin{array}{l} \text{予後と登録 / 追跡期間を} \\ \text{考慮した値} \end{array} \right)}{\ln(\text{HR})^2}$$

α と β はトレードオフ

$$N = \frac{(Z_\alpha + Z_\beta)^2 \cdot \text{バラツキ}^2}{\Delta^2}$$

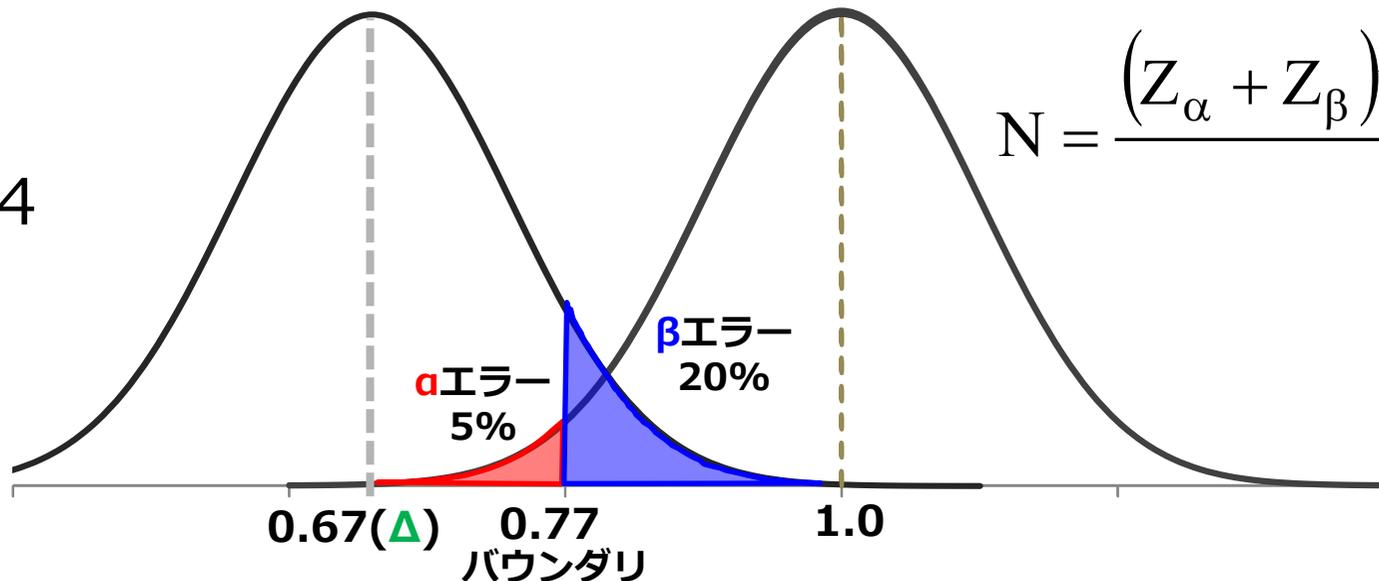


Nと Δ を固定すると、 α と β を同時に小さくできない

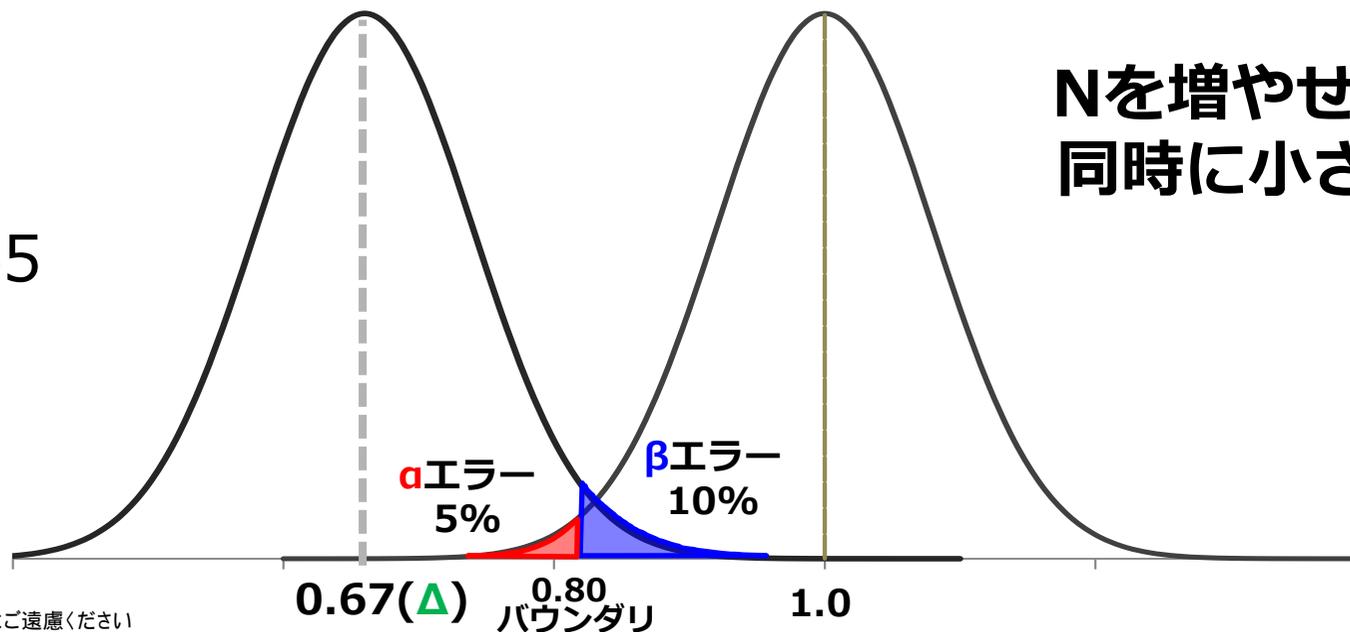
Nを増やせば α を固定してPowerをあげられる

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \text{バラツキ}^2}{\Delta^2}$$

N=174



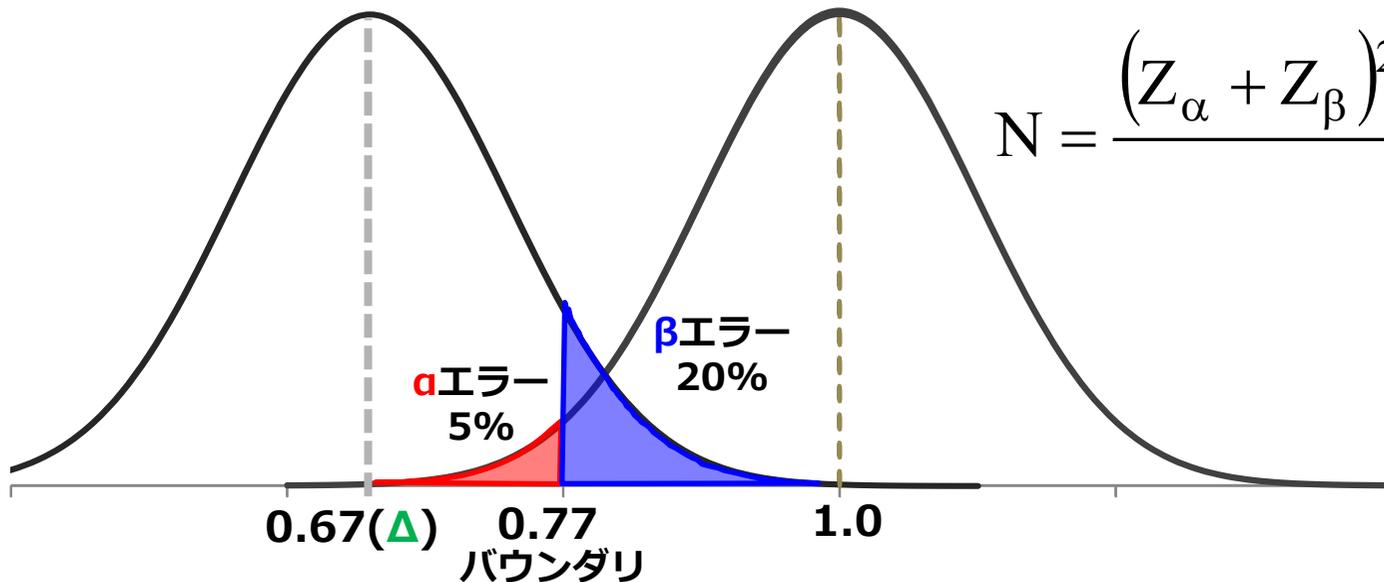
N=245



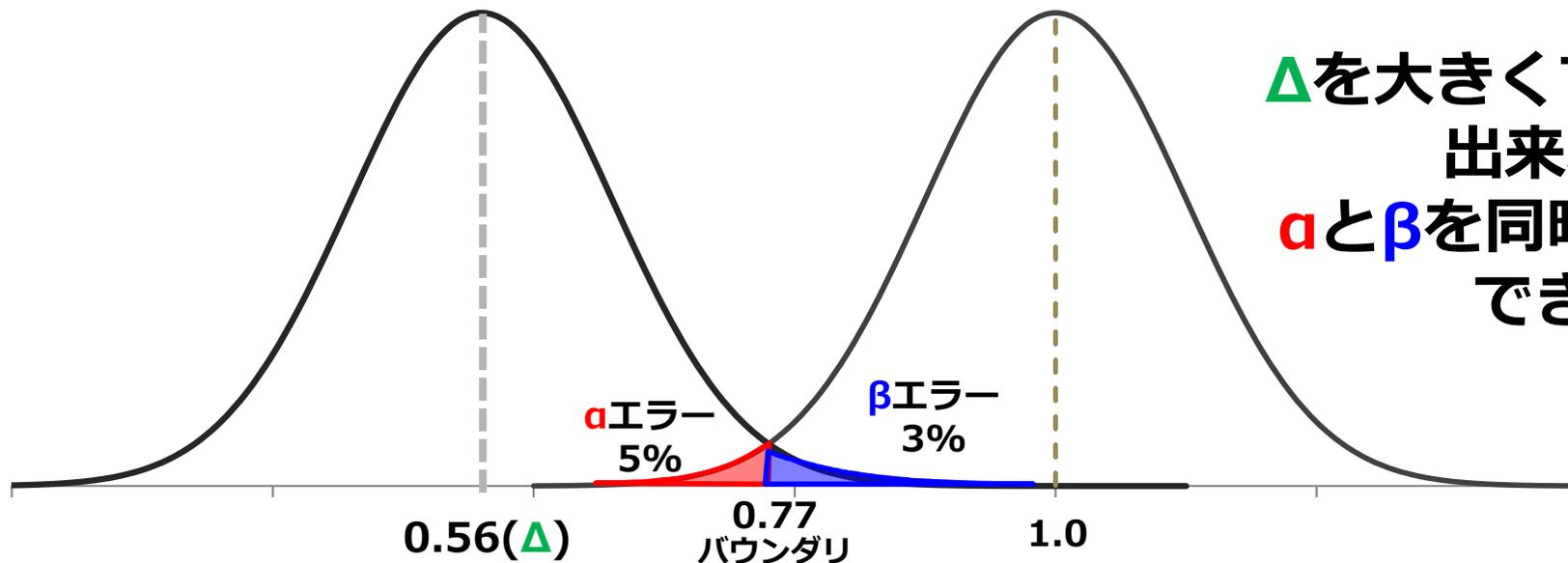
Nを増やせば α と β を同時に小さくできる

Δ を大きく出来れば α を固定してPowerをあげられる

$$N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \text{バラツキ}^2}{\Delta^2}$$



Δ を大きくすることが出来れば α と β を同時に小さくできる



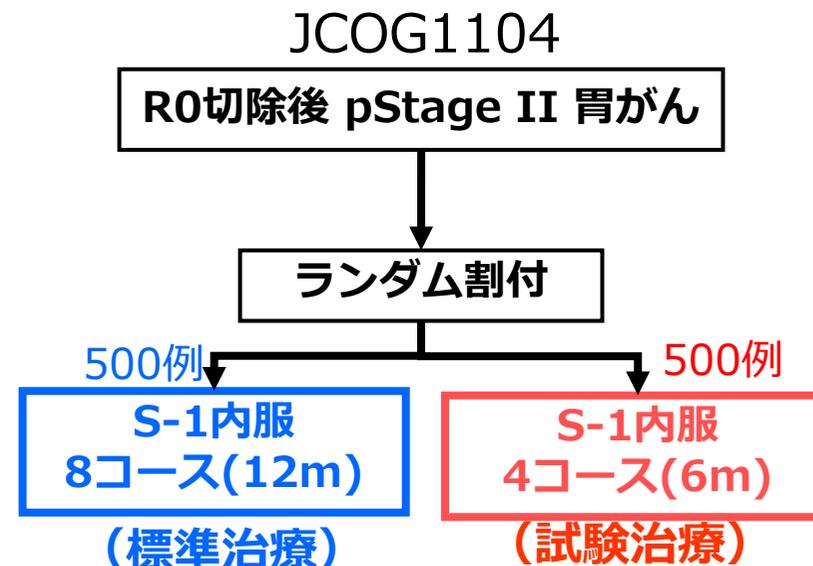
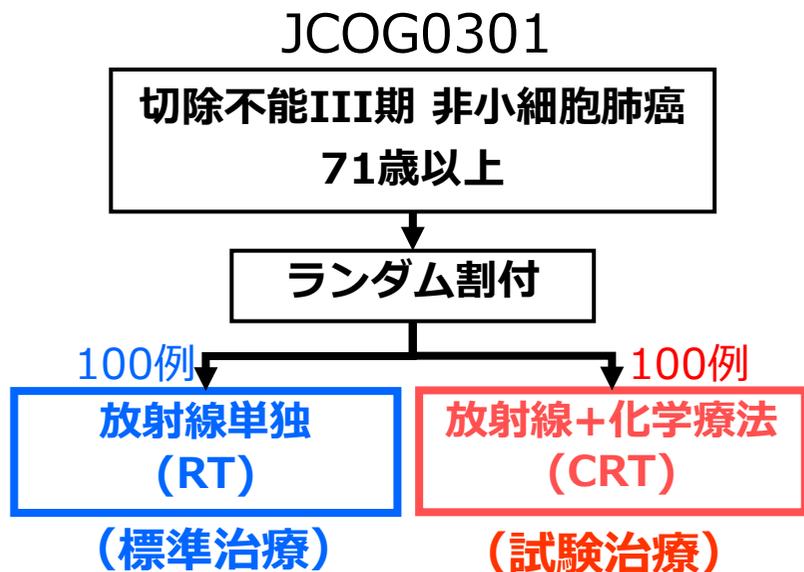
暗記！サンプルサイズ計算に最低限必要なパラメータ

- 治療効果の大きさ： Δ (デルタ)
- バラツキ
- 有意水準： α (アルファ)
 - P値がこれを下回ったら「差がある」と判断する値
- 検出力： $1-\beta$ (1-ベータ, power)
 - 本当に効果のある治療を、正しく効果があると判断する確率

パラメータの設定時に考えること

- デザイン：優越性 or 非劣性
- 有意水準：片側 or 両側、常に5%？
- 検出力
- 治療効果の大きさ（臨床的意義 or 見込み）

優越性試験と非劣性試験



優越性試験

- 試験治療は有効性で上回らなければいけない
- 試験治療は安全性等で標準治療に劣る(**Toxic new**)

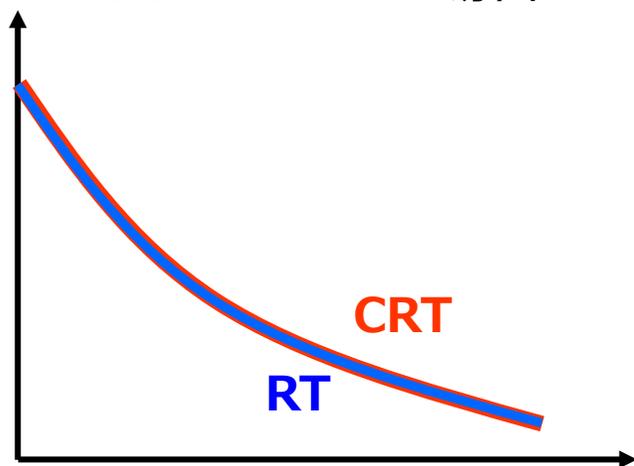
非劣性試験

- 試験治療は有効性で一定以上劣らなければ良い
- 試験治療は安全性等で標準治療に優る(**Less toxic new**)

優越性/非劣性の決め方

- 有効性のエンドポイント（生存曲線）が重なった状況を考えて決める
 - 標準治療を選択する → 優越性試験
 - 試験治療を選択する → 非劣性試験

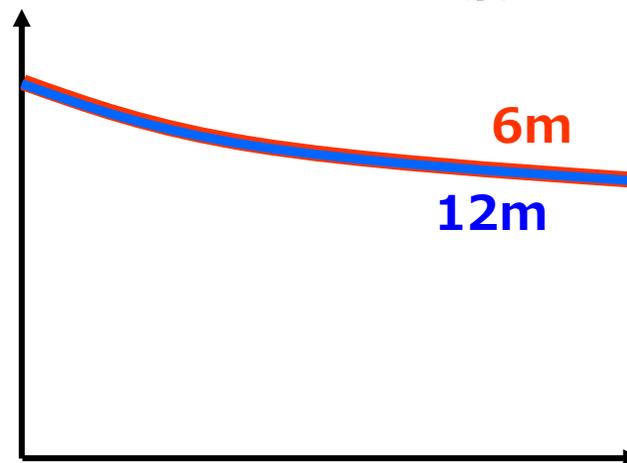
JCOG0301の場合



毒性が強く、治療に手間がかかる
CRT群 (Toxic new)と**RT群**のOSが
同じなら、標準治療は**RT群**
→ 優越性試験

営利目的でのご利用はご遠慮ください

JCOG1104の場合



治療期間が短くなり、安全性で優る**6m**
群 (Less toxic new)と**12m群**のRFS
が同じなら、標準治療は**6m群**
→ 非劣性試験

アプリを使ったサンプルサイズ計算①

JCOG0301のサンプルサイズ設計

- 片側 $\alpha=5\%$
- 検出力=80%
- 登録期間=4年、追跡期間=1.5年
- RT vs CRTのMST : 10m vs 15m (HR=0.67)

パラメータ入力

結果

docomo LTE 9:25 55%

サンプルサイズ計算

サンプルサイズ

検出力

結果表示

2群の生存期間の比較（優越性、生存割合に基づく）

片側検定なら1、両側検定なら2

α (有意水準)

2群の生存期間の比較（非劣性、生存割合に基づく）

検出力

2群の生存期間の比較（優越性、MSTに基づく）

登録期間(年)

フォローアップ期間(年)

2群の生存期間の比較（非劣性、MSTに基づく）

対照群のMST(月)

試験群のMST(月)

1 ハザード比

0.6667

0.05

1群あたりのサンプルサイズ

87

0.8

2群を合わせたサンプルサイズ

174

4

2群を合わせた期待イベント数

151

1.5

ギリギリ有意になる（棄却限界値に対応する）ときのハザード比

0.7652

10

ギリギリ有意になる（棄却限界値に対応する）ときの試験群のMST

13.0683

15

結果表示

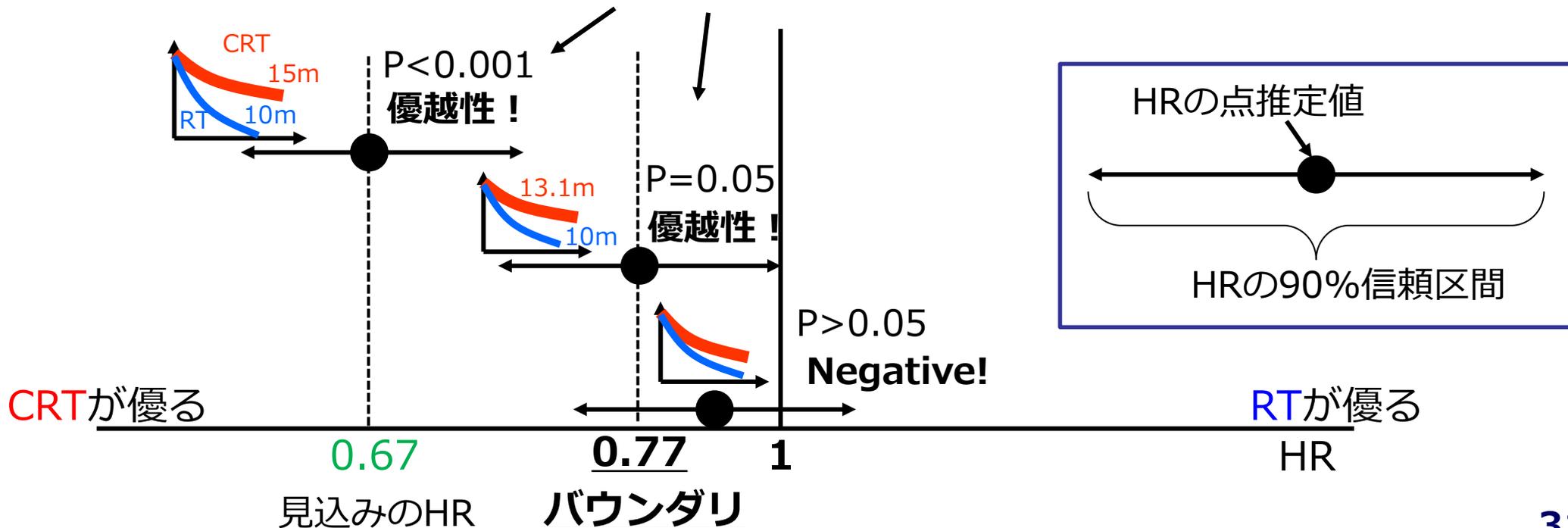
統計的有意差とバウンダリ：優越性試験の場合

- 片側p値 <0.05 =ハザード比(HR)の両側90%CIが帰無仮説(**HR=1**)以下

JCOG0301のサンプルサイズ設計

- 片側 $\alpha=5\%$
- 検出力=80%
- RT vs CRTのMST：10m vs 15m (HR=0.67)
- 必要解析対象数（期待イベント数）：計174例（151）

HR=1ではなく、CRTが上回っていることが示された



アプリを使ったサンプルサイズ計算②

JCOG1104のサンプルサイズ設計

- 片側 $\alpha=5\%$
- 検出力=80%
- 登録期間=5年、追跡期間=3年
- 12m vs 6m の3yRFS : 85% vs 85% (HR=1)
- HRの非劣性マージン : 1.37 (3yRFSで5%に相当)

パラメータ入力

結果

docomo LTE 9:25 55%

サンプルサイズ計算

2群の生存期間の比較 (優越性、生存割合に基づく)

2群の生存期間の比較 (非劣性、生存割合に基づく)

2群の生存期間の比較 (優越性、MSTに基づく)

2群の生存期間の比較 (非劣性、MSTに基づく)

片側検定なら1、両側検定なら2

α (有意水準)

検出力

登録期間(年)

フォローアップ期間(年)

考慮する生存割合の年数(何年時点での生存割合か?)

対照群の生存割合

試験群の生存割合

対照群に与える生存割合のハンデ(対照群から差し引く割合)

1

0.05

0.8

5

3

3

0.85

0.85

0.05

許容ハザード比

1群あたりのサンプルサイズ

2群を合わせたサンプルサイズ

2群を合わせた期待イベント数

ギリギリ有意になる(棄却限界値に対応する)ときのハザード比

ギリギリ有意になる(棄却限界値に対応する)ときの試験群の生存割合

1.3730

482

964

246

1.1133

0.8345

結果表示

営利目的でのご利用はご遠慮ください

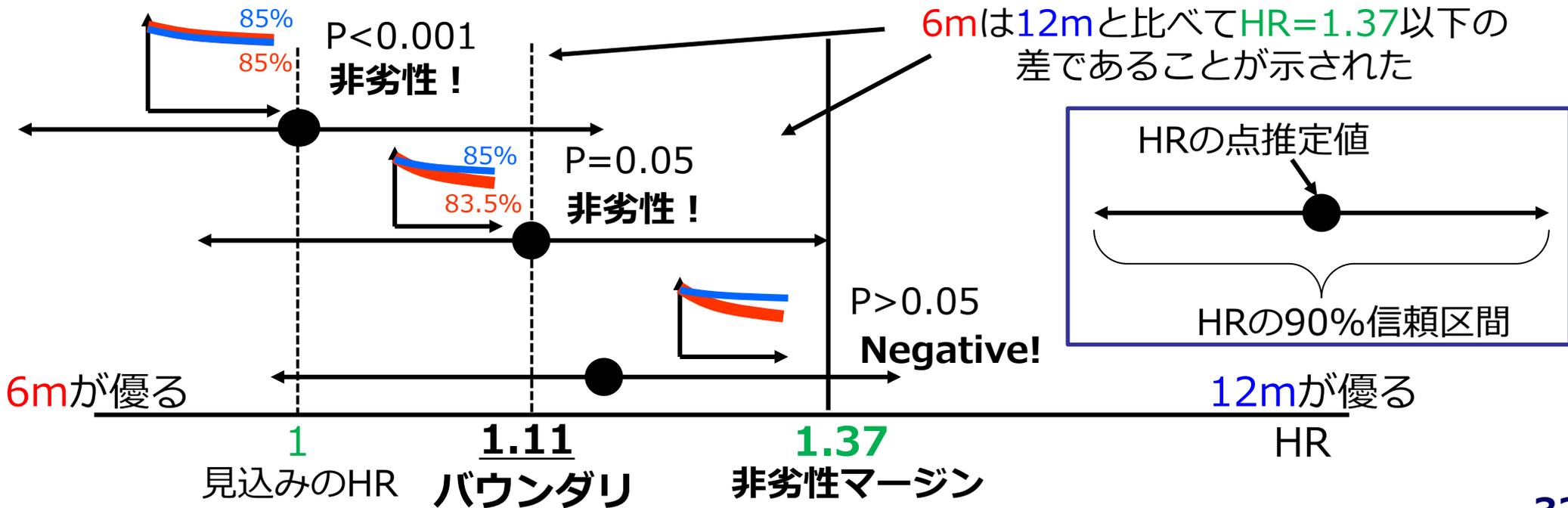
www.icrweb.jp

統計的有意差とバウンダリ：非劣性試験の場合

- 非劣性の片側p値 <0.05 =HRの両側90%CIが帰無仮説(非劣性マージン)以下

JCOG1104のサンプルサイズ設計

- 片側 $\alpha=5\%$
- 検出力=80%
- 12m vs 6m の3yRFS : 85% vs 85% (HR=1)
- HRの非劣性マージン : **1.37** (3yRFSで5%に相当)
- 必要解析対象数 (期待イベント数) : 計964例 (246)

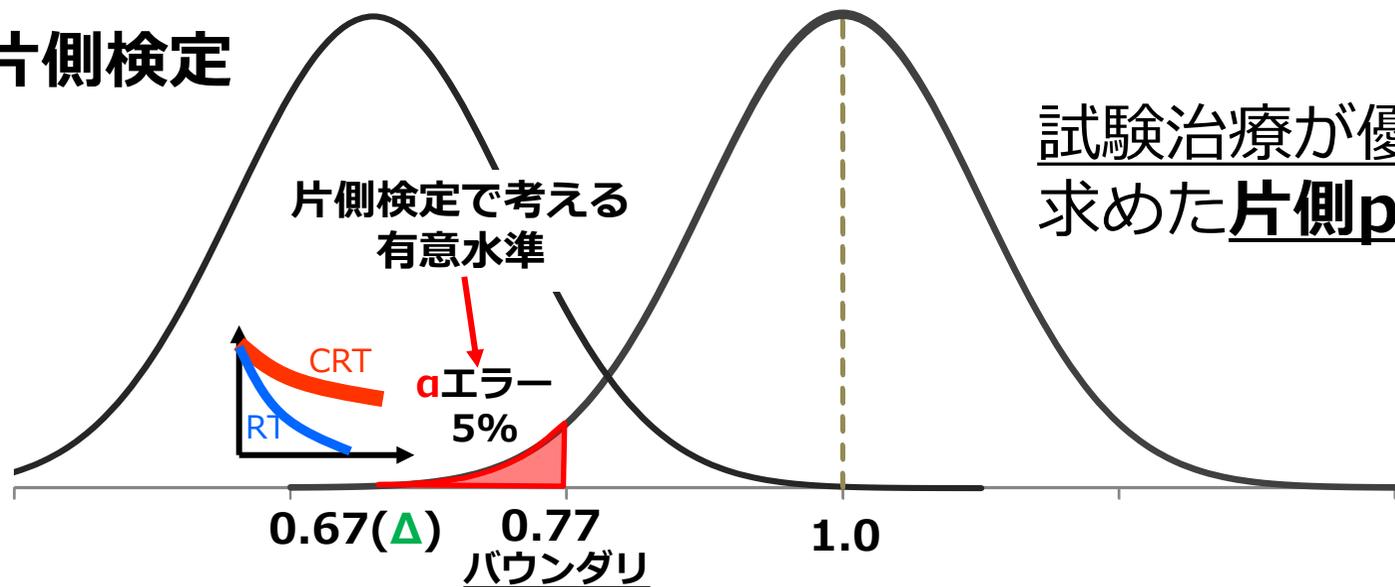


パラメータの設定時に考えること

- デザイン：優越性 or 非劣性
- 有意水準：片側 or 両側、常に5%？
- 検出力
- 治療効果の大きさ（臨床的意義 or 見込み）

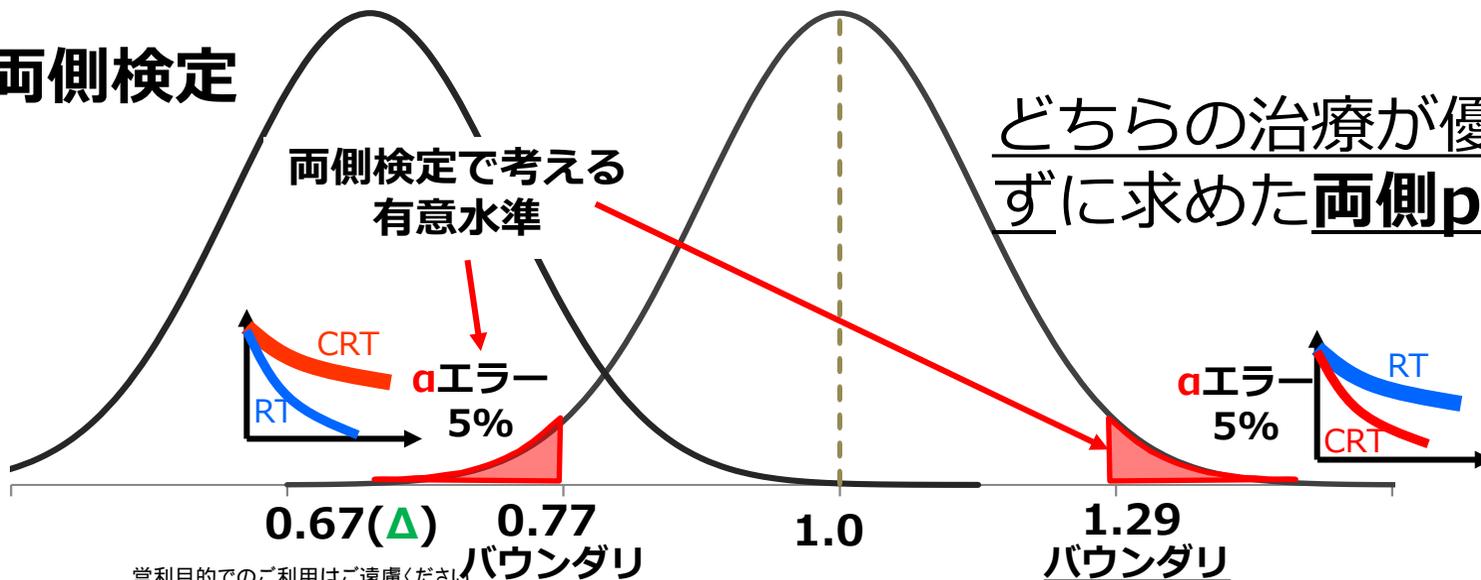
有意水準 α ：片側検定と両側検定

片側検定



試験治療が優る方向だけ考慮して求めた片側p値に基づいて行う

両側検定



どちらの治療が優っているかによらずに求めた両側p値に基づいて行う

片側検定か両側検定を決める2通りの考え方

推奨!

A) 治療法のリスク・ベネフィットを考慮し、意思決定に応じて決める

- JCOG0301(優越性) : RT vs CRTの比較の場合 : 片側検定
 - Toxic newであるCRTはRTに対して優らなければならない
 - CRTがRTに有意に劣ることに興味はなく、中間解析でCRTがRTに有意に劣らなくとも無効中止する
- JCOG1104(非劣性) : 12m vs 6mの比較の場合 : 片側検定
 - Less toxicである6mは12mに対して劣らなければ良い (優っていても良い)
 - 優ってはいけない場合は非劣性ではなく同等性試験 (例: 降圧薬のジェネリック)
- リスク・ベネフィットが同程度の治療法同士の比較の場合 (優越性[差があることを証明するための試験]) : 両側検定
 - どちらの治療法が優ってもよく、有意差がなければどちらの治療法も標準治療

B) 試験治療が標準治療に劣る可能性がある場合は常に両側検定とすべき

- 試験治療が劣ることも証明するために両側検定
- 中間解析で毒性の強い試験治療が負けていても、有意に劣るまで無効中止しない

α は常に5%ではない：状況に応じた α の設定

- 両側 $\alpha=0.05$ or 片側 $\alpha = 0.025$
 - Global standard (**Phase IIIの標準[ICH E9]**)
 - 従うことに理由はいらない
 - α エラーは患者側のリスクなので、出来るだけ小さく
- 片側 $\alpha = 0.05\sim 0.2$
 - **phase II**：間違っても有意になっても次にphase IIIで検証すればよい
 - 早期の段階で、誤って有望な治療法を捨ててしまうリスクを考慮
- 両側 $\alpha = 0.1\sim 0.2$
 - **背景因子**の群間比較をするときの偏りの評価
 - $P > 0.05$ でも endpoint に影響を及ぼしうる → 一律「N.S.」は誤り
 - レトロスペクティブな研究等の予後因子/予測因子の探索
 - サンプルサイズやイベントが少ない場合

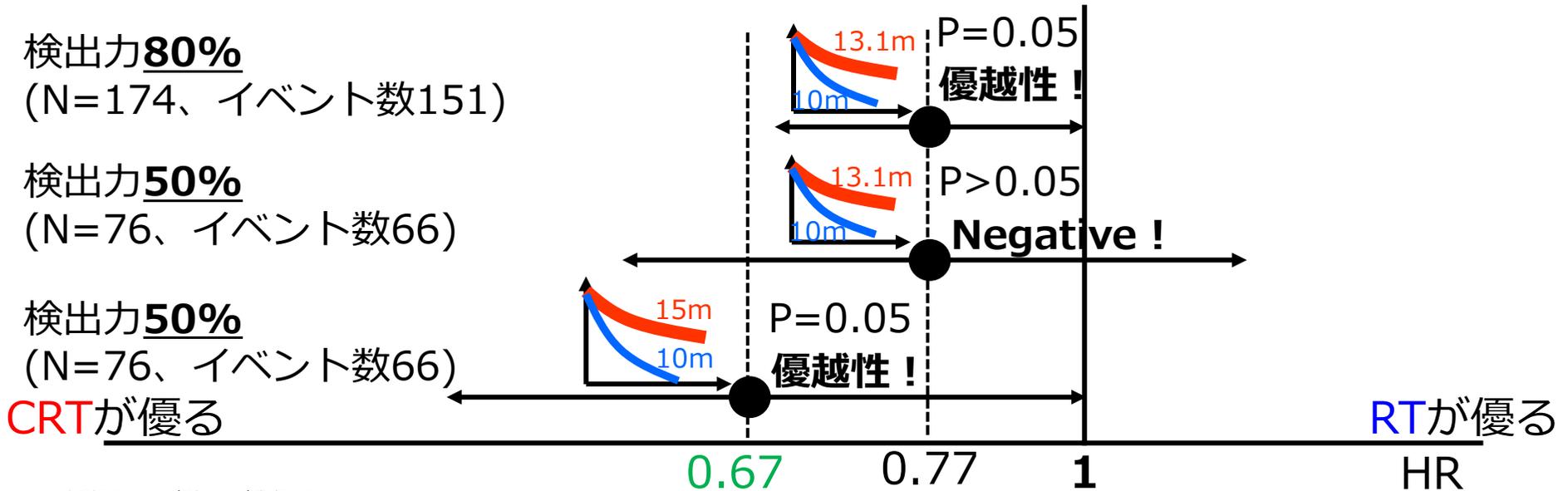
パラメータの設定時に考えること

- デザイン：優越性 or 非劣性
- 有意水準：片側 or 両側、常に5%？
- 検出力
- 治療効果の大きさ（臨床的意義 or 見込み）

検出力の設定：80%以上がデフォルト

- β エラーは研究者側のリスクなので、 α エラーほど小さくする必要はない
 - 有意にならなくとも、現状より悪くなることはない
- 検出力50% = 見込みの Δ がバウンダリの試験
 - コインを振って決めるようなものなので非倫理的
 - どこまで検出力を下げて良いかについてのコンセンサスはない

JCOG0301の例（片側 $\alpha=5\%$ 、MST 10m vs 15m [HR=0.67]）



パラメータの設定時に考えること

- デザイン：優越性 or 非劣性
- 有意水準：片側 or 両側、常に5%？
- 検出力
- 治療効果の大きさ（臨床的意義 or 見込み）

Δを決めるための考え方

優越性試験・非劣性試験共通

A) 比較する治療法のリスク・ベネフィットを考慮して決める

- 臨床的に意味のある差を Δ に設定する
 - 毒性の差が大きければ大きな Δ 、小さければ小さな Δ
- 臨床的に意味のある差は研究者のコンセンサスで決める
 - どのくらいの差であれば新治療を受け入れるか

B) 得られる上乗せ効果 (見込み) を考慮して決める

- 勝つ見込みが大きければ、大きな Δ を設定

非劣性試験の場合（非劣性マージンの決め方）

C) プラセボに一定以上優っていることの担保するように決める

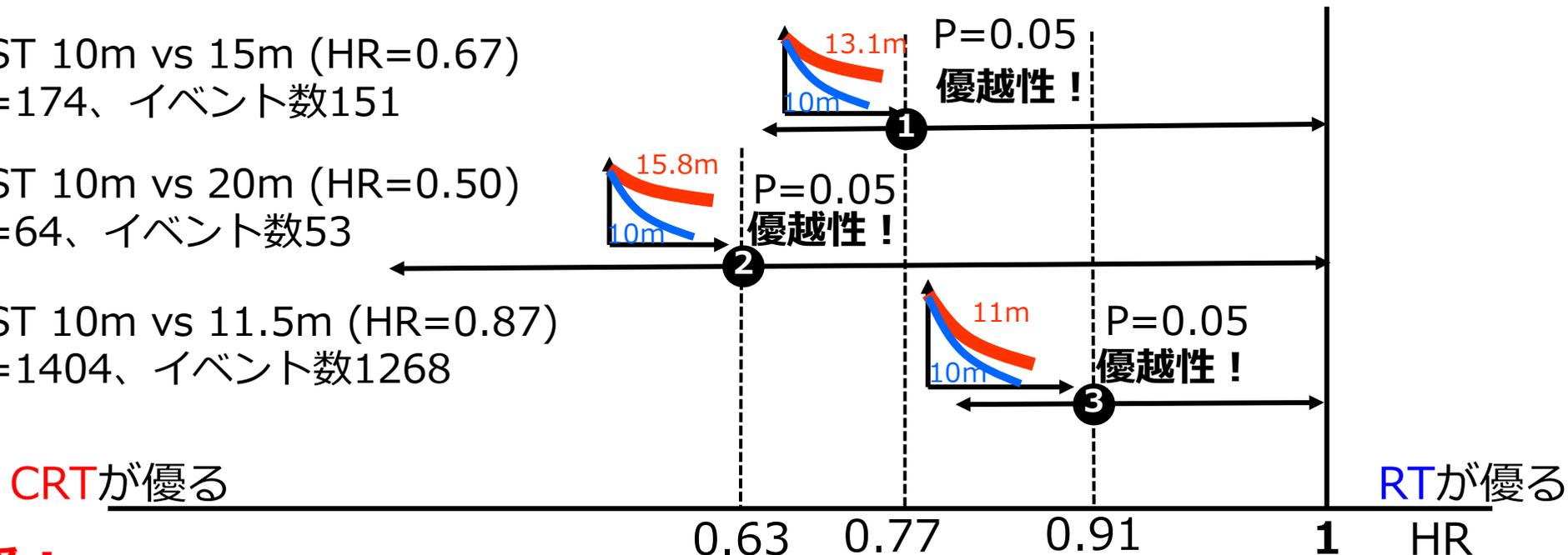
Δの設定と統計的有意差

JCOG0301の例（片側 $\alpha=5\%$ 、検出力80%）

MST 10m vs 15m (HR=0.67)
N=174、イベント数151

MST 10m vs 20m (HR=0.50)
N=64、イベント数53

MST 10m vs 11.5m (HR=0.87)
N=1404、イベント数1268



推奨！

- ① 臨床的に意味のある差を検出
- ② 必要解析対象数は少なくなるが、臨床的に意味のある差が得られても統計的に有意と判断できなくなる
- ③ 臨床的に意味のない差まで統計的に有意と判断してしまう

サンプルサイズ計算

～割合の群間比較の場合～

割合の検定のサンプルサイズ計算に必要な情報

- 有意水準： α
- 検出力： $1-\beta$ (= power)
- 治療効果の大きさ： Δ
 - Δ は絶対値（標準治療群と試験治療群の割合の差）
 - 各群の割合の大きさから**バラツキ**を計算

$$\chi^2 \text{検定} : N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \left[\begin{array}{l} \text{標準群の割合}(1 - \text{標準群の割合}) \\ + \text{試験群の割合}(1 - \text{試験群の割合}) \end{array} \right]^2}{\text{割合の差}^2}$$

まとめ

- 予後に関係する背景因子が治療群間で偏ることで、見かけ上の関連が生じる or 見かけ上関連が消える現象を交絡という
 - 交絡は**ランダム化**によって排除できる
- サンプルサイズに必要なパラメータは α 、 β 、 Δ 、バラツキ
- サンプルサイズ計算は統計家・臨床家の共同作業
 - 特に Δ の決定は主に臨床家が責任を持つべきパラメータ
 - Δ の決定にはclinical questionを明確にすることが必須
 - 優越性 or 非劣性
 - 片側 α or 両側 α
 - 臨床的に意味のある差 (Δ)