

サンプルサイズ計算の考え方

JCOGデータセンター統計部門

水澤 純基

JCOG臨床試験セミナー中級編

2015. 10.10 (土)

注：わかりやすさを優先しているため、統計的には正確ではない表現も含まれます

ICRweb 2015年9月末時点で会員数45000人以上の 臨床研究 e-ラーニング サイト [icrweb.jp]

厚生労働省「e-learningシステムICRwebを用いた臨床研究・治験に携わる 人材の育成方法に関する研究（H-24-臨研基-一般-001）」による臨床研究サイト

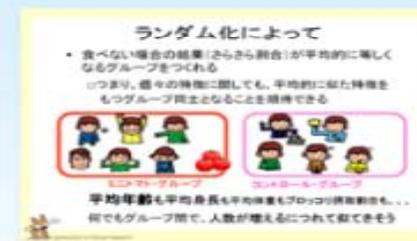


臨床研究に携わる人のeラーニングサイト
ICR 臨床研究入門

- ICRwebのご紹介
- リンク集
- お知らせ
- お問い合わせ
- サイトマップ

HOME 講座一覧 規制ガイドライン 用語集 臨床研究Tips 求人情報 サイトご利用方法 よくあるご質問

「ICR臨床研究入門」は臨床研究に携わる
すべての人を対象としたeラーニングサイトです。



厚生労働省・文部科学省 推薦

ユーザーログイン

ユーザーID

パスワード

[パスワードをお忘れの方はこちら](#)

[新規ユーザー登録](#)

ICR臨床研究入門(Introduction to Clinical Research)とは？

ICRwebは実際に臨床研究を実施する医学研究者だけではなく、それを支援するCRC（臨床研究コーディネーター）をはじめとする臨床研究専門職の方々、倫理審査委員会の委員や倫理審査委員会事務局、研究の事務的なお手伝いをする方々すべてを対象に臨床研究に必要な知識を提供するe-learningサイトです。基礎研究者、疫学者、統計学者、臨床研究のことをもっと勉強したい一般の方々も対象としています。... [続きはこちら](#)

- [新規ユーザー登録](#)
- [ICR臨床研究入門のご紹介](#)

「人を対象とする医学系研究に関する倫理指針」に対応した研究教育は[こちら](#)

コース案内

[注目・最新の講義](#)

[講義サンプル](#)

[ほかの臨床研究教育関連サイト](#)

コース一覧

[臨床研究の基礎知識講座](#)

[全講座履修コース](#)



臨床研究の基礎知識コース(旧 臨床研究入門初級編) **必修**

このコースは、臨床研究の倫理指針の教育義務に対応するコースとなっており、すべての方に必ず受けていただく必要のあるコースです。多くの施設で倫理審査委員会に研究計画書を提出する際の必須要件として利用していただいているコースです。

対象 研究者、CRC、データマネージャ、倫理審査委員、倫理審査委員会事務局等

iTunes プレビュー

iTunes is the world's easiest way to organize and add to your digital media collection.

お使いのコンピュータで iTunes が見つかりません。 e-Trial Co.,Ltd. による無料 App ICRweb をダウンロードするには、iTunes を入手してください。

すでに iTunes をお持ちの方は、「すでに iTunes を持っている」をクリックして iTunes を開いてください。

無料ダウンロード

ICRweb [View More by This Developer](#)

開発: e-Trial Co.,Ltd.

App を購入、ダウンロードするには iTunes を開いてください。

説明

ICR臨床研究入門（略称：ICRweb）は、「臨床研究教育プログラムの作成と普及」を目的として、インターネットを介した教育プログラムの提供、臨床研究に関する最新情報や役立つ情報の提供、および臨床研究に携わる人々への情報交換の場の提供を目指しています。

[ICRweb のサポート](#) ...さらに見る

バージョン 1.2.0 の新機能

- 統計計算ツールの追加及び既存ツールのブラッシュアップ
- 療育指針情報ページに検索機能を追加
- 講義詳細画面に講師名を表示

[iTunes で見る](#)

無料

カテゴリ: 教育

更新: 2015年7月10日

バージョン: 1.2.0

サイズ: 3.3 MB

言語: 英語

販売元: e-Trial Co.,Ltd.

© 厚生労働科学研究費補助金 Te-learningシステムICRweb を用いた臨床研究・治験に携わる人材の育成方法に関する研究」班（研究代表者 山本清一郎）

以下の理由により 12+ のレーティングがついています。

軽度な悪意または治療情報

互換性: iOS 7.1 以降。iPhone、iPad および iPod touch 対応。この App は iPhone 5 に最適化されています。

カスタマー評価

このアプリケーションの現行バージョンの平均評価を出すための十分なデータがありません。

iPhone スクリーンショット

Carrier 4:58 PM

ICR臨床研究入門

講座一覧

ICRからのお知らせ ¹ 最新状況をPC版へ反映

PC版サイトはこちら

キャリア 11:38 AM

ダウンロード済み動画一覧 [一括削除](#)

動画名	再生時間	操作
生物統計講座セミナー 1 仮設検定1	0:42:00	📄 🗑️ 🐾
生物統計講座セミナー 3 生存時間解析	1:07:00	📄 🗑️ 🐾
研究の実践に必要な知識（精神・神経領域） 2 臨床試験の実験作法	0:54:00	📄 🗑️ 🐾
臨床研究コーディネーター講座1-導入研修 3 CRCの役割と業務	0:44:00	📄 🗑️ 🐾
臨床研究の方法論トピック 1. がんの臨床研究論文を読むに必要な統計学	1:24:00	📄 🗑️ 🐾

• App Store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！

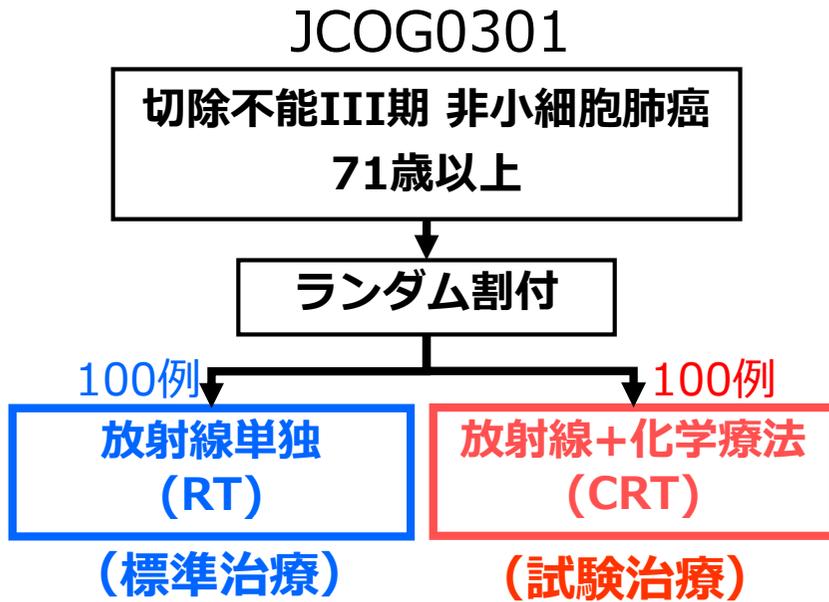


講義の目標

- 臨床家としてサンプルサイズ計算に必要なこと何か？を身につける
 - 統計家に相談する際に何を考えておくべきか
 - サンプルサイズ計算に必要な情報は何か
 - 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

サンプルサイズ計算法の考え方

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

and 16.9 months (p value=0.0179). In the radiotherapy group, thrombocytopenia (12.5%) than in the chemoradiotherapy group were similar.

chemoradiotherapy group.

group—the Japan Clinical Oncology Group (JCOG9812)—was assessed with low-dose carboplatin on survival than with unresectable NSCLC. The trial was stopped early after 4 years of patient accrual in the chemotherapy group.

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

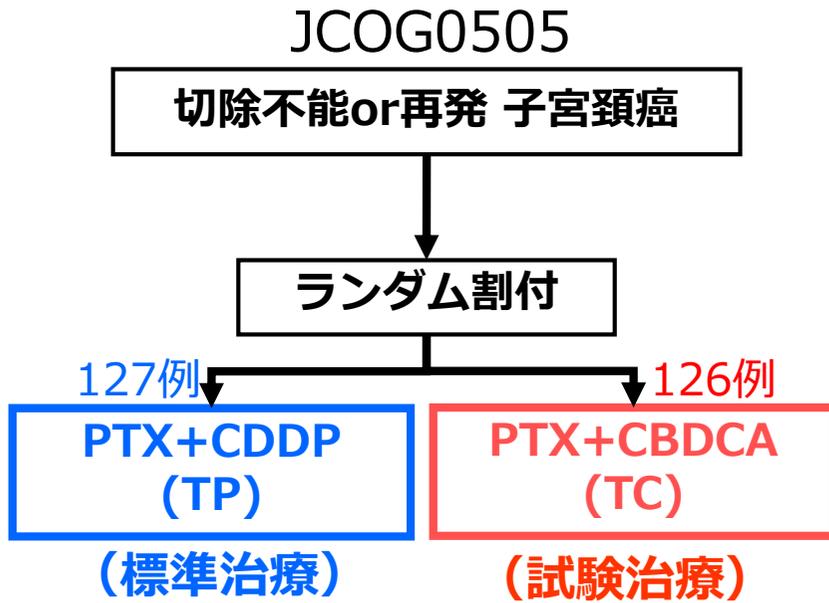


Published Online
May 22, 2012
DOI:10.1016/S1473-2045(12)70178-X
See Online/Comment
DOI:10.1016/S1473-2045(12)70178-X
Department of Thoracic Oncology, Kinki-chuo Chest Medical Center, Osaka, Japan (S Atagi MD); Department of Medical Oncology, Otemae Hospital, Osaka, Japan (M Kawahara MD); Department of Internal Medicine, Niigata Cancer Center Hospital, Niigata, Japan (A Yokoyama MD); Department of Respiratory Medicine and Medical Oncology, Yokohama Municipal Citizen's Hospital, Kanagawa, Japan (H Okamoto MD); Division of Thoracic Oncology, Shizuoka Cancer Center, Shizuoka, Japan (N Yamamoto MD); Division of Thoracic Oncology, National Cancer Center Hospital East, Chiba, Japan (Y Ohe MD); Division of Respiratory Medicine and Medical Oncology, Gifu Municipal Hospital, Gifu, Japan (T Sawa MD); Department of Radiology, Nagoya City University Graduate School of Medical Sciences, Aichi, Japan (S Ishikura MD); Japan Clinical Oncology Group Data Center, National Cancer Center, Tokyo, Japan (T Shibata MSc, H Fukuda MD); Medical Oncology Division, Kinki University School of Medicine, Osaka, Japan (N Saijo MD); and Department of Thoracic Oncology, National Cancer Center Hospital, Tokyo, Japan (T Tamura MD)

論文のサンプルサイズ記載例：非劣性試験

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT



Paclitaxel Plus Carboplatin Versus Paclitaxel Plus Cisplatin in Metastatic or Recurrent Cervical Cancer: The Open-Label Randomized Phase III Trial JCOG0505

Ryo Kitagawa, Noriyuki Katsumata, Taro Shibata, Toshiharu Kamura, Takahiro Kasamatsu, Toru Nakanishi, Sadako Nishimura, Kimio Ushijima, Masashi Takano, Toyomi Satoh, and Hiroyuki Yoshikawa

See accompanying article doi: 10.1200/JCO.2014.57.7122

ABSTRACT

Purpose

In metastatic or recurrent cervical cancer, cisplatin-based chemotherapy is standard. The JCOG0505 randomized phase III trial evaluated the clinical benefits of carboplatin-based regimen.

Patients and Methods

Eligible patients had metastatic or recurrent cervical cancer and had \leq one platinum-containing treatment and no prior taxane. Patients were randomly assigned either to conventional paclitaxel plus cisplatin (TP; paclitaxel 135 mg/m² over 24 hours on day 1 and cisplatin 50 mg/m² on day 2, repeated every 3 weeks) or paclitaxel plus carboplatin (TC; paclitaxel 175 mg/m² over 3 hours and carboplatin area under curve 5 mg/mL/min on day 1, repeated every 3 weeks). Primary end point was overall survival (OS). Planned sample size was 250 patients to confirm the noninferiority of TC versus TP with the threshold hazard ratio (HR) of 1.29.

Results

Between February 2006 and November 2009, 253 patients were enrolled. The HR of OS was 0.994 (90% CI, 0.79 to 1.25; noninferiority $P = .032$ by stratified Cox regression). Median OS was 18.3 months with TP versus 17.5 months with TC. Among patients who had not received prior cisplatin, OS was shorter with TC (13.0 v 23.2 months; HR, 1.571; 95% CI, 1.06 to 2.32). One treatment-related death occurred with TC. Duration of hospitalization periods was significantly longer with TC ($P < .001$).

Ryo Kitagawa, NTT Medical Center Tokyo; Taro Shibata, National Cancer Center Hospital, Tokyo; Noriyuki Katsumata, Nippon Medical School, Museshikosugi Hospital, Kanagawa; Toshiharu Kamura and Kimio Ushijima, Kurume University, Fukuoka; Toru Nakanishi, Aichi Cancer Center Hospital, Aichi; Sadako Nishimura, Osaka City General Hospital, Osaka; Masashi Takano, National Defense Medical College, Saitama; and Toyomi Satoh and Hiroyuki Yoshikawa, University of Tsukuba, Ibaraki, Japan.

Published online ahead of print at www.jco.org on March 2, 2015.

Processed as a Rapid Communication manuscript.

Written on behalf of the Japan Clinical Oncology Group and Gynecologic Cancer Study Group investigators.

Statistical Design and Analysis

Initially, with an accrual time of 2.5 years and minimum follow-up period of 1 year, the required number of OS events was 209 and the planned sample size was 250 according to the Schoenfeld and Richter method¹⁷ to confirm the noninferiority of TC compared with TP, with a one-sided α level of 0.05 and power of at least 70%, with noninferiority margin of 1.29, corresponding to 2 months in OS, assuming a median OS in the TP group of 9 months based on outcome in the GOG 169 trial (SCC only).

metastatic or recurrent cervical received platinum agents.

be administered over 24 hours to analog when combined with cisplatin inpatient hospital stay for

has been reported to be a less analog than cisplatin for cervical these agents have not been compared in phase III trials. Carboplatin induces less nausea/vomiting, and less than cisplatin.¹⁰ The combination and paclitaxel allows for hydration over 3 hours, and carbohydration. In the first multi-center II trial of paclitaxel and metastatic or advanced cervical cancer, we found an overall RR of 43 to 75), median PFS of 5.3

サンプルサイズ計算の基本的な考え方

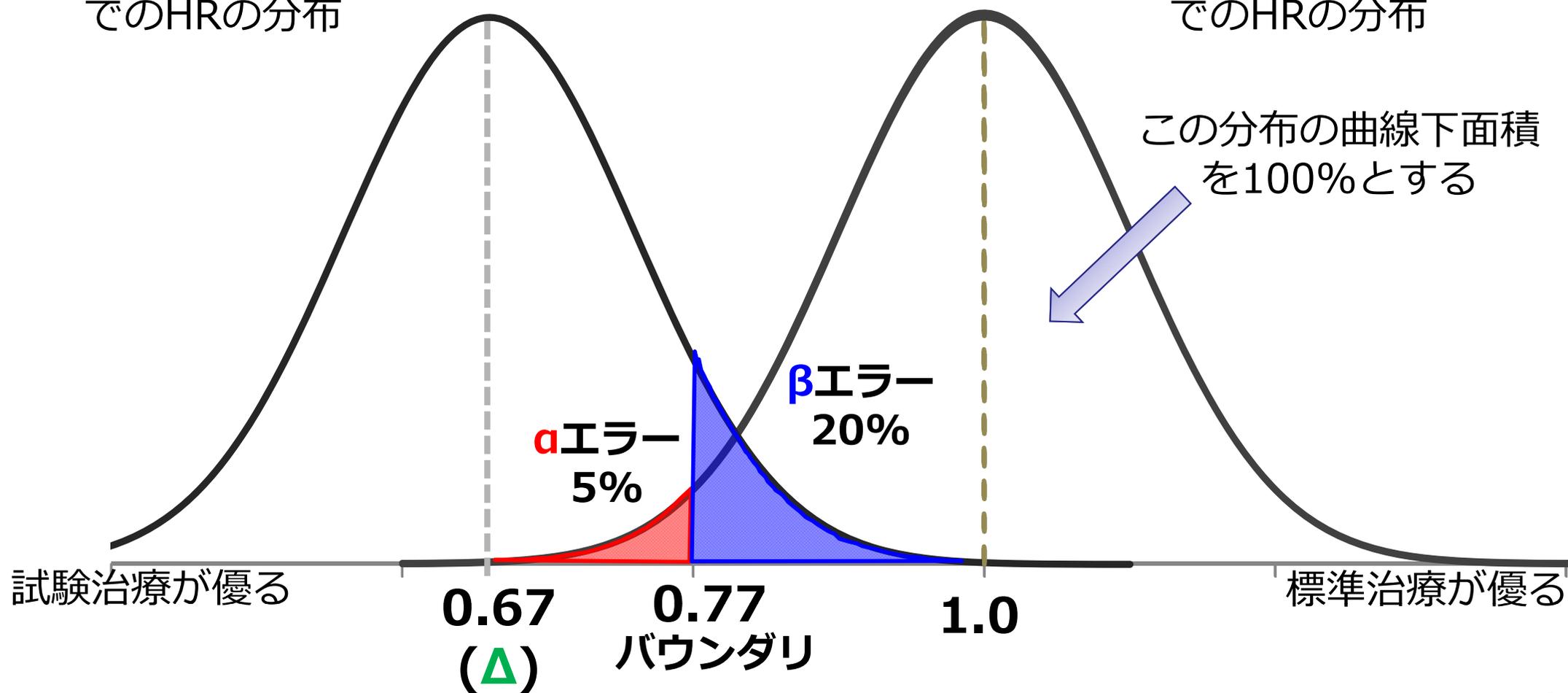
サンプルサイズ計算で求める必要解析対象数とは、
以下の条件を満たすように定める

1. 帰無仮説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. 新治療の真の効果量が 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
(**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

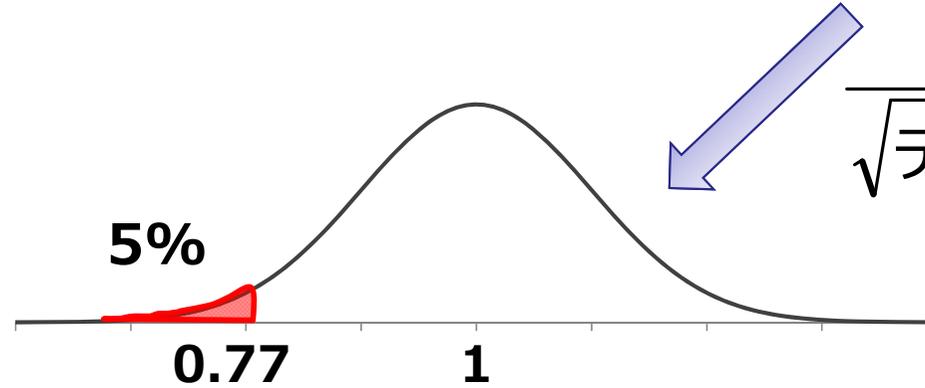
帰無仮説(HR=1)の下
でのHRの分布



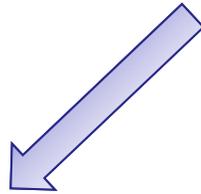
サンプルサイズと検定統計量の分布の関係

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

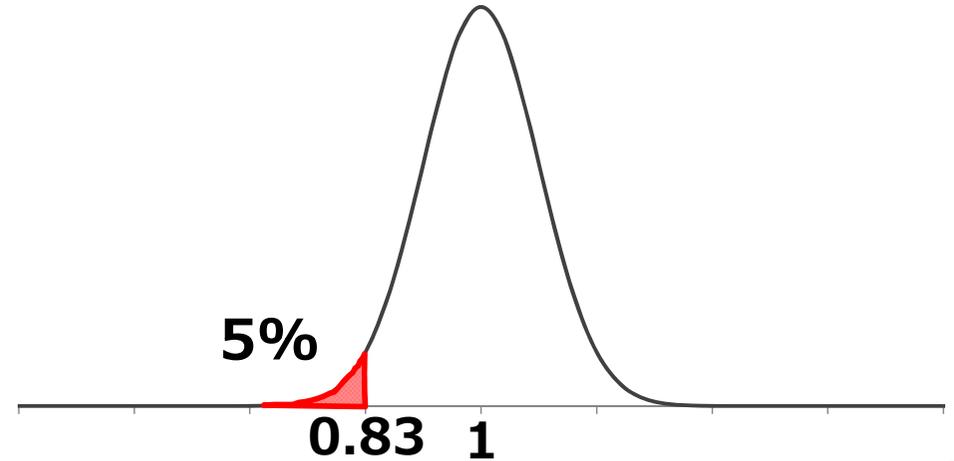
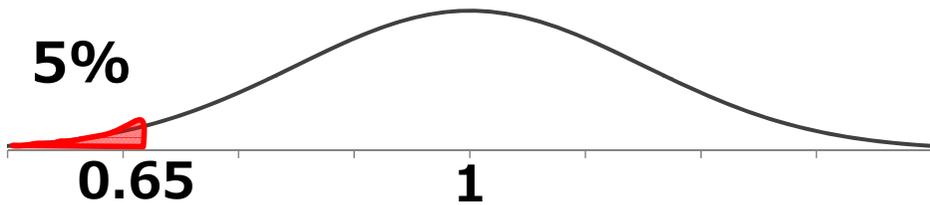
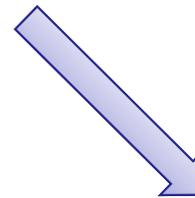
$$\frac{\Delta}{\sqrt{\text{データのバラツキ}^2 / N}}$$



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

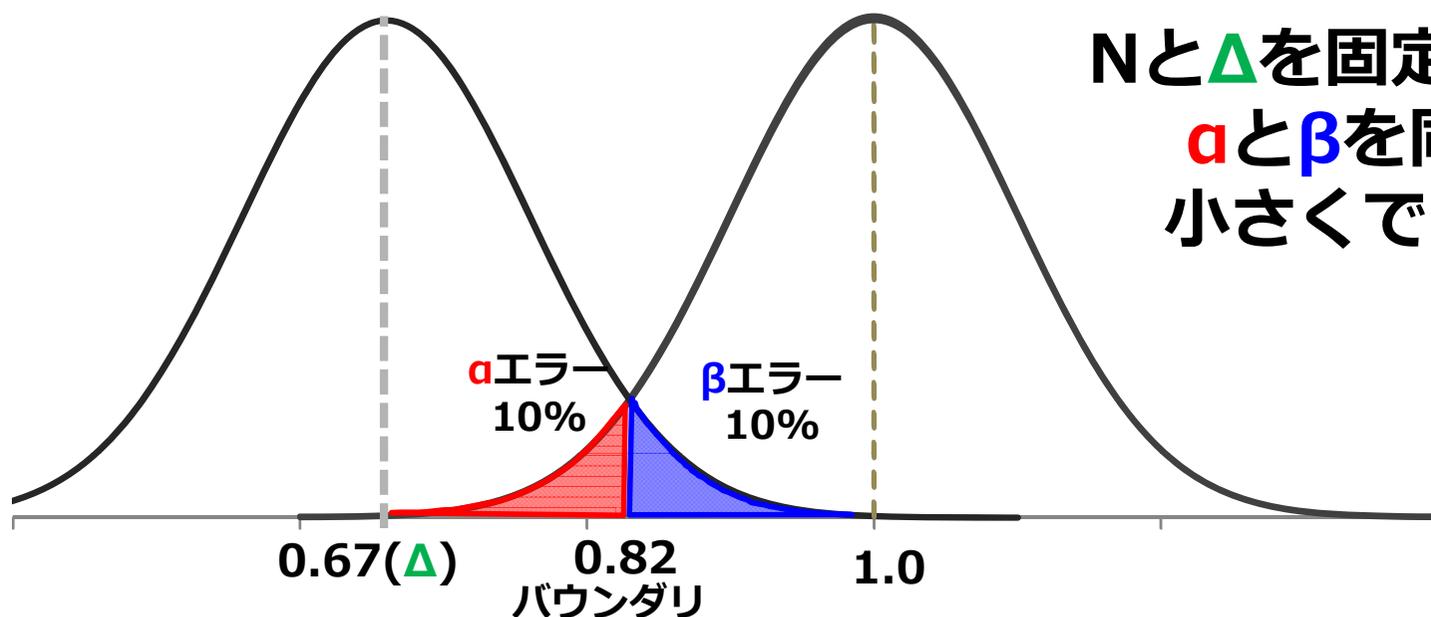
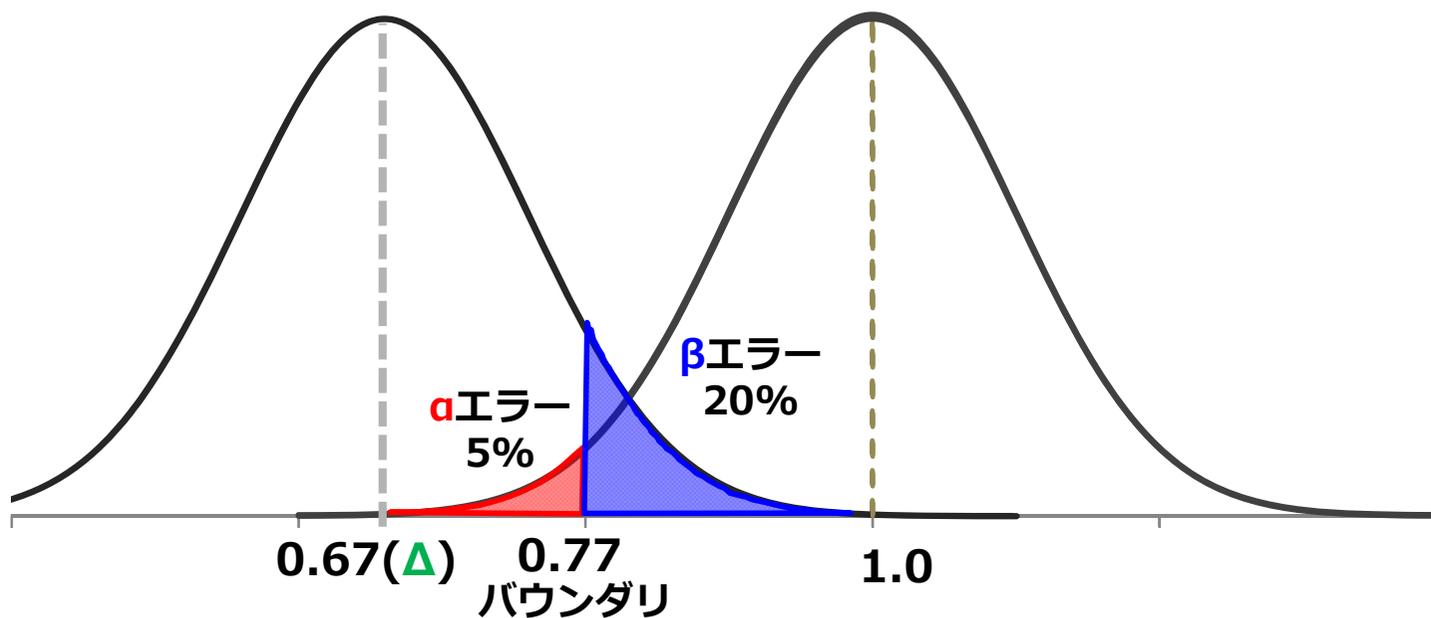
$$t\text{検定} : N = \frac{(Z_{\alpha} + Z_{\beta})^2 \cdot \text{データのバラツキ}^2}{\text{平均値の差}^2}$$

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

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

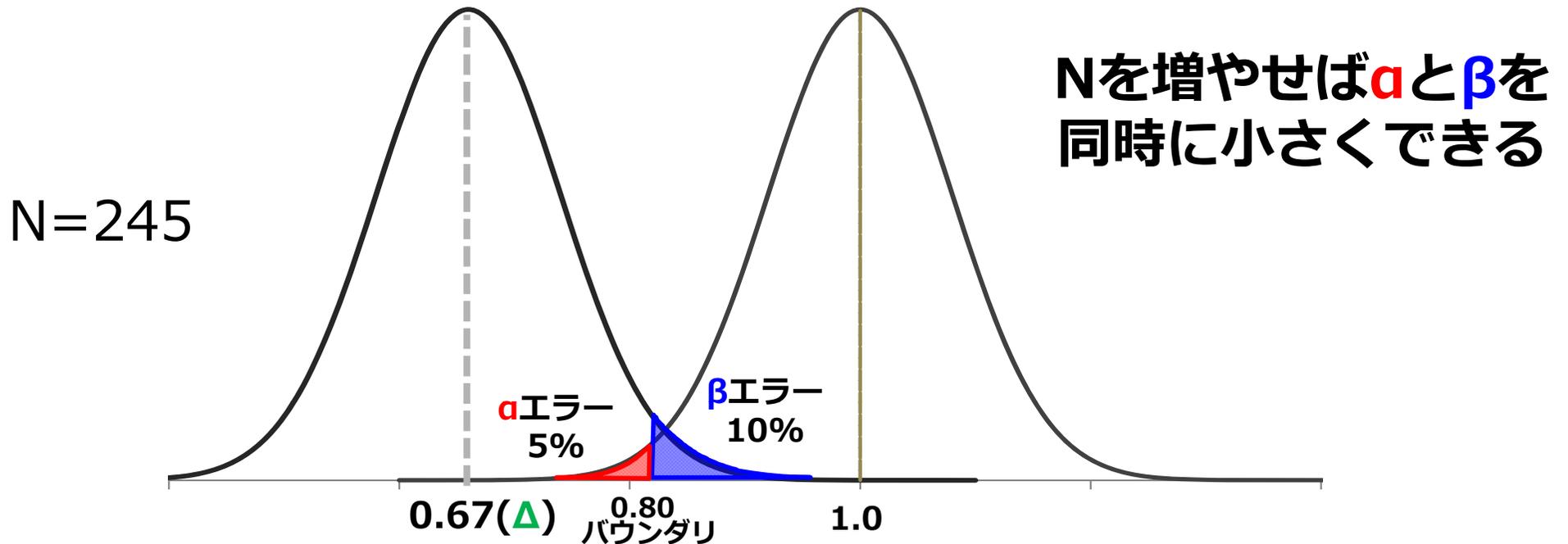
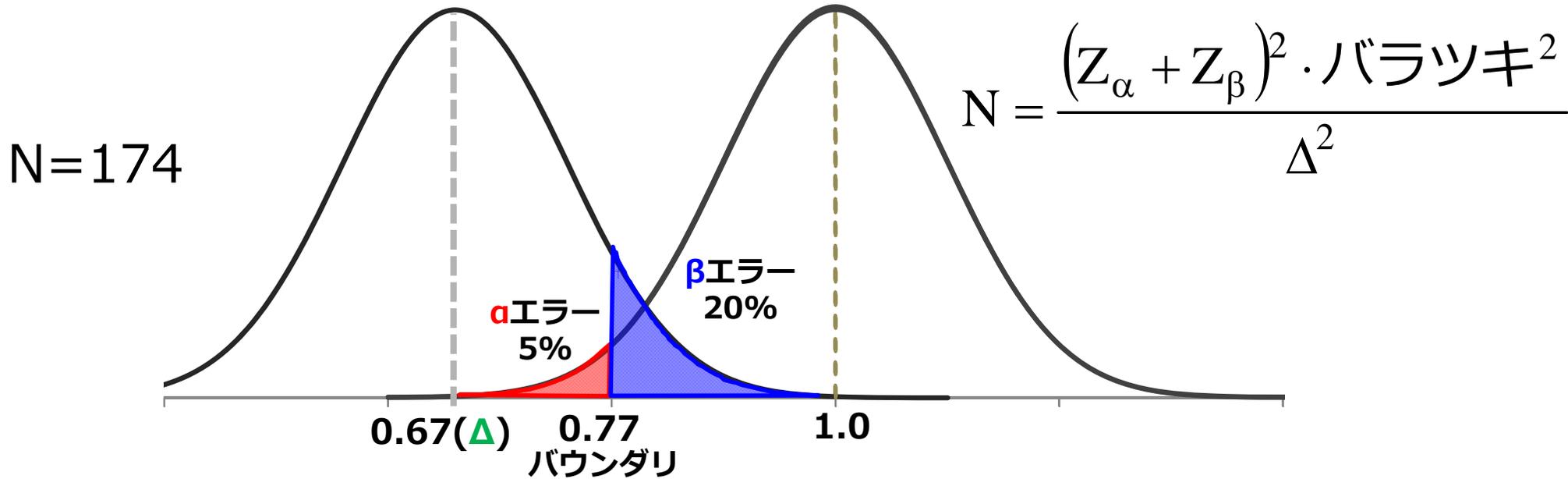
αとβはトレードオフ

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



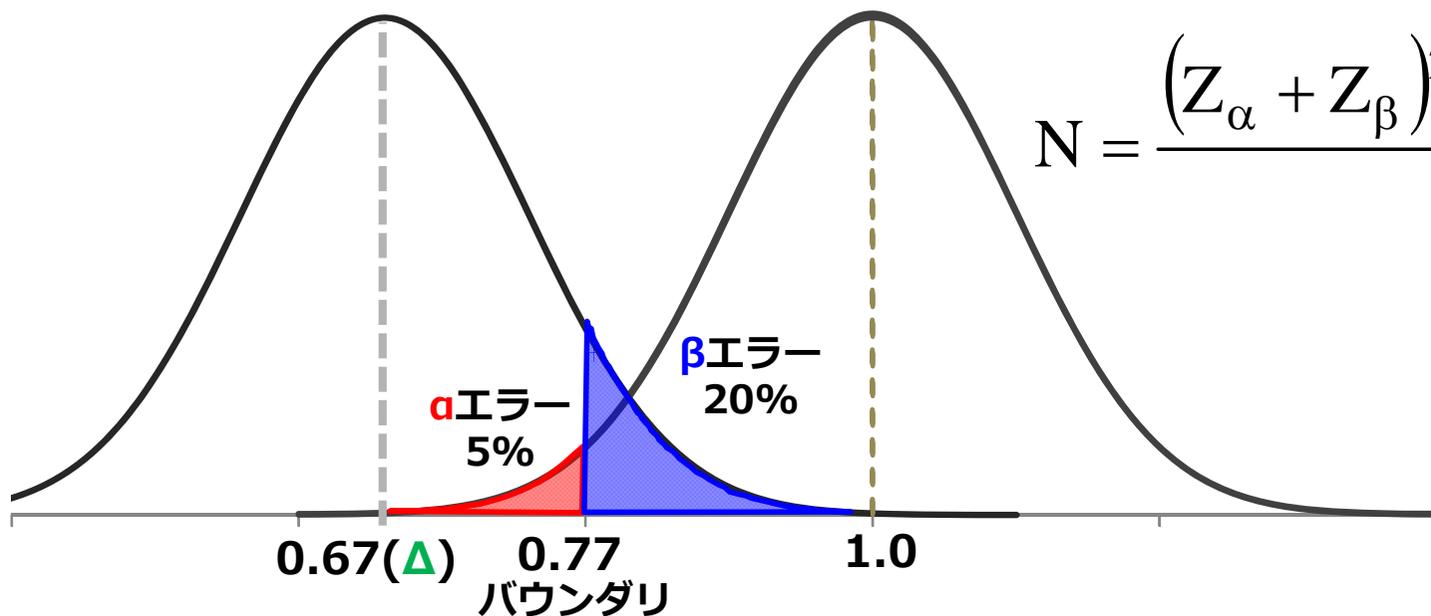
NとΔを固定すると、
αとβを同時に
小さくできない

Nを増やせば α を固定してPowerをあげられる

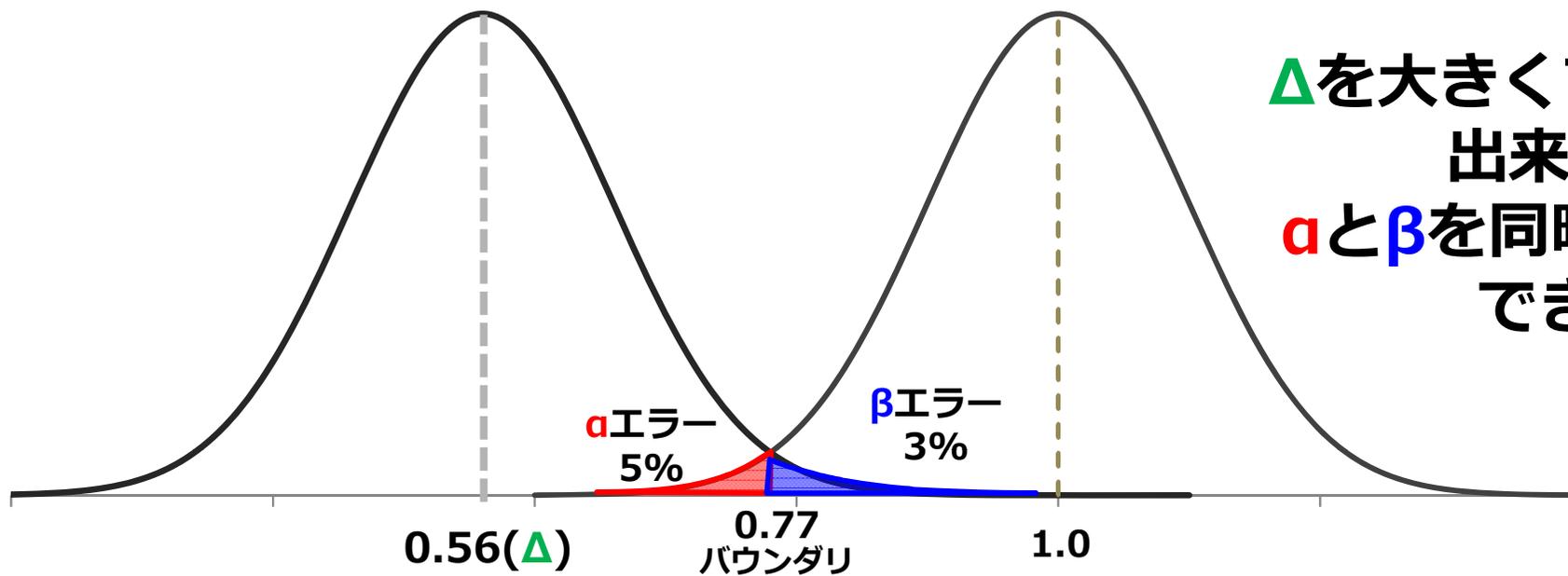


△を大きく出来ればαを固定してPowerをあげられる

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



△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



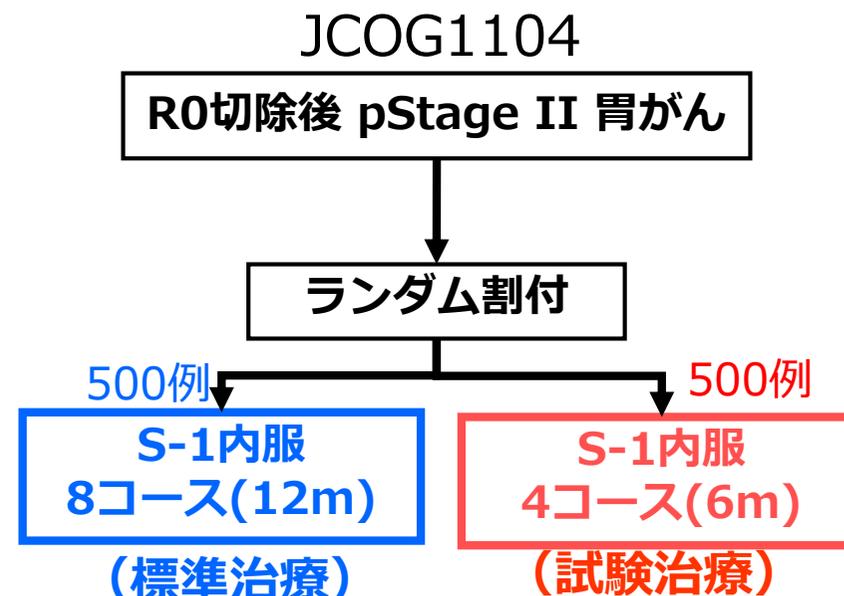
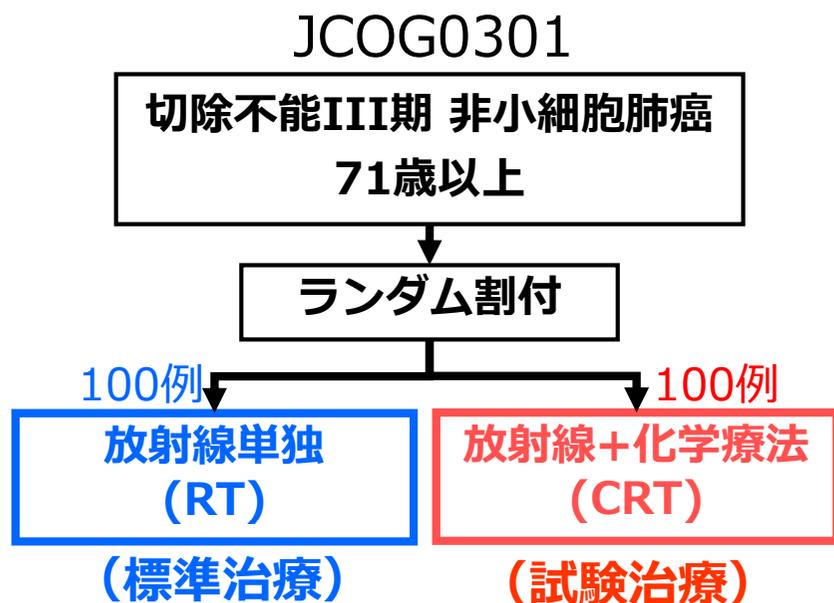
暗記！サンプルサイズ計算に最低限必要なパラメータ

- 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**)

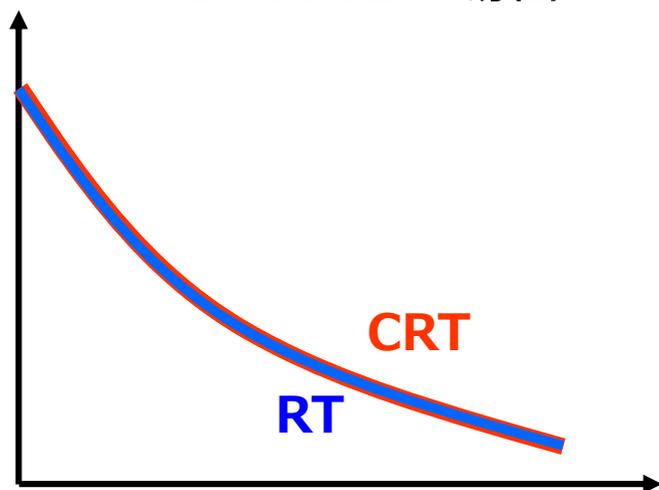
非劣性試験

- 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

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

演習1：アプリを使ってサンプルサイズ計算

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	1	ハザード比	0.6667
	α (有意水準)	0.05	1群あたりのサンプルサイズ	87
2群の生存期間の比較 (非劣性、生存割合に基づく)	検出力	0.8	2群を合わせたサンプルサイズ	174
2群の生存期間の比較 (優越性、MSTに基づく)	登録期間(年)	4	2群を合わせた期待イベント数	151
2群の生存期間の比較 (非劣性、MSTに基づく)	フォローアップ期間(年)	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
	対照群のMST(月)	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
	試験群のMST(月)	15		

結果表示

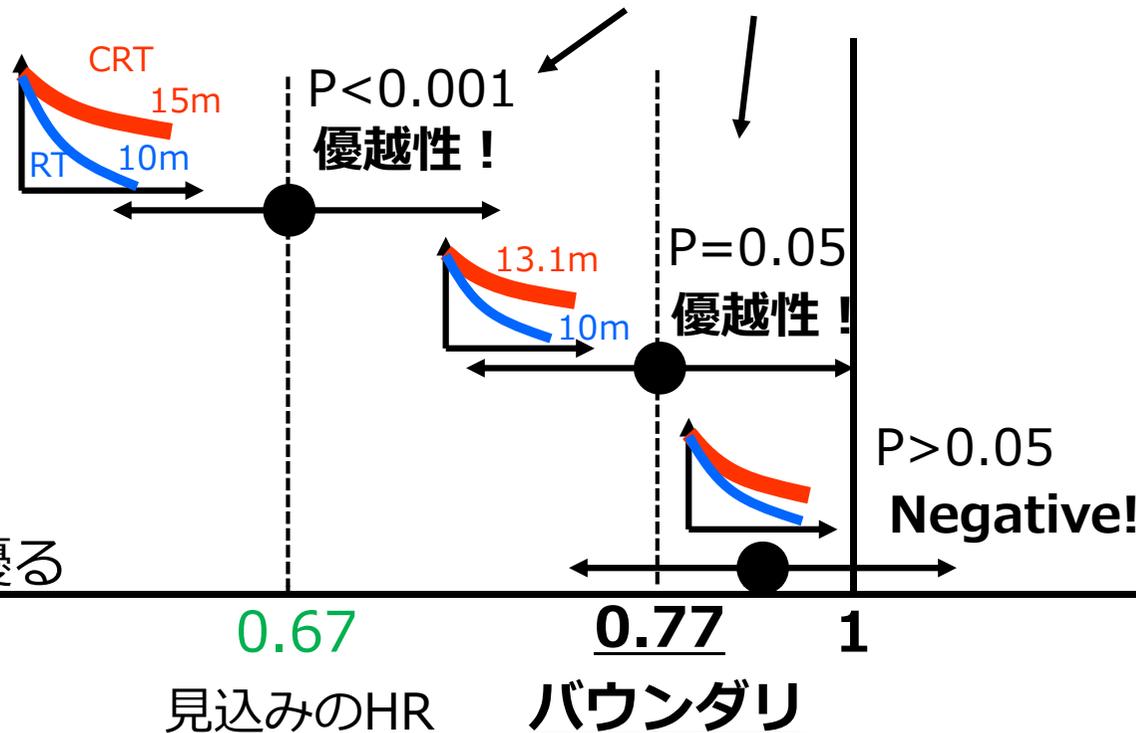
統計的有意差とバウンダリ：優越性試験の場合

- 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



演習2：アプリを使ってサンプルサイズ計算

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



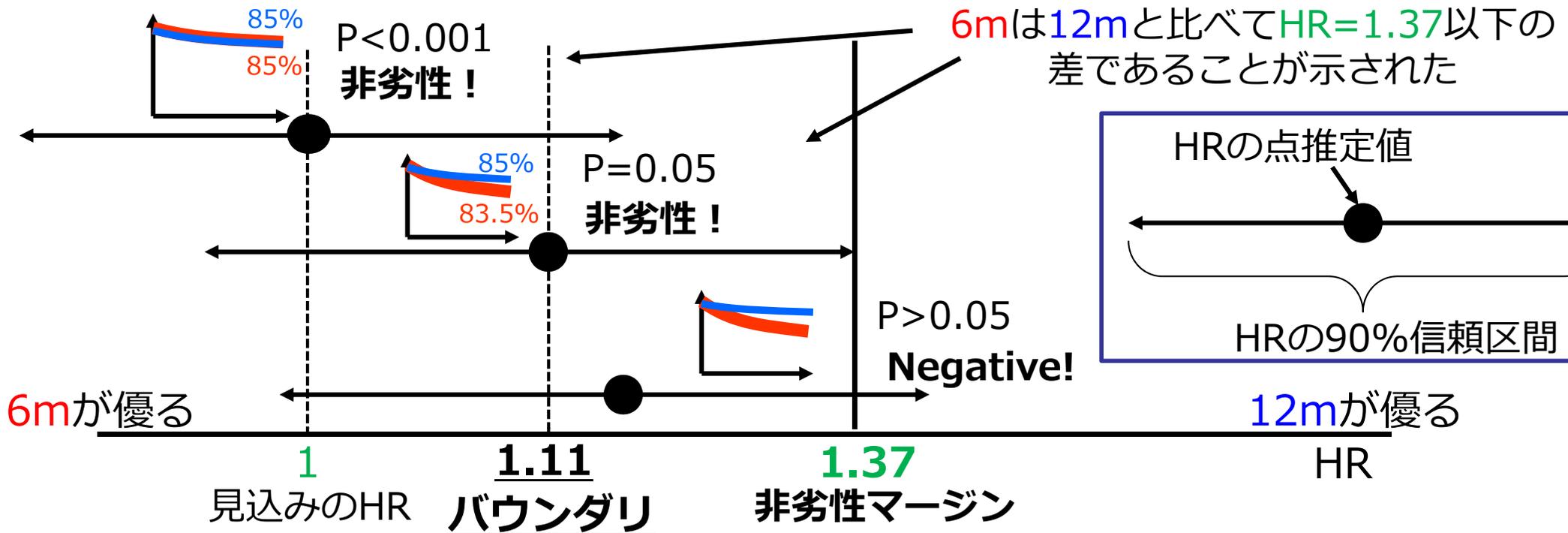
パラメータ入力		結果	
	サンプルサイズ	検出力	結果表示
片側検定なら1、両側検定なら2	1		許容ハザード比
α (有意水準)	0.05		1.3730
検出力	0.8		1群あたりのサンプルサイズ
登録期間(年)	5		482
フォローアップ期間(年)	3		2群を合わせたサンプルサイズ
考慮する生存割合の年数(何年時点での生存割合か?)	3		964
対照群の生存割合	0.85		2群を合わせた期待イベント数
試験群の生存割合	0.85		246
対照群に与える生存割合のハンデ(対照群から差し引く割合)	0.05		ギリギリ有意になる（棄却限界値に対応する）ときのハザード比
			1.1133
			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
			0.8345

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)

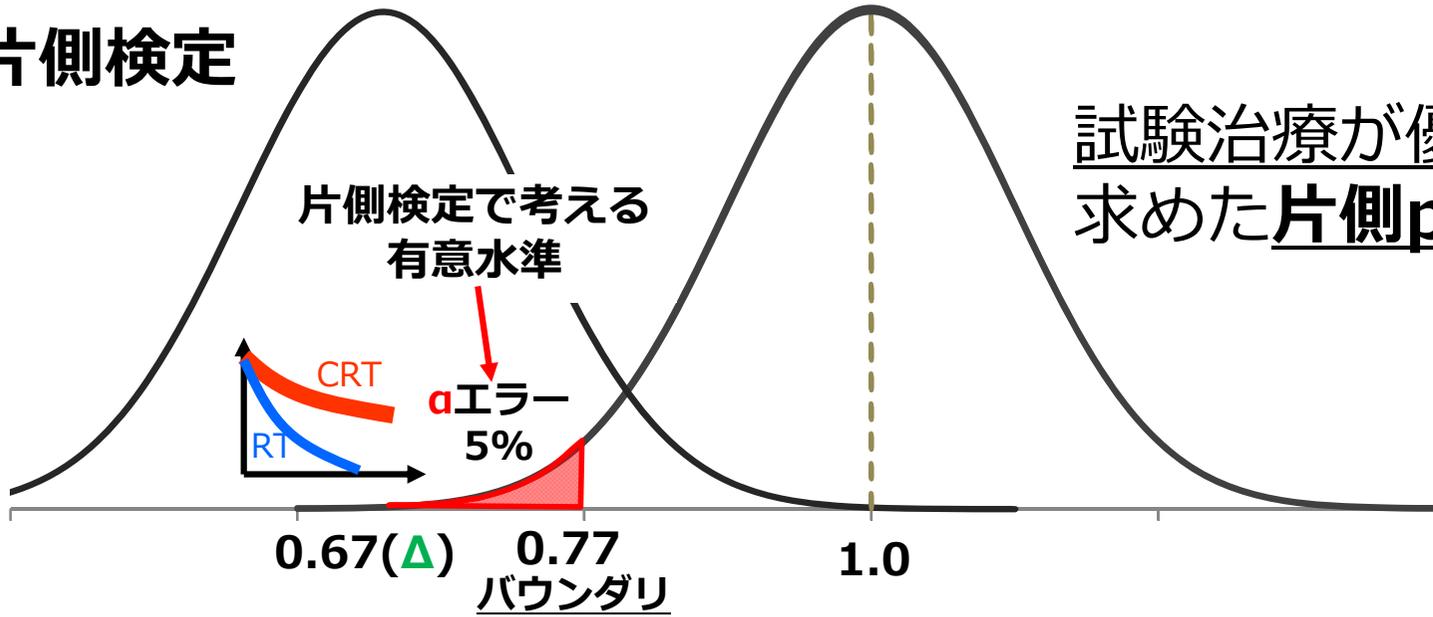


パラメータの設定時に考えること

- 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）

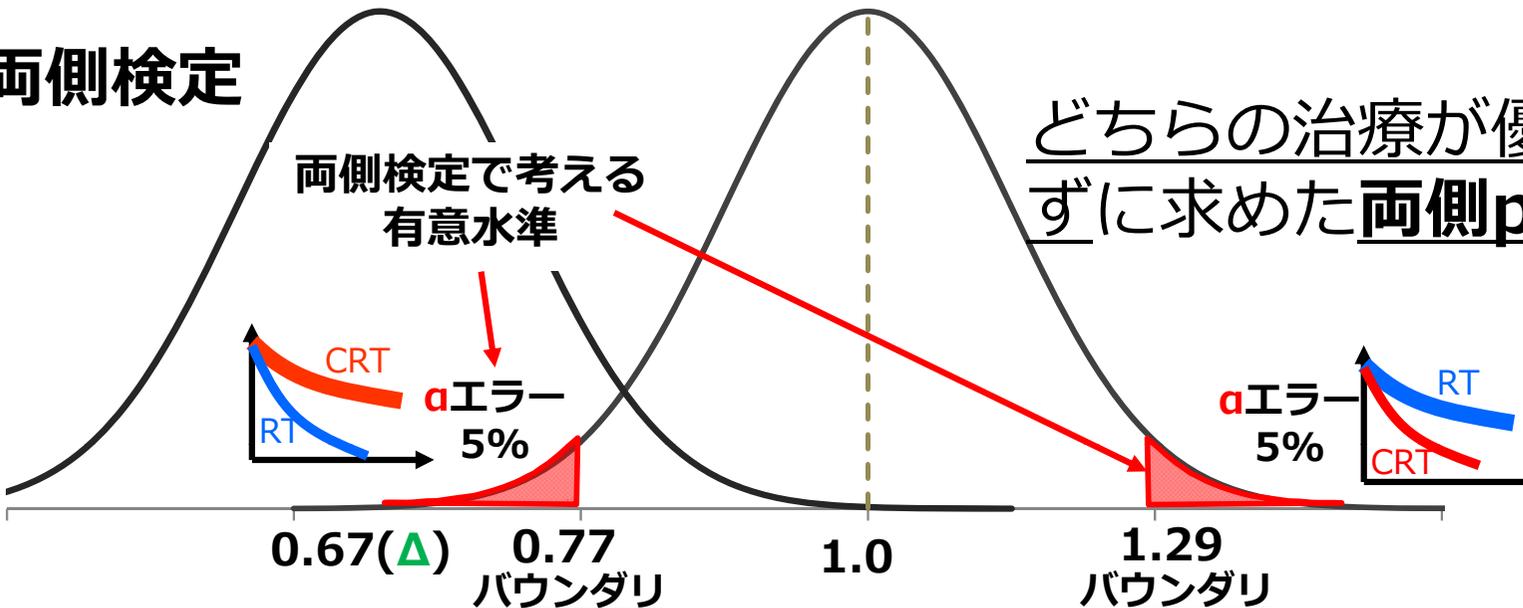
有意水準 α ：片側検定と両側検定

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!

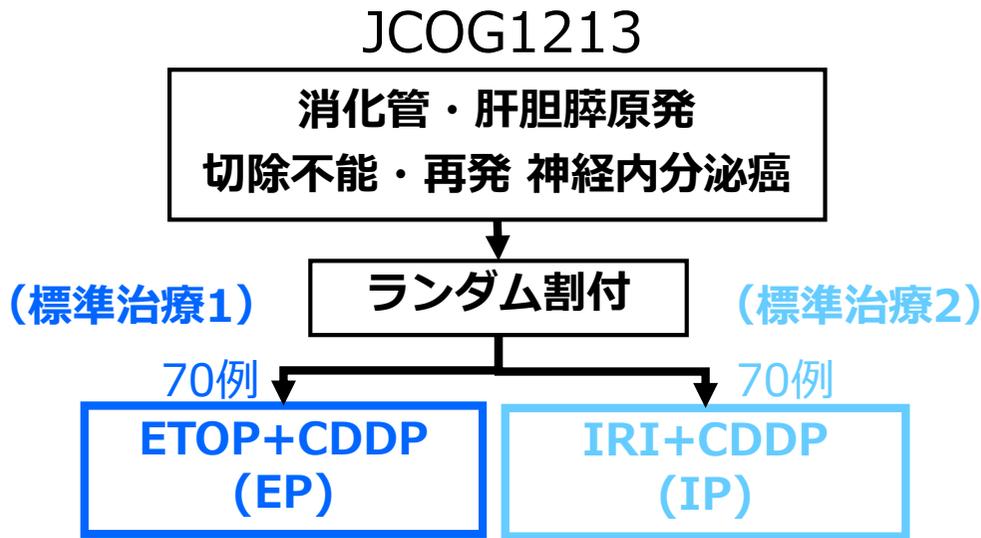
A) 治療法のリスク・ベネフィットを考慮し、意思決定に応じて決める

- 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 (JCOG1213)

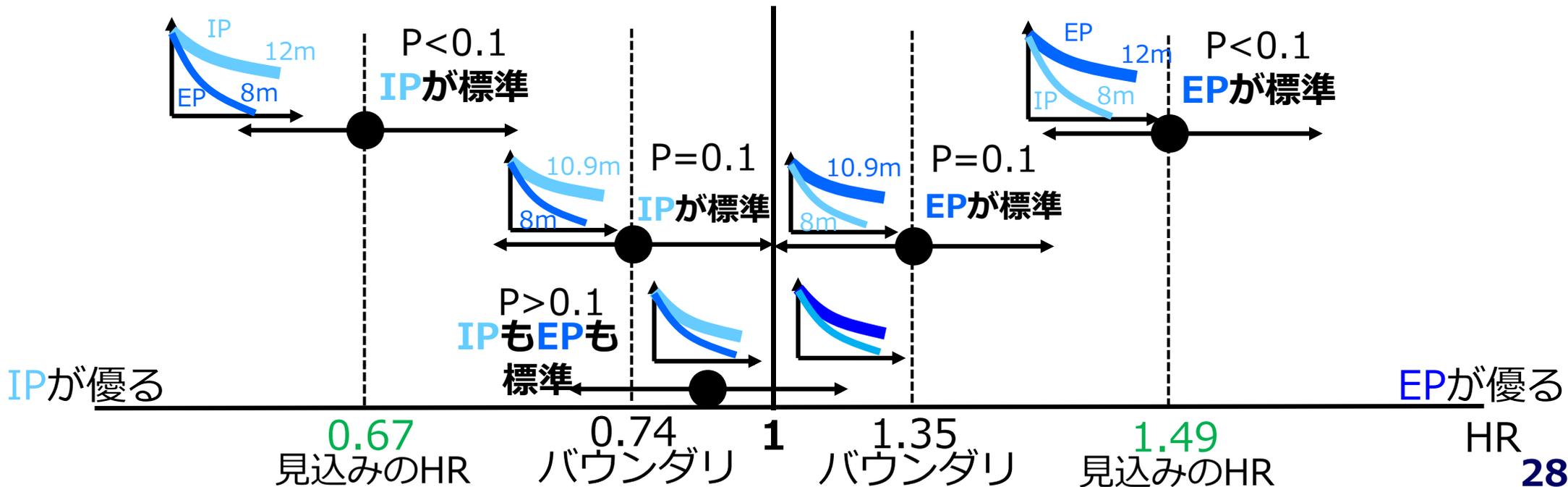
B) 試験治療が標準治療に劣る可能性がある場合は常に両側検定とすべき

- 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

両側 α を設定しているJCOG試験例



- 有効性・安全性とも優劣をつけがたい
2つの標準治療の比較
- どちらが有意に優ることにも興味がある
- **両側 $\alpha=10\%$** 、検出力=80%
- MST 8m vs 12m (HR=0.67)



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]**)：従うのに理由はいらない
 - 両側→片側に変更するとサンプルサイズが少なくなるわけではない
 - 両側5%に対応するのは片側2.5% (サンプルサイズは同じ)
 - 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
- 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
 - サブグループ間で治療効果に違いがあるか (**交互作用**) の検定
- 両側 or 片側 $\alpha= 0.0001\sim 0.005$
 - **Phase III**：多重性を考慮した**中間解析**での有意水準
 - 試験の途中では、よほどの差がない限り有効中止しない

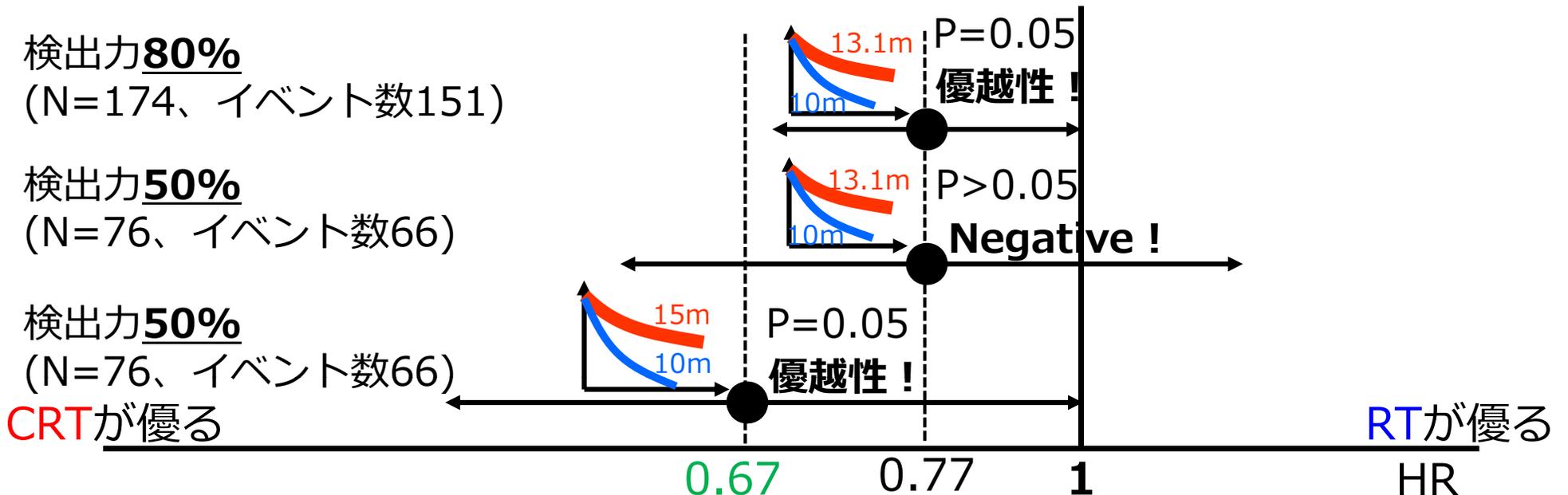
パラメータの設定時に考えること

- 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

JCOG0301計画段階の仮想例[優越性試験]

RT群

- MST=**10m**と想定
- RTは1日2Gy,週5日間,60Gy
- AnyGrade3-4 \leq 10%
- Grade4のAEは0%

CRT群[低用量連日CBDCA+RT]

- Phase IIの結果、MST=**20m**
- RTは1日2Gy,週5日間,60Gy
- CBDCAは1日1回 30mg/m²を
RT開始前1時間以内に静注
- G4好中球減少=25%、
G3のFN=12.5%
- 5mの上乗せがあれば、毒性や手間に見合う
(=臨床的に意味のある最小限の差)

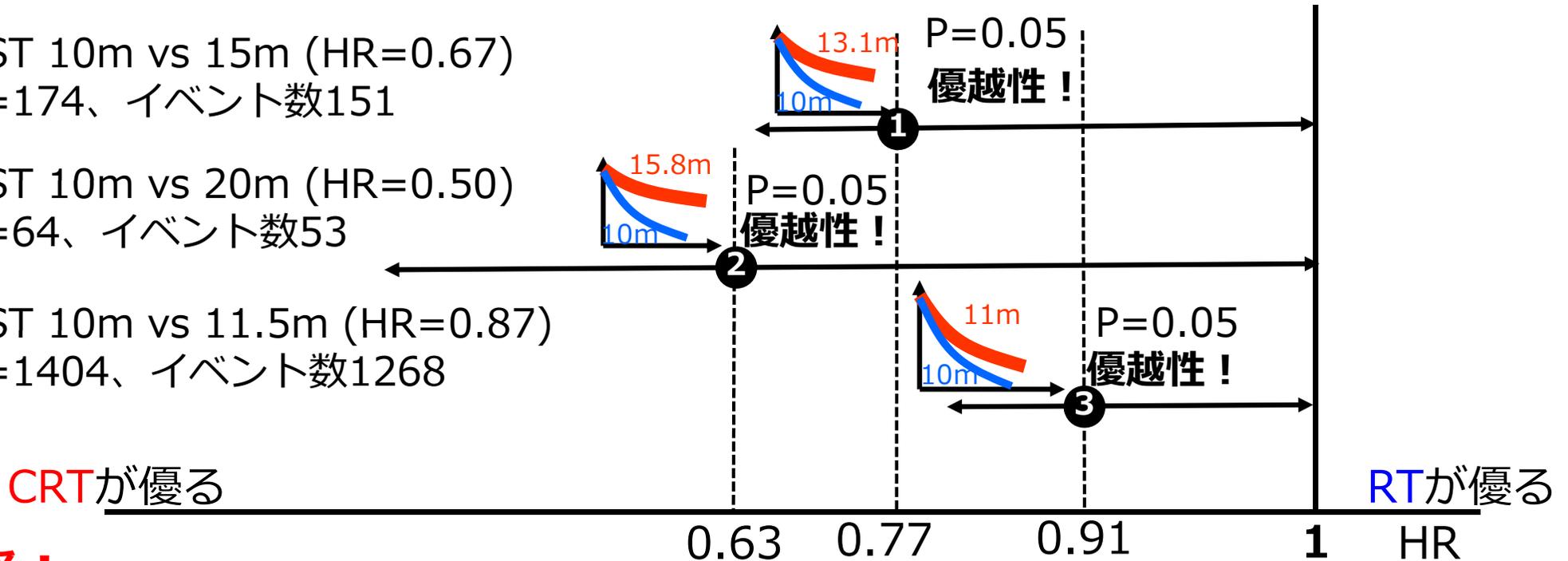
Δの設定と統計的有意差

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

JCOG1104試験計画の背景

ACTS-GC

胃切除後 II-IIIIB期 胃癌
20-80歳

ランダム割付

529例

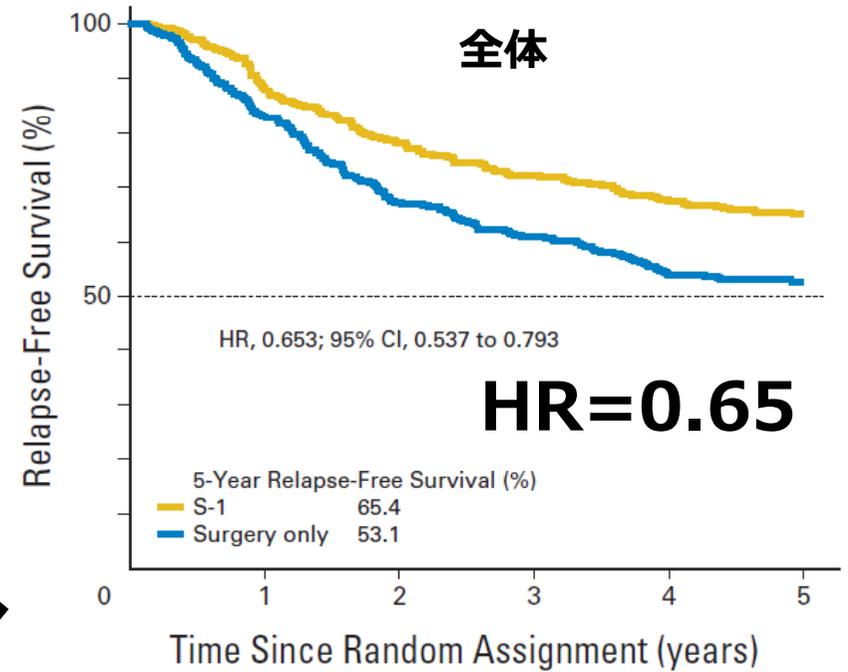
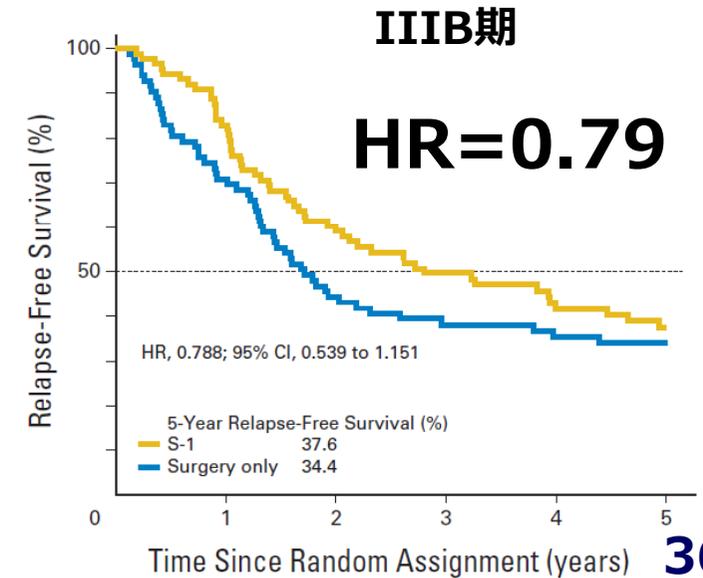
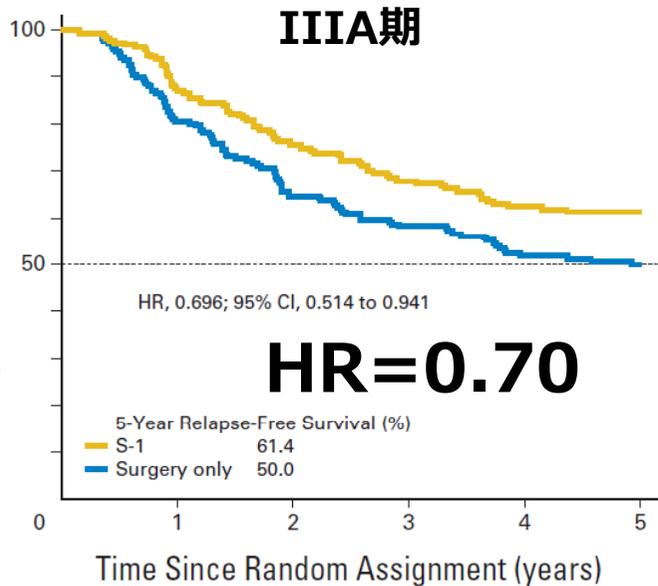
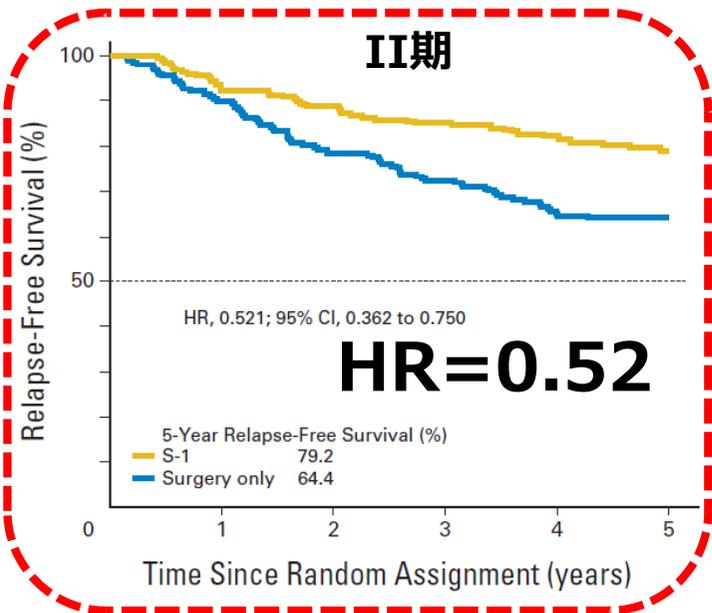
530例

経過観察
(Surgery only)

S-1 1年内服
(S-1)

(標準治療)

(新治療)



JCOG1104計画段階の仮想例[非劣性試験]

S-1 12m投与群

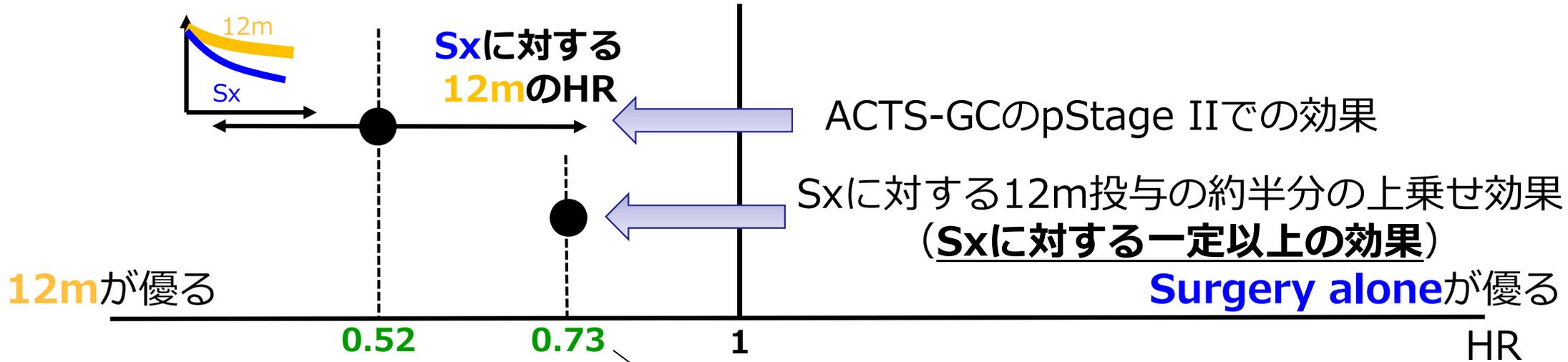
- 3yRFS=**85%**と想定
- 1年間の消化器毒性
 - Grade2以上の食欲不振、嘔気、下痢、疲労、皮膚炎 $\geq 20\%$
- 1年分のコスト

S-1 6m投与群

- 3yRFS=**85%**と想定
- 6m以降におこるAEは消失
 - 通院などの患者負担も軽減
- コストは半分

- 3yRFSで5%以下の差(HRの非劣性マージン=1.37)なら6m投与にメリットあり
- Surgery aloneに対する一定以上の効果は担保されている？

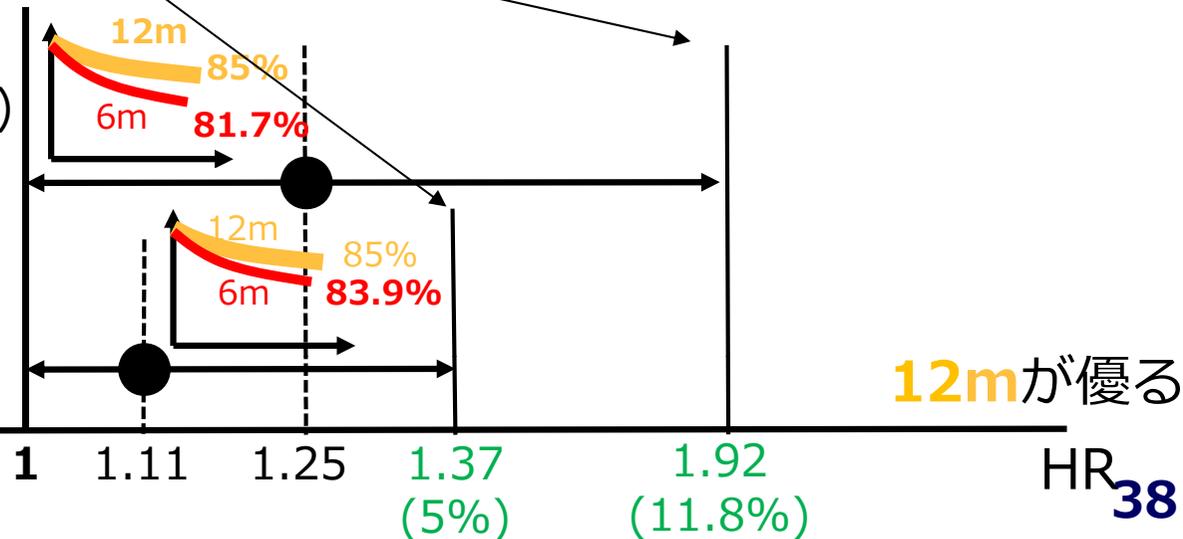
Surgery aloneに対する一定以上の効果とは



6mは12mと比べてHR=1.92以下の差
(Sxに優っていることが示されたに過ぎない)

6mは12mと比べてHR=1.37以下の差
(12mのSxに対する一定以上の効果を担保)

6mが優る



勝つ見込みのある非劣性試験デザイン

VOLUME 25 · NUMBER 31 · NOVEMBER 1 2007

JOURNAL OF CLINICAL ONCOLOGY

SPECIAL ARTICLE

Randomized Clinical Trial Design for Assessing Noninferiority When Superiority Is Expected

Boris Freidlin, Edward L. Korn, Stephen L. George, and Robert Gray

ABSTRACT

The randomized clinical trial (RCT) is the gold standard for definitive evaluation of new therapies. RCTs designed to show that the therapeutic efficacy of a new therapy is not unacceptably inferior to that of standard therapy are called noninferiority trials. Traditionally, noninferiority trials have required very large sample sizes. Sometimes, a new treatment regimen with a favorable toxicity and/or tolerability profile is also expected to have some modest improvement in efficacy. In such specialized settings we describe a hybrid trial-design approach that requires a dramatically smaller sample size than that of a standard noninferiority design. This hybrid design can naturally incorporate a formal test of superiority as well as noninferiority.

J Clin Oncol 25:5019-5023. © 2007 by American Society of Clinical Oncology

INTRODUCTION

The majority of randomized clinical trials (RCTs) are superiority trials, designed to show that one treatment arm (experimental arm) is superior to another (standard arm) with respect to some clinically relevant outcome. The objectives of these trials are often contrasted with noninferiority trials designed to show that a new treatment arm is noninferior to the standard therapy with respect to some clinical outcome. The latter approach may be justified by a favorable toxicity profile, convenience, or the cost associated with the new regimen relative to the standard arm.

The goal of RCTs is to provide definitive evaluation of new regimens. When the new regimen is incremental in nature (eg, addition of a new drug to the standard backbone), a superiority trial is appropriate. A definitive evaluation must provide sufficiently convincing evidence to distinguish between the conclusion that the new regimen is superior to the standard versus the conclusion that new regimen does not provide clinically meaningful benefit. When the new regimen was developed to provide a better safety or tolerability profile while preserving most of the efficacy of the standard regimen (eg, by reducing a dose of a toxic component), a noninferiority trial is appropriate. A definitive evaluation must be able to provide compelling evidence to distinguish between the conclusion that the new regimen preserves the benefit of the standard regimen (noninferiority) versus the conclusion that the new regimen does not preserve a clinically important fraction of the standard arm benefit (eg, the dose-

reduced regimen produces only one half of the complete remissions induced by the full-dose regimen).

Whether the trial is a superiority trial or a noninferiority trial, the sample size needed to provide sufficiently compelling evidence to address its goal is determined mainly by the distance (on the treatment effect scale) between the negative and the positive conclusions (this difference is called the target treatment effect). In noninferiority trials, the target treatment effect generally is smaller than in superiority trials. Thus, larger sample sizes generally are required to achieve the necessary level of evidence. The increase in sample size can be quite dramatic. This often makes noninferiority designs infeasible because of the inability to complete the study in a reasonable time frame.

Sometimes the traditional classification of trials as either superiority trials or noninferiority trials is inadequate, and there are some features of each type of trial present. This occurs in the special case when a new treatment regimen with a favorable toxicity and/or tolerability profile is also expected to have some modest improvement in efficacy. For example, lenalidomide, an immunomodulatory analog of thalidomide, is believed to have a more potent anti-multiple myeloma effect than thalidomide, with lower systemic toxicity.¹ We argue that in this case the design should take into consideration the expected marginal benefit of the new regimen; this leads to a hybrid design with a dramatically smaller sample size than the standard noninferiority design. Similar designs have been mentioned in the statistical literature² and implemented in two recent studies.^{3,4} Furthermore, we argue that a definitive evaluation

- 小さな非劣性マージンを設定した非劣性試験と小さな差を設定した優越性試験のサンプルサイズはほぼ同等
- きちんとデザインされた非劣性試験は非常に大規模になる
- 新治療の毒性が低く、かつ有効性も上回ることを期待される場合、実現可能なサンプルサイズを設定できないか？

From the Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD; Department of Biostatistics and Bioinformatics and the Cancer and Leukemia Group B Statistical Center, Duke University School of Medicine, Durham, NC; and Eastern Cooperative Oncology Group, Harvard School of Public Health, Boston, MA. Submitted March 23, 2007; accepted July 30, 2007.

Supported in part by Grant No. U10CA033601 from the National Cancer Institute (S.L.G.).

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

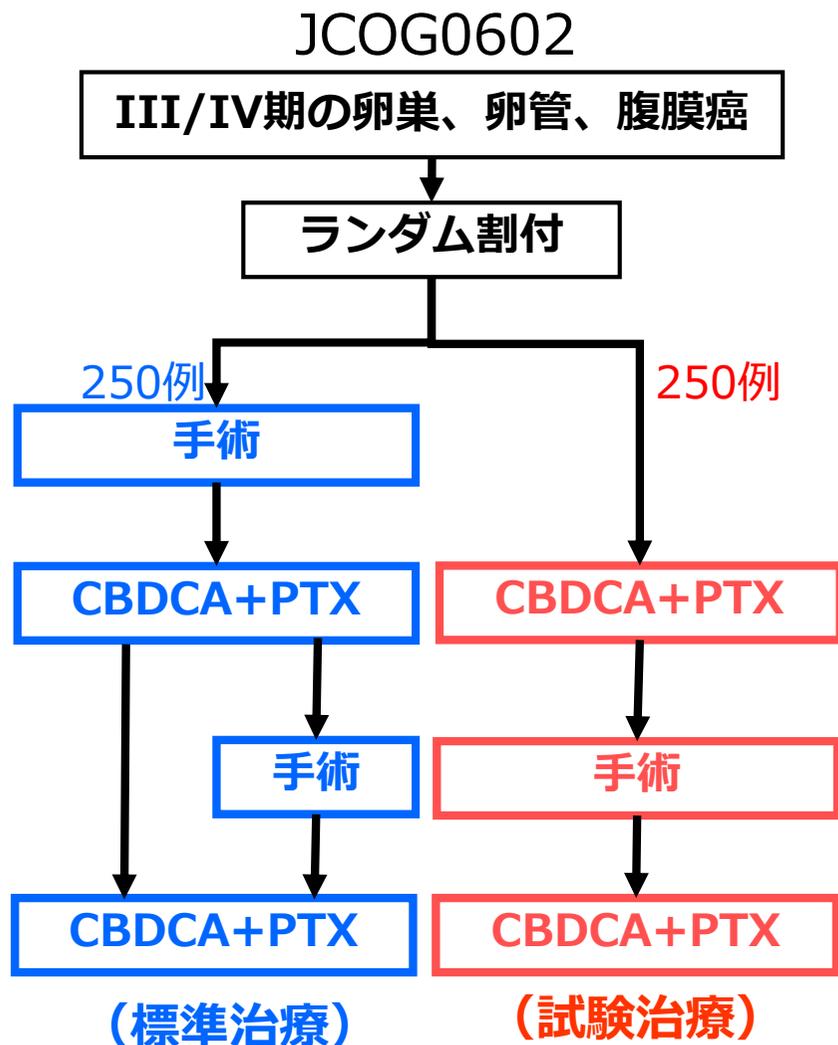
Address reprint requests to Boris Freidlin, PhD, Biometric Research Branch, 6130 Executive Blvd, EPN-8122, MSC 7434 National Cancer Institute, Bethesda, MD 20892; e-mail: freidlinb@ctep.nci.nih.gov.

© 2007 by American Society of Clinical Oncology

0732-183X/07/2531-5019/\$20.00

DOI: 10.1200/JCO.2007.11.8711

勝つ見込みのある非劣性試験デザインの例



非劣性デザインのrationale

- OSが同じなら**試験治療**を選択
 - 手術が1回で済む
 - NAC後の手術により、手術関連の合併症の軽減

勝つ見込みのrationale

- 早期のNAC
- 改善した状態で手術を受けることで残存腫瘍<1cmの手術が可能

演習3：アプリを使ってサンプルサイズ計算

JCOG0602のサンプルサイズ設計

- 片側 $\alpha=5\%$
- 検出力=80%
- 登録期間=3年、追跡期間=5年
- 標準治療 vs 試験治療 の3yOS : 25% vs 30.3% (HR=0.86)
- HRの非劣性マージン : 1.161 (3yOSで5%に相当)

パラメータ入力

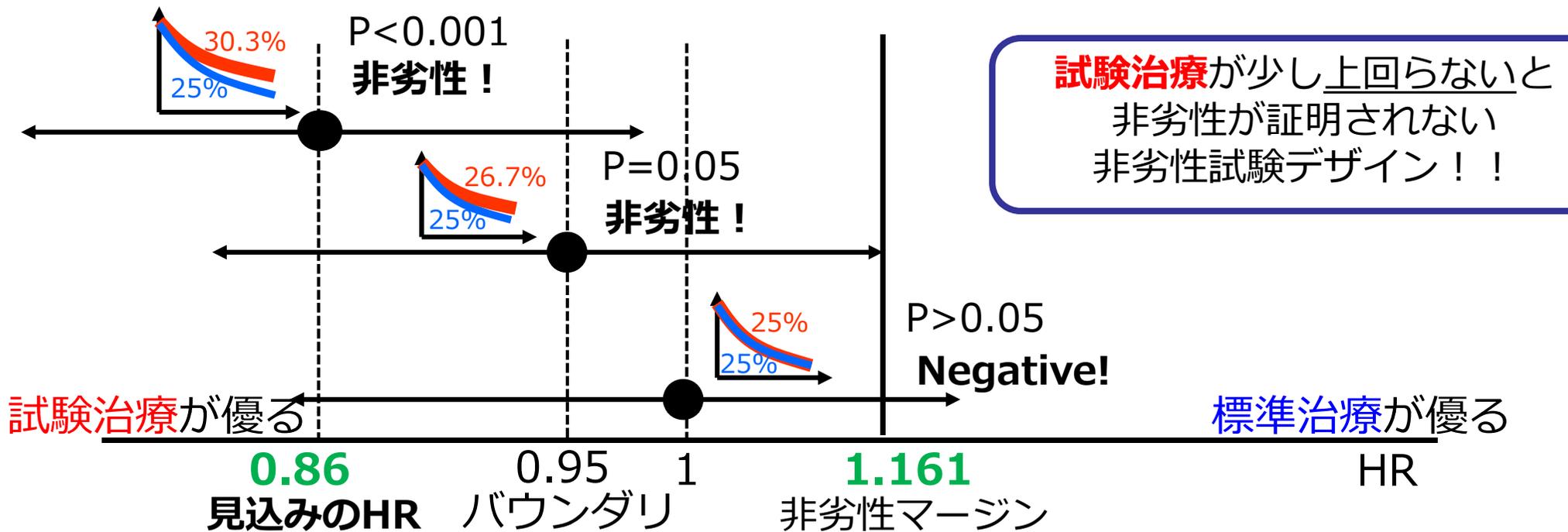
結果

パラメータ入力	値	結果	値
片側検定なら1、両側検定なら2	1	許容ハザード比	1.1610
α (有意水準)	0.05	1群あたりのサンプルサイズ	149
検出力	0.8	2群を合わせたサンプルサイズ	298
登録期間(年)	3	2群を合わせた期待イベント数	278
フォローアップ期間(年)	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	0.9531
考慮する生存割合の年数(何年時点での生存割合か?)	3	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	0.2668
対照群の生存割合	0.25		
試験群の生存割合	0.303		
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)	0.05		

JCOG0602のバウンダリ

JCOG0602のサンプルサイズ設計

- 片側 $\alpha=5\%$
- 検出力=80%
- 標準治療 vs 試験治療 の3yOS : 25% vs 30.3% (HR=0.86)
- HRの非劣性マージン : 1.161 (3yOSで5%に相当)
- 必要解析対象数 (期待イベント数) : 計298例 (278)
 - 標準vs試験のHR=1の場合、必要解析対象数 (イベント数) : 1174 (1110)



どのような試験デザインにしますか？

プラチナ製剤使用後のIIIB/IV期非扁平非小細胞肺癌に対する
ドセタキセル VS **ニボルマブ** (単純にするため、コストは考えないものとします)

ドセタキセル

- 毒性強い
 - Any AE grade3-4 50%以上
 - G3-4疲労 10%
 - G3-4好中球減少 30%
 - FN 10%
 - Any AE grade1-4 86%

- MST=8m程度

- ① 優越性試験
- ② 非劣性試験
- ③ 勝つ見込みの非劣性試験

ニボルマブ

- 毒性弱い
 - Any AE grade3-4 7%
 - G3-4疲労 1%
 - G3-4白血球減少 1%
 - その他 0%
 - Any AE grade1-4 58%

- MST=12m程度

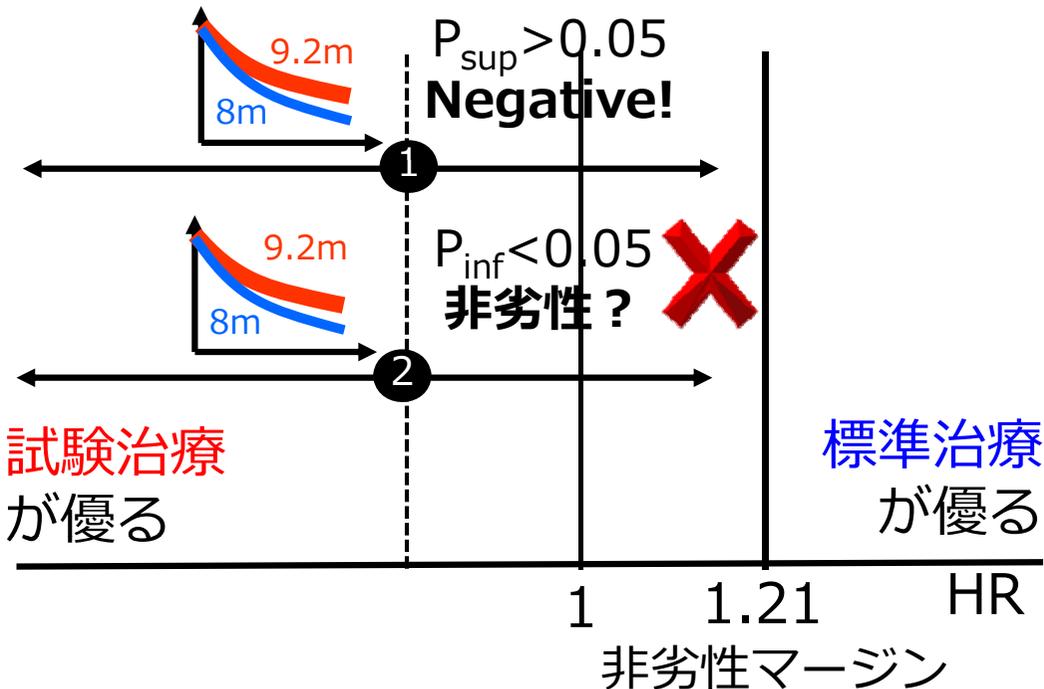
優越性狙えるのに非劣性は抵抗あるな・・・
 非劣性試験では優越性を主張できないでしょ？



非劣性デザインの設定で優越性の主張も可能

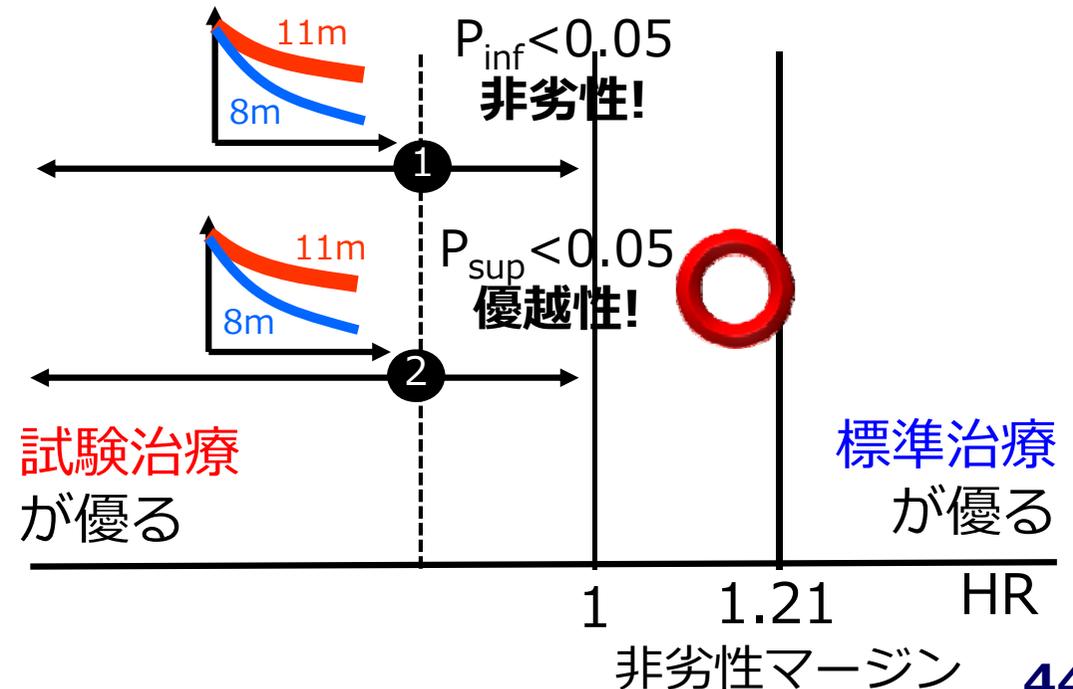
優越性デザインの場合

優越性が示されなかった時に、非劣性仮説を後付け解析するのはダメ



非劣性デザインの場合

非劣性が示された時に、優越性仮説を解析するのはOK



サンプルサイズ計算

～生存時間の群間比較の場合～

生存時間解析のサンプルサイズの算出に必要な情報

- 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 (5yOSやMST) ではなく相対値 (HR)
 - 指数曲線を仮定して5yOSやMSTの Δ をHRに変換する
- 対象集団の予後の見積もり
- 登録期間
- 追跡期間

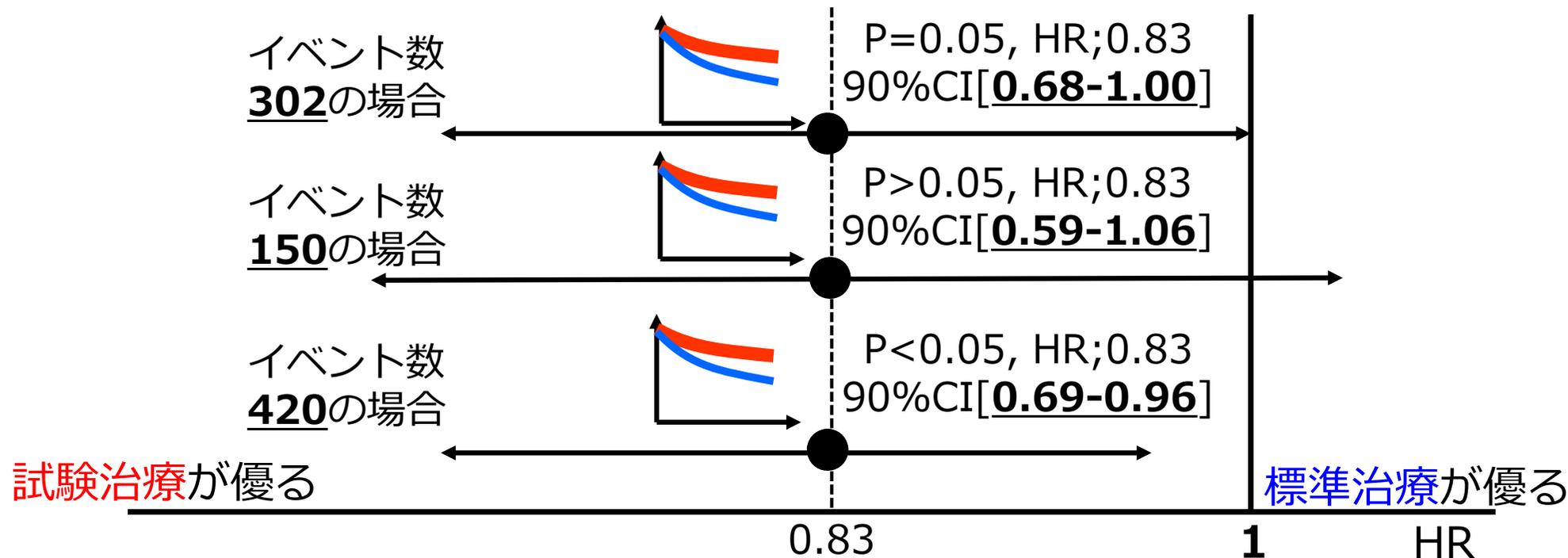
サンプルサイズ計算式
のバラツキに対応

生存時間解析の検出力はイベント数で決まる

- 10000例登録してイベント数100
- 1000例登録してイベント数100
- 100例登録してイベント数100

全て検出力は同じ
(HRの点推定値が同じ
なら、信頼区間も同じ)

「何例登録したか？」より、「何例イベントが起きたか？」が重要
信頼区間の幅はイベント数で決まる



生存時間解析のサンプルサイズ計算手順

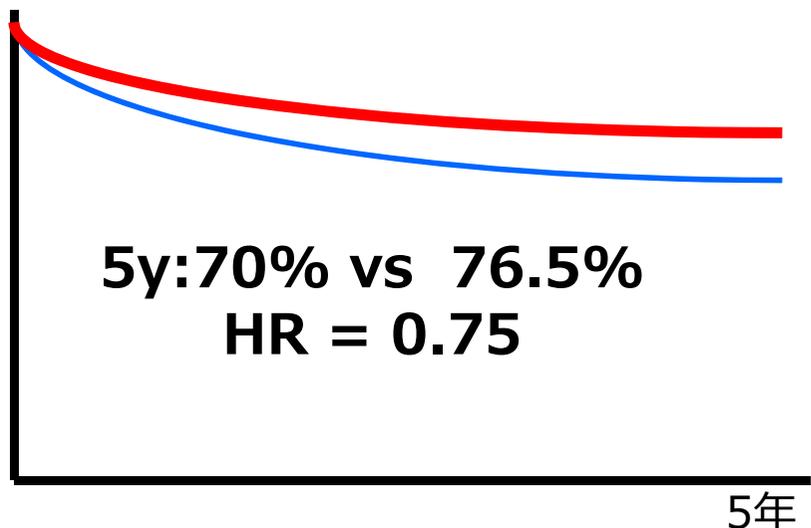
$$N = \frac{(Z_{\alpha} + Z_{\beta})^2}{\ln(HR)^2} \cdot \left(\begin{array}{l} \text{予後と登録 / 追跡期間を} \\ \text{考慮した値} \end{array} \right)$$

① α 、 β 、 Δ で
必要イベント数を求める

② 必要イベント数を得る
ために必要なNを求める

380イベントを観察するために必要な解析対象数

片側 $\alpha=2.5\%$, 検出力 80% , $HR=0.75$ の場合、約380イベント必要



・ 群をプールした5y $\doteq 73\%$

→ 登録した患者のうち、27%に
イベントが起きれば380イベント

→ 必要解析対象数 = $380 \div 0.27 \doteq 1400$ 例



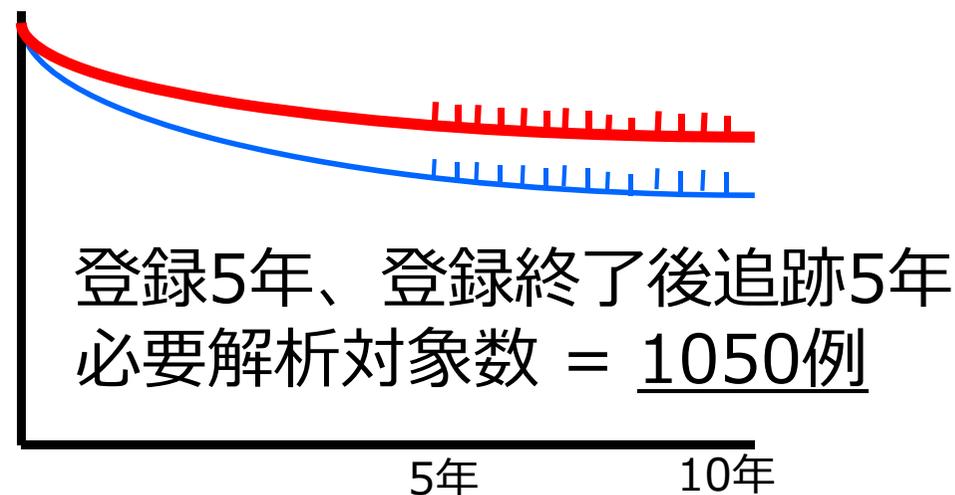
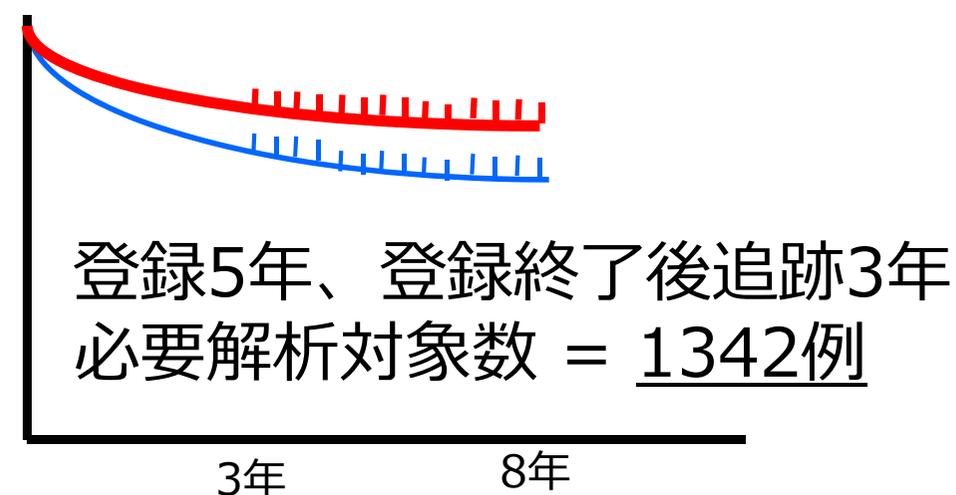
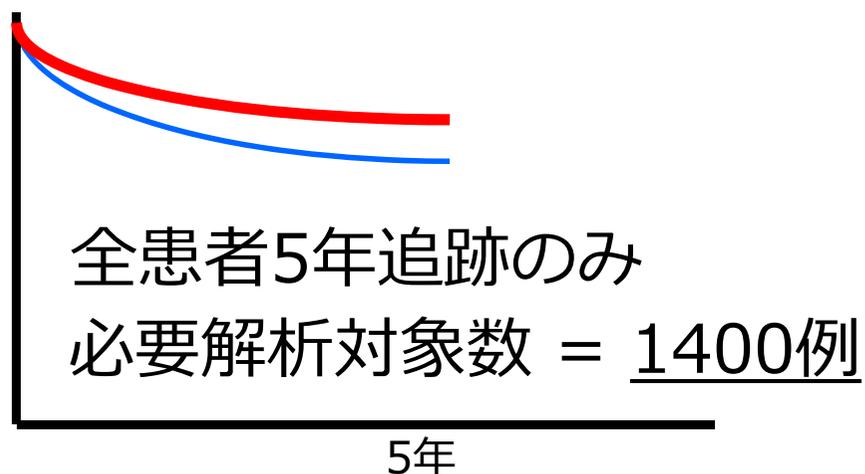
・ 登録された患者はほぼ全員イベント有

→ 必要解析対象数 \doteq 必要イベント数

→ 必要解析対象数 $\doteq 380$ 例

登録/追跡期間中に起こるイベントを考慮

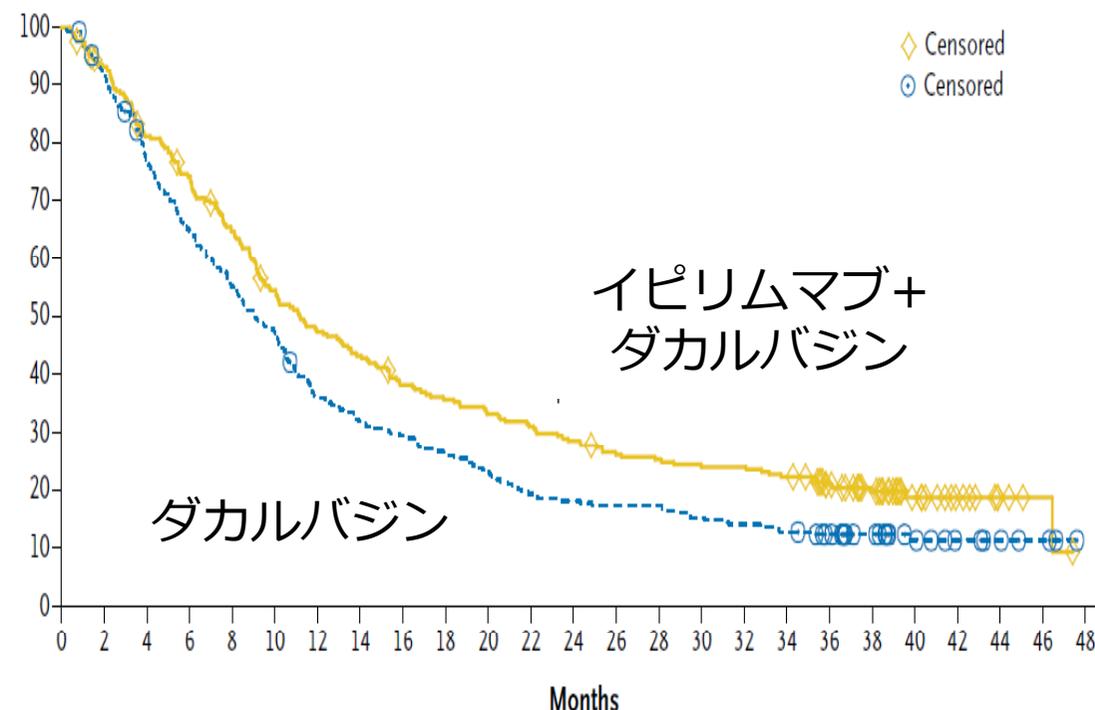
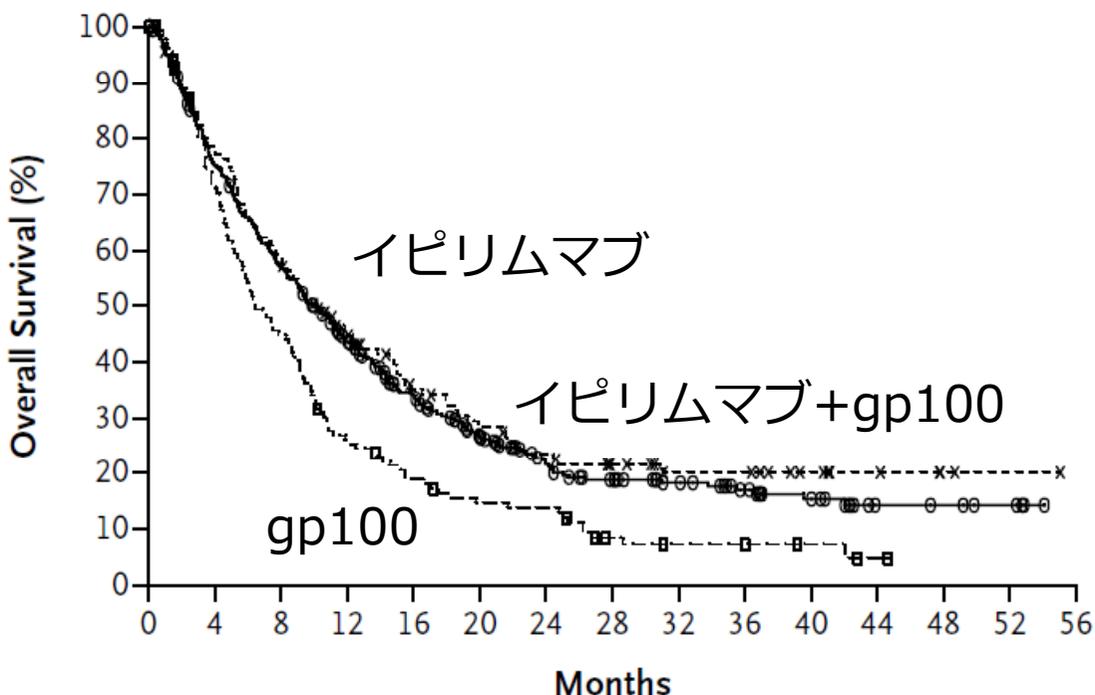
片側 $\alpha=2.5\%$, 検出力 80% , $HR=0.75$ の場合、約380イベント必要
 5y: 70% vs 76.5% ($HR=0.75$)



標準的な生存時間解析のサンプルサイズ計算法の仮定

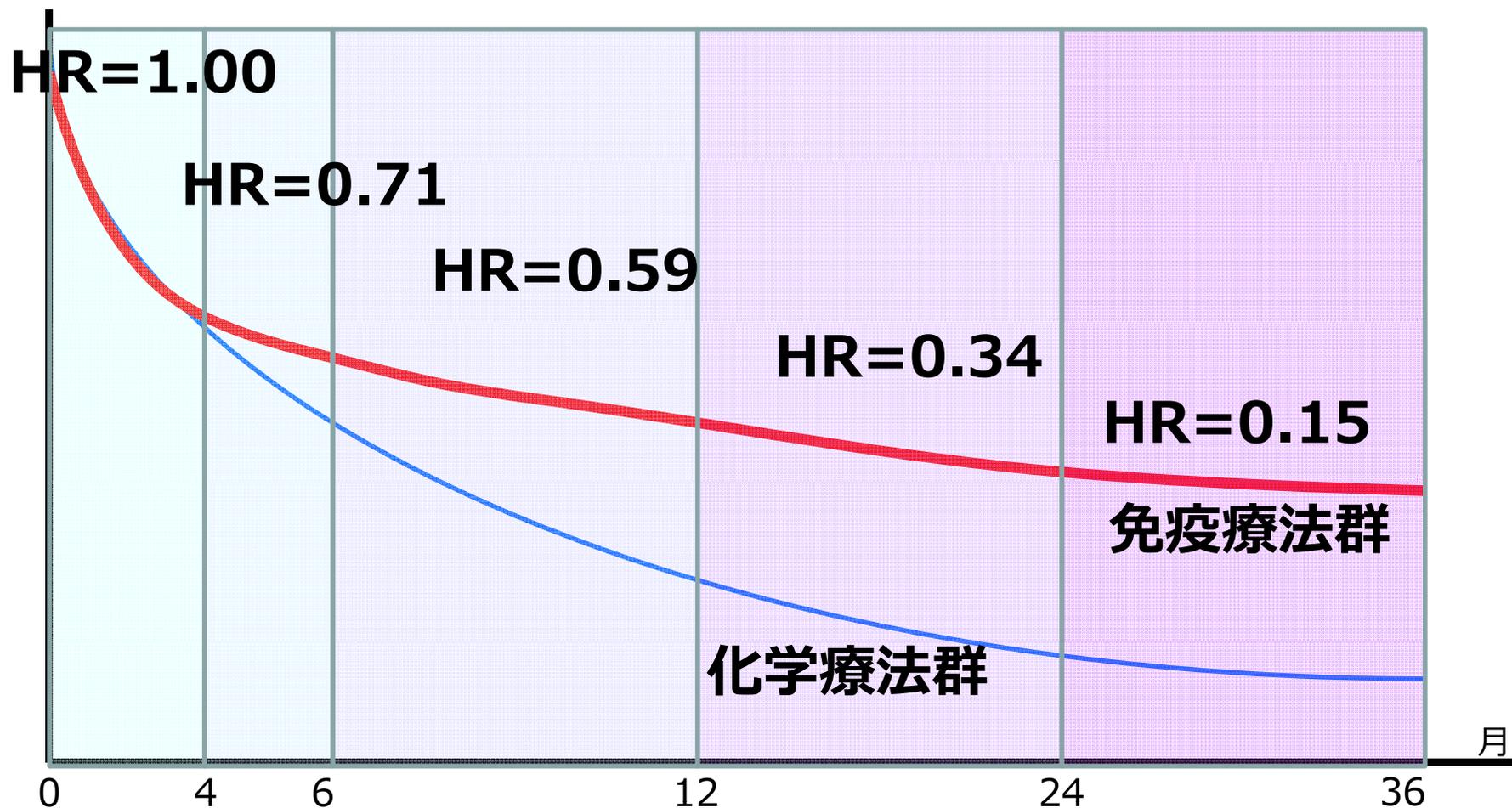
- 両群ともに指数曲線を仮定している
 - 指数曲線が成り立たない場合は？
 - 予後の良い対象は曲線が途中でプラトーになり検出力不足
- 比例ハザード性が成り立っている
 - 比例ハザード性が成り立っていない場合は？
 - ある時点まで曲線が重なり、ある時点以降離れて行く場合は検出力不足

応用：免疫療法のサンプルサイズ計算



- 比例ハザード性が成り立っていない
 - 最初の4か月は効果なし
 - 一定数、治癒する患者が存在
- 両群とも指数曲線を仮定したサンプルサイズ計算では検出力不足に

曲線の区間ごとに別々の指数曲線を想定



まとめ

- サンプルサイズに必要なパラメータは α 、 β 、 Δ 、バラツキ
- 生存時間の群間比較ではNではなくイベント数が重要
- 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)
- 細かい計算法は覚える必要はない
 - 計算はソフトを使えばできる
 - デザインやパラメータの妥当性を統計家と吟味する