

がんを対象とした非劣性試験

JCOGデータセンター 統計部門

若林 将史

JCOG臨床試験セミナー中級編

2023. 10.7 (土)

非劣性って何だろう？

- 『有効性が実薬よりも少なくともある程度以上劣らなければ、薬として世の中に出す価値があるかもしれない』
 - 椿広計, 藤田利治, 佐藤俊哉 編集「これからの臨床試験」より
- 『有効性以外にメリットがあるのであれば、有用であると結論付けるうえで、新薬がより有効であることを示すことは必ずしも必要ではない』
 - JCOGデータセンター訳「米国SWOGに学ぶ がん臨床試験の実践」より
- 『「負けていない」つまり引き分け狙い（中略）「劣性（有意に負けていること）が証明されなかった」ということとは全く違う』
 - 里見清一著「誰も教えてくれなかった癌臨床試験の正しい解釈」より

JCOGにおける非劣性試験

Survival outcomes following laparoscopic versus open D3 dissection for stage II or III colon cancer (JCOG0404): a phase 3, randomised controlled trial

Seigo Kitano, Masafumi Inomata, Junki Mizusawa, Hiroshi Katayama, Masahiko Watanabe, Seichiro Shiochi, Fumi Kashi, Yoshihisa Saida, Hiroshi Hasegawa, Tomonori Akagi, Kenichi Sugihara, Yasuki Fukunaga, Kohji Murata, Masazumi Okajima, Yoshitaka Moriya, Yasuhito Shimoda

Summary
Background Although benefits of laparoscopic surgery compared with open surgery survival of patients undergoing laparoscopic surgery for colon cancer require unclear. We did a randomised controlled trial to establish non-inferiority of laparoscopic surgery.

Methods We did an open-label, multi-institutional, randomised, two-arm phase 3 trial. Patients aged 20–75 years who had histologically proven colon cancer, tumours sigmoid, or rectosigmoid colon; T3 or deeper lesions without involvement of metastasis stage M0; and tumour size of 8 cm or smaller were included. Only an operator or instructor. Patients were randomly assigned (1:1) preoperatively to open or a laparoscopic route, via phone call or fax to the Japan Clinical Oncology Group. Randomisation used a minimisation method with a biased-coins algorithm (ascum, ascending or sigmoid, rectosigmoid) and institution. The primary endpoint was overall survival. The non-inferiority margin for the hazard ratio (registered with UMIN Clinical Trials Registry, number C000001015, and ClinicalTrials.gov, number NCT01070732) was 0.73.

Findings Between Oct 1, 2004, and March 27, 2009, 1057 patients were randomised (n=528) or laparoscopic surgery (n=529). 5-year overall survival was 90.4% (95% CI 89.1–91.8) for laparoscopic surgery. Laparoscopic D3 surgery was not superior to open surgery (HR 1.06, 95% CI 0.79–1.41; p=0.73). 65 (13%) patients in the laparoscopic surgery group had grade 2–4 adverse events. Grade 2–4 (15.3%) in the open surgery group vs 14 (3%) in the laparoscopic surgery group; one small intestine bowel obstruction (16.3%) vs 11 (2%). Two treatment-related deaths: one patient died 7 days after surgery (probably due to myocardial infarction), one patient died 7 days after surgery (probably due to myocardial infarction), one patient died 7 days after surgery (probably due to myocardial infarction), one patient died 7 days after surgery (probably due to myocardial infarction), one patient died 7 days after surgery (probably due to myocardial infarction).

Interpretation Laparoscopic D3 surgery was not non-inferior to open D3 surgery in stage II or III colon cancer. However, because overall survival in both groups was similar, laparoscopic D3 surgery could be an acceptable treatment option for patients with stage II or III colon cancer.

Funding National Cancer Center Research and Development Fund, Grants-In-Aid for Scientific Research, Labour Sciences Research Grant for Clinical Cancer Research from the Ministry of Health, Labour, and Welfare of Japan.

Introduction
In the USA and Europe, laparoscopic surgery for colon cancer is accepted as a standard treatment, which has some benefits over open surgery including decreased pain, improved postoperative pulmonary function, reduced postoperative ileus, improved incidence of wound infection, faster recovery, and shorter hospital stay. In addition to these short-term outcomes, the long-term outcome after laparoscopic surgery for colon cancer is comparable to that after open surgery, as shown by the results of several randomised controlled trials.^{1–4} However, these trials were done before the standard procedure in the USA and Europe—to complete mesorectal excision with central vascular ligation—had emerged as a technical weakness: the mesorectal excision is the anatomical plane of dissection that separates the mesorectum from the rest of the mesorectum. The D3 dissection, anatomical lymph node dissection of lymph feeding artery and mesorectum to the rest of the mesorectum, is the anatomical plane of dissection that separates the mesorectum from the rest of the mesorectum. The D3 dissection, anatomical lymph node dissection of lymph feeding artery and mesorectum to the rest of the mesorectum, is the anatomical plane of dissection that separates the mesorectum from the rest of the mesorectum.

www.thelancet.com/gastrohep Published online February 1, 2017 [http://dx.doi.org/10.1016/S2468-2667\(16\)30131-1](http://dx.doi.org/10.1016/S2468-2667(16)30131-1)

Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, non-inferiority trial

Hiroshi Saji, Morihito Okada, Masahiro Tsuboi, Ryu Nakajima, Kenji Suzuki, Keiji Aokage, Tadashi Aoki, Jiro Okami, Ichiro Yoshino, Naohito Okumura, Masafumi Yamaguchi, Norihiko Ikeda, Masashi Wakabayashi, Kenichi Nakamura, Haruhiko Fukuda, Shintaro Tetsuya, Mitsudomi, Shun-ichi Watanabe, Hisao Asamura, on behalf of the West Japan Oncology Group and Japan Clinical Oncology Group

Summary
Background Lobectomy is the standard of care for early-stage non-small-cell lung cancer (NSCLC). The clinical benefits of segmentectomy have not been investigated in a randomised trial setting. We aim to establish non-inferiority of segmentectomy to lobectomy in patients with small-sized peripheral NSCLC.

Methods We conducted this randomised, controlled, non-inferiority trial at 70 institutions in Japan. Clinical stage IA NSCLC (tumour diameter ≤ 2 cm; consolidation-to-tumour ratio ≥ 0.5) were randomised to either lobectomy or segmentectomy. Randomisation was done via the minimisation method for the institution, histological type, sex, age, and thin-section CT findings. Treatment allocation was concealed from investigators and patients. The primary endpoint was overall survival for all randomly assigned patients. Secondary endpoints were postoperative respiratory function (6 months and 12 months), relapse proportion of local relapse, adverse events, proportion of segmentectomy completion, duration of chest tube placement, duration of surgery, amount of blood loss, and the number of staples used. Overall survival was analysed on an intention-to-treat basis with a non-inferiority margin of 0.73. The hazard ratio (HR) and 95% CI of the hazard ratio (HR) and estimated using a stratified Cox regression model. The non-inferiority margin was 0.73. The hazard ratio (HR) and 95% CI of the hazard ratio (HR) and estimated using a stratified Cox regression model. The non-inferiority margin was 0.73.

Findings Between Aug. 10, 2009, and Oct. 21, 2014, 1106 patients (intention-to-treat population) were randomised to lobectomy (n=554) or segmentectomy (n=552). Patient baseline clinicopathological factors were well balanced between the two groups. In the segmentectomy group, 22 patients were switched to lobectomy and one patient underwent resection. At a median follow-up of 7.3 years (range 0–10.9), the 5-year overall survival was 94.7% for segmentectomy and 91.1% for lobectomy (95% CI 88.4–93.2); superiority and non-inferiority were confirmed using a stratified Cox regression model (HR 0.663; 95% CI 0.474–0.927; one-sided p=0.0082 for superiority). Improved overall survival was observed consistently across subgroups in the segmentectomy group. At 1 year follow-up, the significant difference in the reduced forced expiratory volume in 1 sec between the two groups was 3.5% (p<0.0001), which did not reach statistical significance of 10%. The 5-year relapse-free survival was 88.0% (95% CI 84.8–90.3) for lobectomy (HR 0.998; 95% CI 0.753–1.323; p=0.9889) and 87.9% (95% CI 84.8–90.3) for lobectomy (HR 0.998; 95% CI 0.753–1.323; p=0.9889). 30-day or 90-day mortality was observed. One or more postoperative complications of grade 2 or 3 were observed in both groups (142 [26%] patients who received lobectomy, 148 [27%] who received segmentectomy).

Interpretation To our knowledge, this study was the first phase 3 trial to show the benefits of segmentectomy in overall survival of patients with small-peripheral NSCLC. The findings suggest that segmentectomy could be the standard surgical procedure for this population of patients.

Funding National Cancer Center Research and the Ministry of Health, Labour, and Welfare of Japan.

Copyright © 2017 Published by Elsevier Ltd. All rights reserved.

Introduction
Lung cancer is the leading cause of cancer-related deaths worldwide, and the incidence has increased over the past two decades.¹ Surgical resection is the gold standard of treatment for early-stage lung cancer, with lobectomy being the standard since 1960.² To date, only one randomised trial has compared lobectomy with sublobectomy.

www.thelancet.com Vol 399 April 23, 2017



original reports

abstract

Weekly Cisplatin Plus Radiation for Postoperative Head and Neck Cancer (JCOG1008): A Multicenter, Noninferiority, Phase II/III Randomized Controlled Trial

Naomi Kijata, MD, PhD¹; Makoto Tahara, MD, PhD²; Anki Mizusawa, MD³; Takashi Kodaira, MD⁴; Hirofumi Fujii, MD⁵; Tomoko Yamazaki, MD⁶; Hiroki Mibani, MD⁷; Shigemichi Iwae, MD⁸; Yasushi Fujimoto, MD⁹; Yusuke Onozawa, MD¹⁰; Nobuhiko Hanai, MD¹¹; Takenori Ogawa, MD¹²; Hiroki Ham, MD¹³; Nobuyuki Monden, MD¹⁴; Eiji Shimura, MD¹⁵; Shujiro Minami, MD¹⁶; Takashi Fujii, MD¹⁷; Kaoru Tanaka, MD¹⁸; Akhiro Homma, MD¹⁹; Seichi Yoshimoto, MD²⁰; Nobuhiko Oridate, MD²¹; Koichi Omori, MD²²; Tsutomu Ueda, MD²³; Kenji Okami, MD²⁴; Ichiro Ota, MD²⁵; Kiyoto Shiga, MD²⁶; Masashi Sugawara, MD²⁷; Takahiro Aekagawa, MD²⁸; Yuki Saito, MD²⁹; Shigeyuki Muroto, MD³⁰; Yasumasa Nishimura, MD³¹; Kenichi Nakamura, MD³²; and Ryuichi Hayashi, MD³³ and the Head and Neck Cancer Study Group of the Japan Clinical Oncology Group (JCOG-HNCSG)

PURPOSE The standard treatment for postoperative high-risk locally advanced squamous cell carcinoma of the head and neck (LA-SCCHN) is chemoradiotherapy with 3-weekly cisplatin (100 mg/m²). However, whether chemoradiotherapy with weekly cisplatin (40 mg/m²) yields comparable efficacy with 3-weekly cisplatin in postoperative high-risk LA-SCCHN is unknown.

PATIENTS AND METHODS In this multi-institutional open-label phase II/III trial, patients with postoperative high-risk LA-SCCHN were randomly assigned to receive either chemoradiotherapy with 3-weekly cisplatin (100 mg/m²) or with weekly cisplatin (40 mg/m²) to confirm the noninferiority of weekly cisplatin. The primary end point of phase II was the proportion of treatment completion, and that of phase III was overall survival. A noninferiority margin of hazard ratio was set at 1.32.

RESULTS Between October 2012 and December 2018, a total of 261 patients were enrolled (3-weekly cisplatin, 132 patients; weekly cisplatin, 129 patients). At the planned third interim analysis in the phase III part, after a median follow-up of 2.2 (interquartile range 1.19–3.56) years, chemoradiotherapy with weekly cisplatin was noninferior to 3-weekly cisplatin in terms of overall survival, with a hazard ratio of 0.69 (99.1% CI, 0.374 to 1.273 [< 1.32], one-sided P for noninferiority = .0027 < .0043). Grade 3 or more neutropenia and infection were less frequent in the weekly arm (3-weekly v weekly, 49% v 35% and 12% v 7%, respectively), as were renal impairment and hearing impairment. No treatment-related death was reported in the 3-weekly arm, and two (1.6%) in the weekly arm.

CONCLUSION Chemoradiotherapy with weekly cisplatin is noninferior to 3-weekly cisplatin for patients with postoperative high-risk LA-SCCHN. These findings suggest that chemoradiotherapy with weekly cisplatin can be a possible treatment option for these patients.

J Clin Oncol 40:1980-1990. © 2022 by American Society of Clinical Oncology
Creative Commons Attribution Non-Commercial No Derivatives 4.0 License

INTRODUCTION
The global incidence of squamous cell carcinoma of the head and neck (SCCHN) is estimated at more than 700,000 cases newly diagnosed annually.¹ A majority of patients present with locally advanced SCCHN (LA-SCCHN), and surgery is a mainstay of treatment for resectable cases. Postoperative chemoradiotherapy with 3-weekly cisplatin at 100 mg/m² is standard treatment for patients with high-risk factors for recurrence.^{2,3} However, the 3-weekly dosage has raised concerns about insufficient cisplatin delivery because of high-dose-related toxicity,^{2,3,5} and chemoradiotherapy with weekly cisplatin is widely used as a possible alternative with a better safety profile.⁶⁻¹³ Although some results have conflicted, meta-analyses of comparisons between the 3-weekly and weekly cisplatin suggest that the two approaches have equal efficacy.¹⁴⁻¹⁹

Here, we conducted a multi-institutional open-label phase II/III trial to investigate whether chemoradiotherapy with weekly cisplatin at 40 mg/m² was noninferior to 3-weekly cisplatin at 100 mg/m² in terms of overall survival (OS) for postoperative high-risk LA-SCCHN (JCOG1008).

本講義の目的

- がんの非劣性試験をデザインする上でのポイントを理解すること
 - 非劣性マージンの決め方
 - ハザード比と生存割合の関係
 - 解析対象集団と分析感度
- 非劣性試験のtopicや最近の動向を把握すること
 - システムティックレビューの紹介

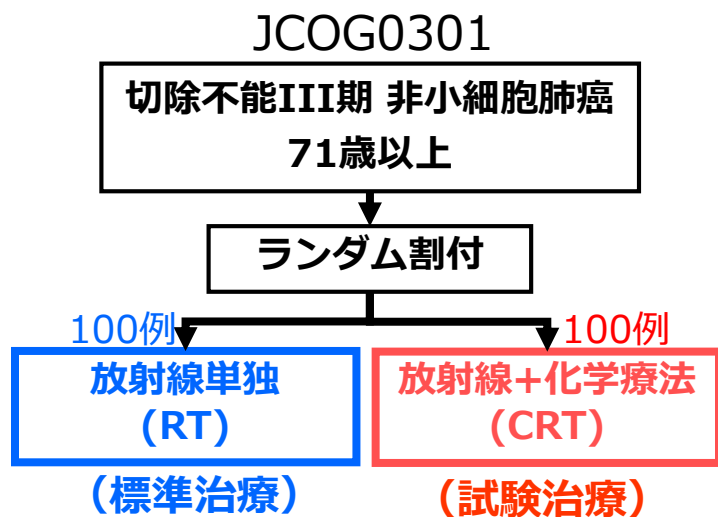
本講義のoutline

- 非劣性試験のデザイン
- 非劣性試験に関するtopic
- システムティックレビュー

本講義のOutline

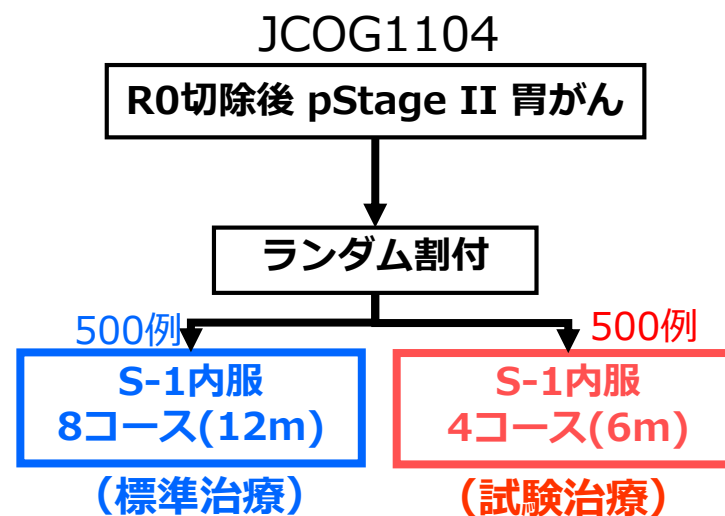
- **非劣性試験のデザイン**
- 非劣性試験に関するtopic
- システムティックレビュー

優越性試験と非劣性試験



優越性試験

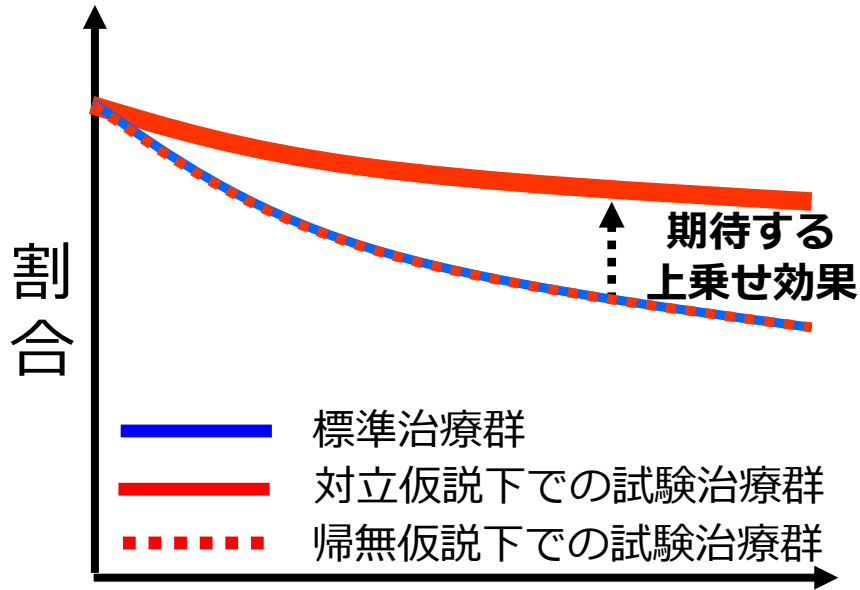
- 試験治療は有効性で上回らなければいけない
- 試験治療は安全性等で標準治療に劣る(Toxic new)



非劣性試験

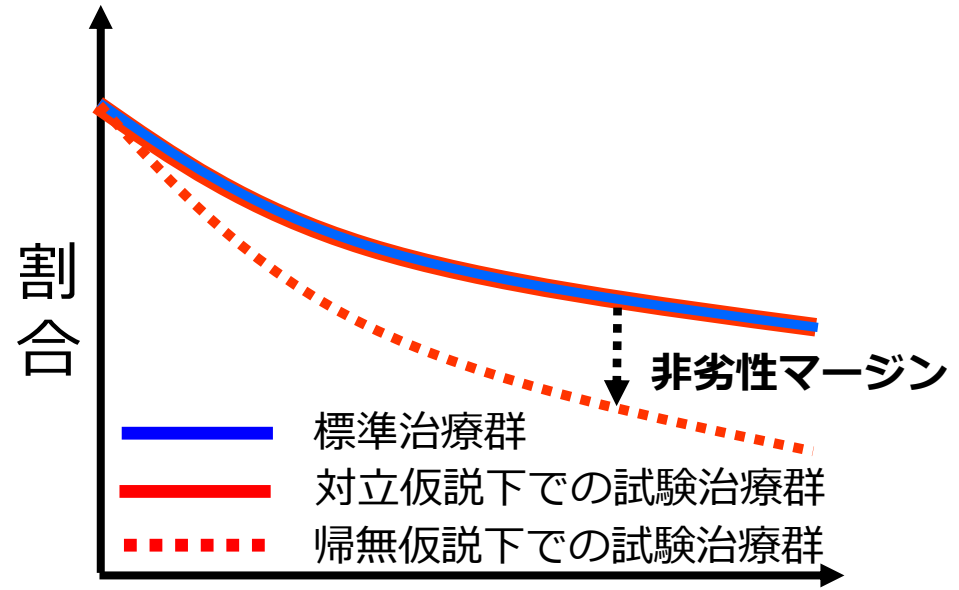
- 試験治療は有効性で一定以上劣らなければ良い
- 試験治療は安全性等で標準治療に優る(Less toxic new)

優越性試験と非劣性試験



優越性試験

- 標準治療に対して上乗せを期待
- 生存曲線が重なっているかどうか（帰無仮説）の検定を行う



非劣性試験

- 標準治療に対して同じであることを期待
- 生存曲線が一定以上劣っていないかどうか（帰無仮説）の検定を行う

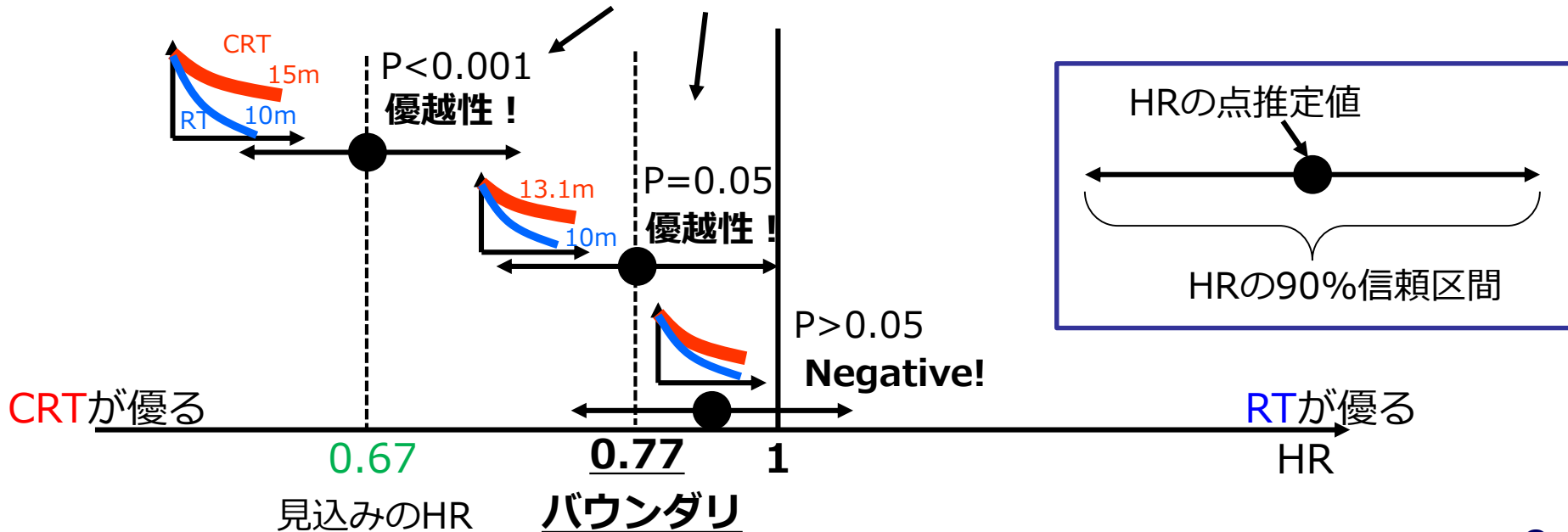
意思決定方法：優越性試験の場合

JCOG0301の例

- 片側 $\alpha=5\%$, 検出力=80%
- RT vs CRTのMST : 10m vs 15m (HR=0.67)

- 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

HR=1ではなく、CRTが上回っていることが示された

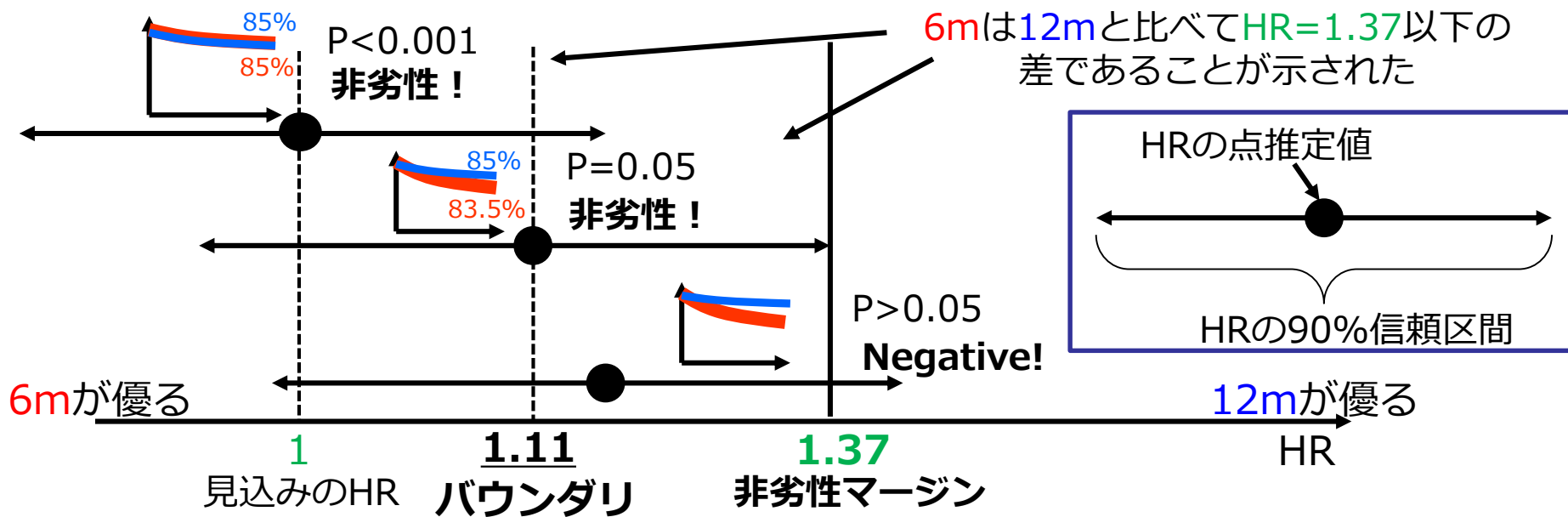


意思決定方法：非劣性試験の場合

JCOG1104の例

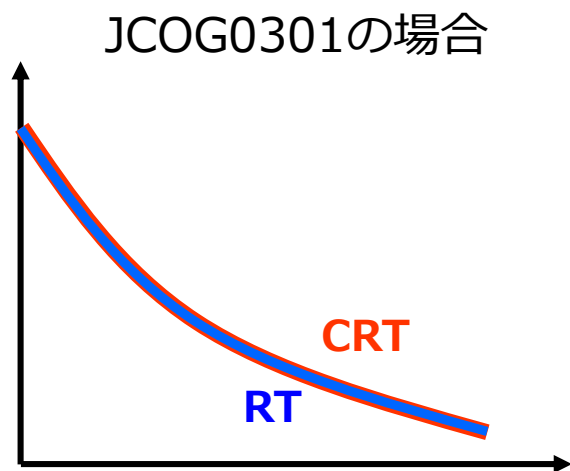
- 片側 $\alpha=5\%$, 検出力=80%
- 12m vs 6m の3yRFS : 85% vs 85% (HR=1)
- HRの非劣性マージン : **1.37** (3yRFSで5%に相当)

- 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

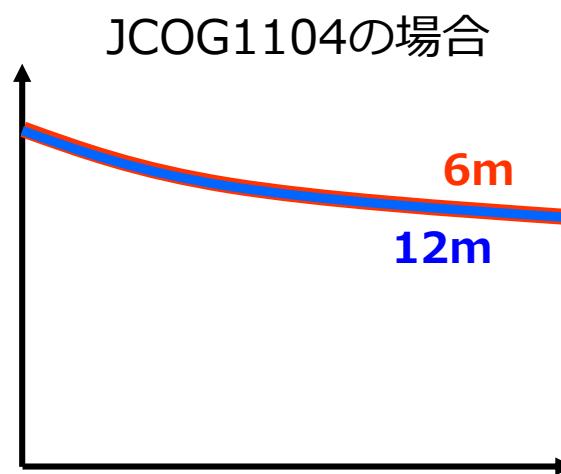


優越性/非劣性の決め方

- 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**

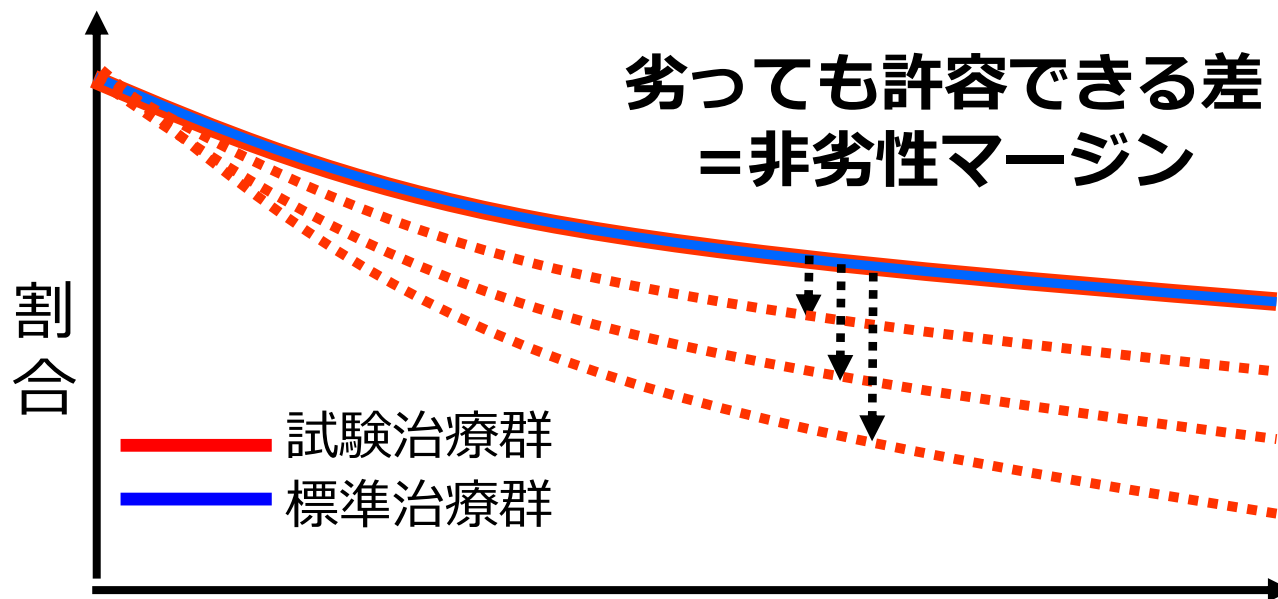


毒性が強く、治療に手間がかかる
CRT群 (Toxic new)と**RT群**のOSが
同じなら、標準治療は**RT群**
→ **優越性試験**



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**

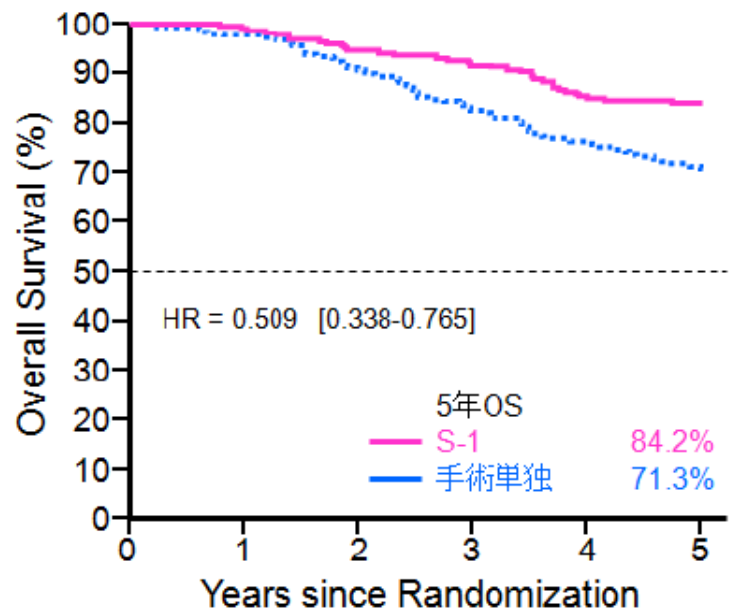
非劣性マージンの決め方



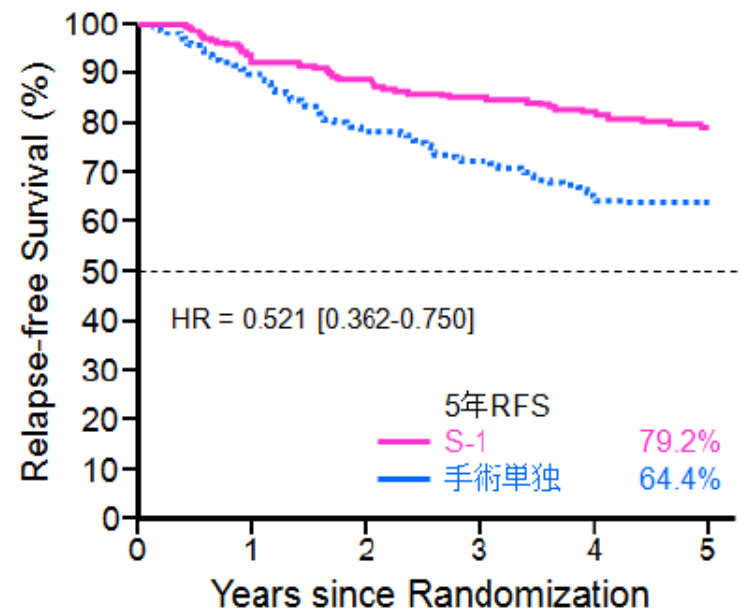
- 臨床の観点を重視
 - 低侵襲性の内容とのトレードオフで研究者が許容できる大きさの考慮
 - 統計の観点を重視
 - リスクの4/3倍 (HR:1.33) といった決め方や、標準治療とプラセボの中間等
 - 非劣性試験を繰り返す or マージンを際限なく下げてしまうと…
 - プラセボ (無治療) を下回るマージンを引くことになる
- バイオクリープ (Bio creep)

非劣性マージンの決め方(JCOG1104を例に)

ACTS-GC試験におけるOSとRFS



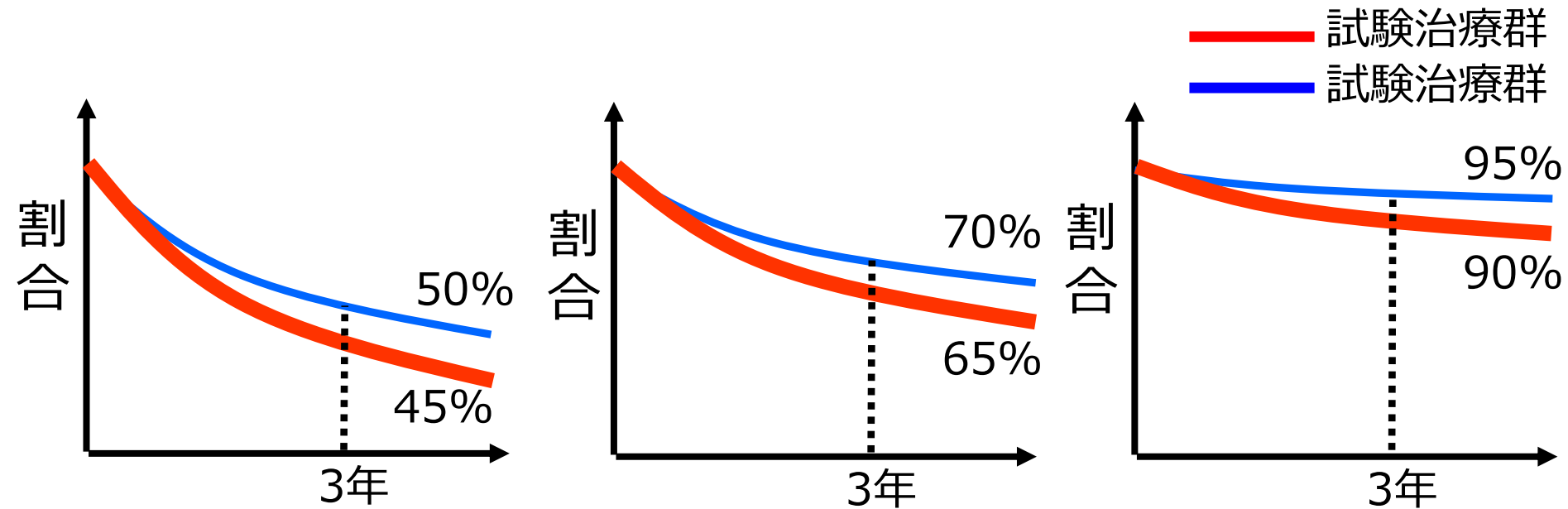
No. at risk	0	1	2	3	4	5
S-1	232	229	219	208	188	163
手術単独	233	226	206	183	162	140



No. at risk	0	1	2	3	4	5
S-1	232	212	199	182	170	101
手術単独	233	206	176	154	132	82

- 既報であるACTS-GC試験の5年OSと3年RFSを参照
 - リスク（再発）に見合うベネフィット（治療の負担の軽減）より-5%に設定
- (仮に) -5%よりも下げた場合に必要となる議論
 - -15%：手術単独群を下回ることを許容することになる（Bio creep）
 - -10%：手術単独群に+4%の上乗せと6か月のS-1投与が釣り合うのか？

ハザード比と生存割合の差の関係

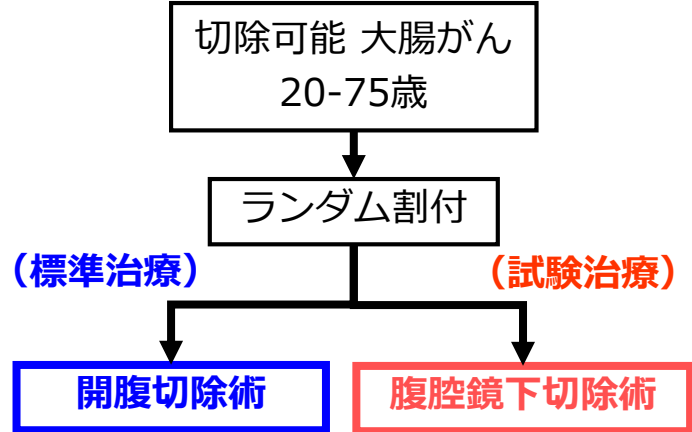


$$HR=1.152 < HR=1.208 < HR=2.054$$

- 予後良好な対象に対して5%の差をつける方が難しい
→相対的な差 (HR) は大きくなる
- 同じ5%でも臨床的な解釈は異なるはず

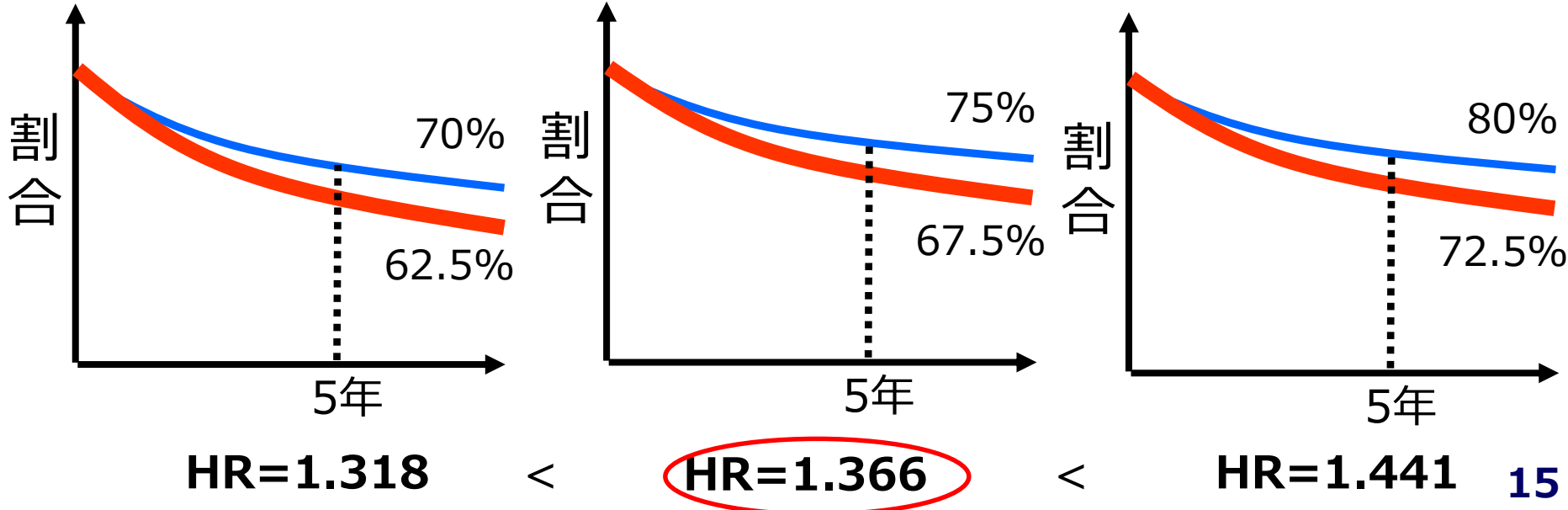
ハザード比 or 生存割合の差？

JCOG0404



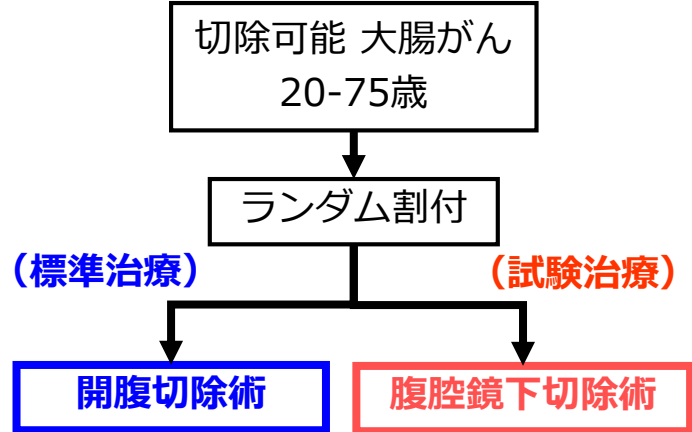
- 【サンプルサイズ計算時に想定したパラメータ】
- ・ 期待5年OS：70-80%
 - ・ 試験治療群に許容される5年OS：-7.5%

「ハザード比<1.441は非劣性の許容域としては大きすぎると考えられるため、許容域を5年生存割合が75%であった場合のハザード比<1.366とする。」



ハザード比 or 生存割合の差？

JCOG0404



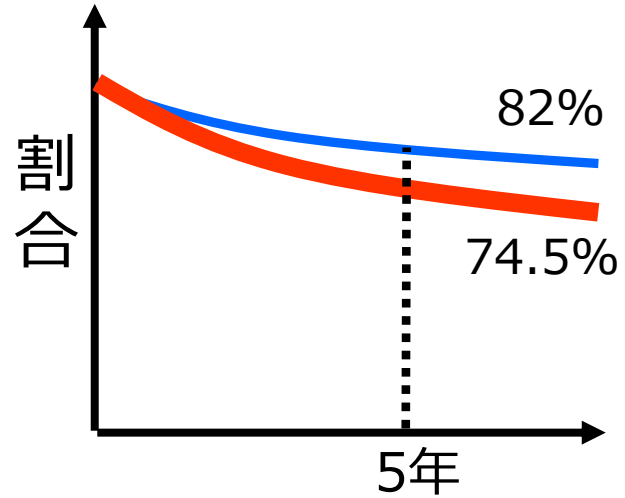
【サンプルサイズ設定】

- ・期待5年OS 75%,非劣性マージン 1.366, サンプルサイズ 818



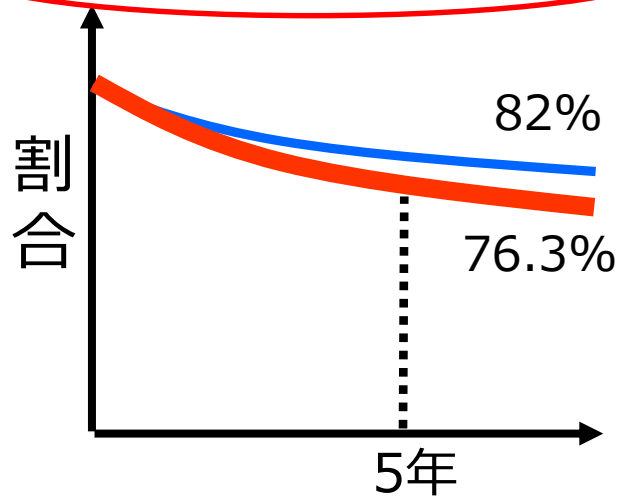
予後良好であったため、
期待5年OSを82%に上方修正したい

【割合の差-7.5%を固定した場合】



HR=1.483
必要サンプルサイズ：640

【HR=1.366を固定した場合】



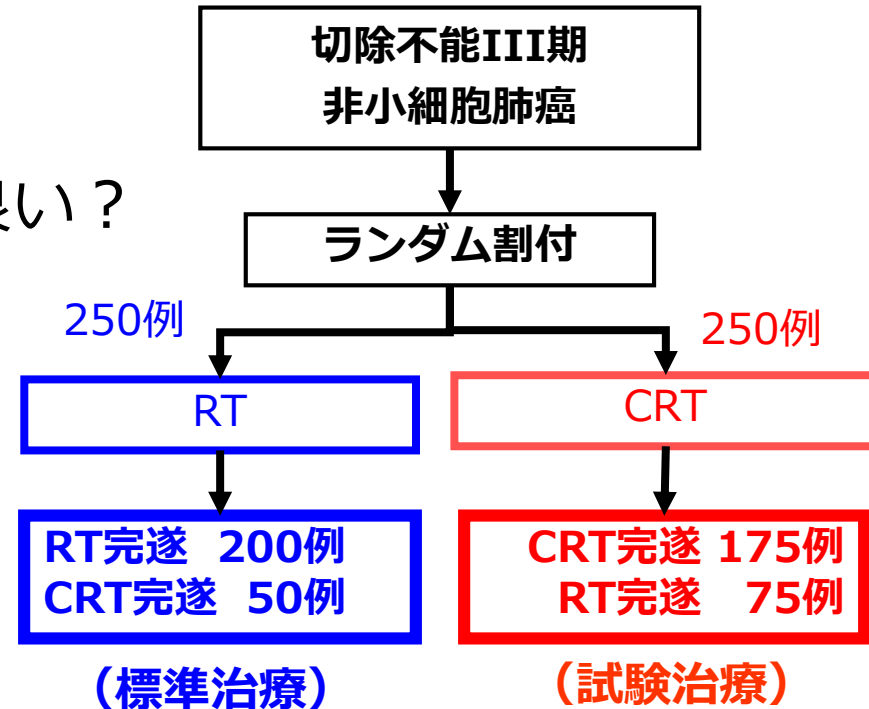
HR=1.366
必要サンプルサイズ：1022

>
<

ITTかPPSか？

注：仮想例

- 簡単のために右のような状況の場合どの解析法を選ぶのが良い？



- ① 治療完遂例どうしで比較

RT 200例 vs. CRT 175例

- ② 実際に行った治療同士で比較

RT (200+75)例 vs. CRT (175+50)例

- ③ ランダム化で割り付けられたどうしで比較

RT 250例 vs. CRT 250例

優越性検定の場合

RTに対するCRTの優越性試験の場合・・・

差がないのに誤って差がある
と書いてしまふ確率up

- ① 治療完遂例どうして比較 RT 200例 vs. CRT 175例
 - 未完遂の人は予後の悪い人なので、残った人は元気な人
- ② 実際に行った治療同士で比較RT(200+75)例 vs. CRT(175+50)例
 - CRT群に割り付けられたけどRTしか出来なかった人は予後の悪い人、RT群に割り付けられたけどCRTをした人は元気な人
→①、②の比較では群間で背景因子が揃わなくなり、ランダム化した意味がなくなってしまう

差がない時は
正しく差がないと書ける

- ③ ランダム化で割り付けられたどうして比較RT 250例 vs. CRT 250例
 - 本当にCRTに効果がある場合、CRT群に割り付けられたけどRTをした人がいると、治療効果が薄まってしまふ

Intention-to treat解析(ITT解析)

- ランダム化によって割り付けられた通りの治療群で行う解析(③の対象で解析する方法) のこと
 - **優越性試験**では、**ITT解析**をすれば、 α エラーを起こす確率は大きくなる
 - 有意になりにくいという意味で「保守的(conservative)」な解析法
- ITT解析でも有意差があったのなら、自信を持って差があると判断できる
 - 「少なくともXXXくらいの治療効果がある！」と言える
 - ITT解析が優越性試験の主たる解析の標準的方法

非劣性試験では？

- 治療未完遂例の存在や治療のクロスオーバー等により群間差が薄まる方向にバイアスがかかる
 - 非劣性が検証されやすくなる
 - 無効な治療が誤って非劣性であると判断される（**分析感度**がない）状況になる
 - 非劣性試験ではITTよりも**PPS（割付られた群の治療を施行した集団）**や**全適格例**による解析の方が望ましいとする考え方もある
 - JCOG試験も基本的にPPS
 - JCOGポリシー07 統計的原則と試験デザイン 5.1節

分析感度

有効な治療と有効性の低いあるいは無効な治療とを区別する力
(ICH E10より)

- 以下の2点が満たされているかどうかで推測される
 - 同様のデザインで行われた過去の試験で、有効/無効を区別できた
 - 試験の実際の施行において有効/無効を区別する力を低下させなかった

→**解析対象集団や適格/除外規準等の試験デザインに大きく異存**
- 分析感度がないと考えられる例
 - 高齢者を対象とした試験で他病死がイベントの多数を占める
 - 術後補助化学療法の試験で、RFSではなく（二次がんを含めた）DFSで比較することで再発の差が薄まる
 - 両群で有害事象中止が頻発し、施行された後治療が群間でほぼ等しい
 - バイオクリープ (Bio creep)

低侵襲性に関する意思決定規準も必要か？

- 低侵襲の規準を設定する場合がある
 - 術後呼吸機能の中央値の差として10% (JCOG0802/WJOG4607L)
 - QoLの差として10 pointsの差
 - Cosmetic outcomeとして40%の差 等々
- 低侵襲であることが明らかと考えられる場合は設けないことが多い
 - 胃癌に対する術後補助化学療法の投与期間の短縮 (JCOG1104)
 - 胃癌に対する脾臓の温存 (切除範囲の縮小) (JCOG0110)
 - 卵巣癌に対する術前補助化学療法 (手術回数の削減) (JCOG0602)
 - 大腸がんに対する補助化学療法の投与経路の違い (JCOG0205) 等々
- 規準は設定しなくても評価項目として検討することが望ましい
 - 有害事象発生割合
 - 手術時間、術中出血量、入院期間 等々

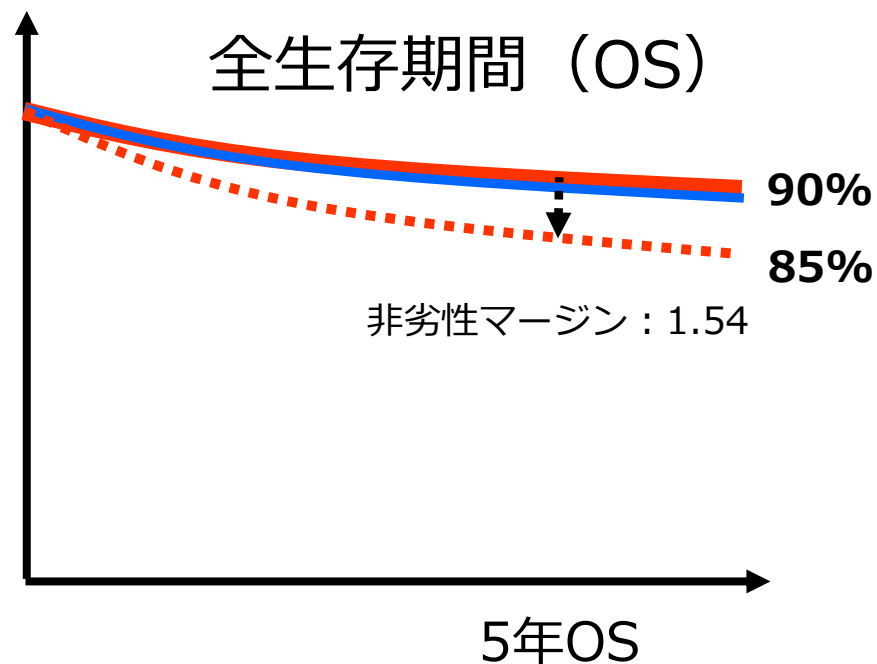
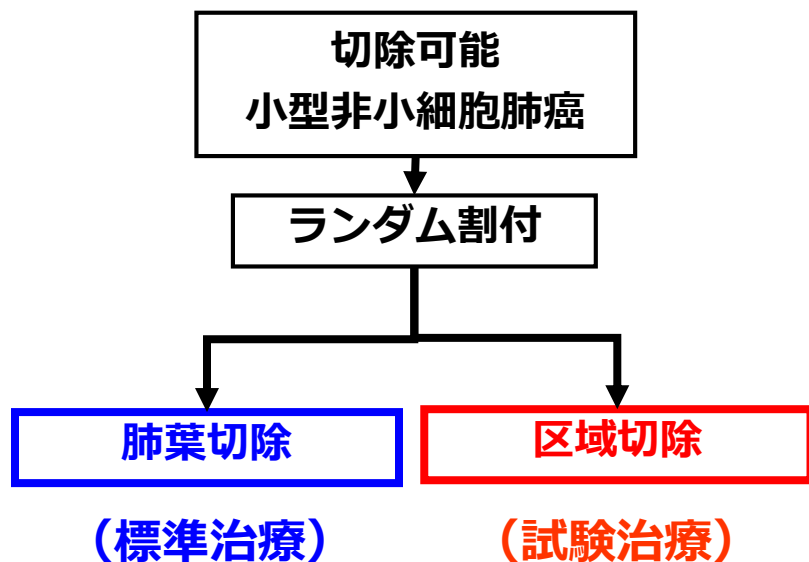
Yoshikawa *et al.* (2019) *Lancet Gastroenterol Hepatol* 4(3):208-216.

Sano *et al.* (2017) *Ann Surg* 265(2):277-283.

Onda *et al.* (2020) *Eur J Cancer* 130:114-125.

Shimada *et al.* (2014) *Eur J Cancer* 50(13):2231-40.

JCOG0802/WJOG4607L



- 呼吸機能を温存し術後のQoLが相対的に向上 (ベネフィット) が期待されるが術後再発のリスク (非劣性試験)
- OSの非劣性が示された新たな標準治療と位置付けることはできない

→術後1年時の呼吸機能で中央値として10%の差が必要

試験の意思決定ルール

1) 区域切除が肺葉切除に対して、**全生存期間において劣っていない事が証明され、かつ優越性が示された場合**

→**区域切除**を新たな標準治療と判断する。

2) 区域切除が肺葉切除に対して、**全生存期間において劣っていない事が証明され、かつ術後 1 年後の呼吸機能で区域切除の低侵襲性が証明された場合**

→**区域切除**をより有用な標準治療と判断する。ただし、肺葉切除を標準治療の 1 つのオプションとするかどうかはその他のエンドポイントから総合的に評価して決定する。

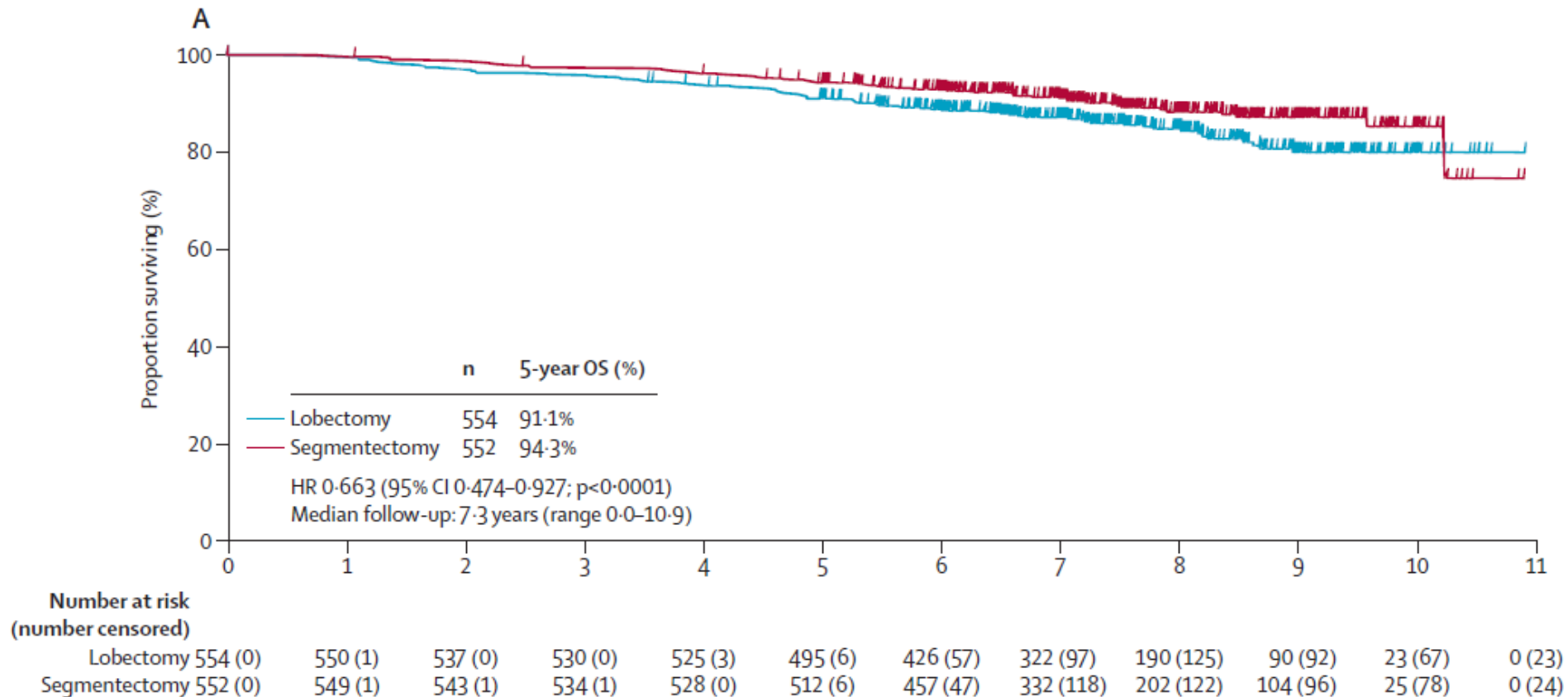
3) 区域切除が肺葉切除に対して、**全生存期間において劣っていない事が証明され、かつ術後 1 年後の呼吸機能で区域切除の低侵襲性が証明されなかった場合**

→引き続き**肺葉切除**をより有用な標準治療と判断する。ただし区域切除を標準治療の 1 つのオプションとするかどうかはその他のエンドポイントから総合的に評価して決定する。

4) 区域切除が肺葉切除に対して、**全生存期間において劣っていない事が証明されなかった場合**

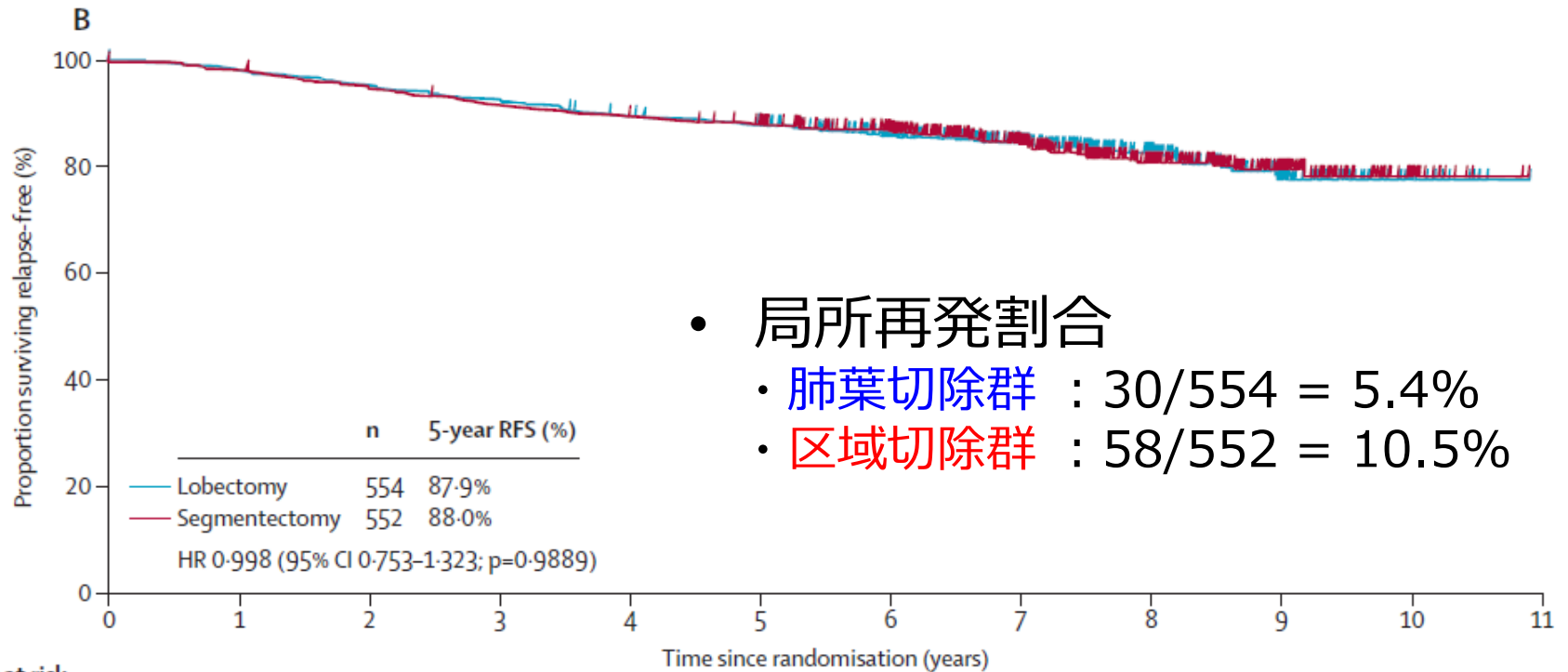
→引き続き**肺葉切除**を標準治療と判断する。

Primary endpoint : OS



Secondary endpoint

【Relapse-free survival (RFS)】



- 局所再発割合

- 肺葉切除群 : 30/554 = 5.4%
- 区域切除群 : 58/552 = 10.5%

	0	1	2	3	4	5	6	7	8	9	10	11
Number at risk (number censored)												
Lobectomy	554 (0)	542 (1)	527 (0)	512 (0)	492 (3)	477 (6)	409 (57)	310 (93)	184 (121)	85 (91)	22 (63)	0 (22)
Segmentectomy	552 (0)	541 (1)	521 (1)	503 (1)	491 (0)	477 (6)	426 (45)	304 (112)	181 (112)	89 (90)	21 (67)	0 (21)

- RFSイベントの内訳

- 肺葉切除群 : 再発44例、**無再発死亡53例** (計97イベント)
- 区域切除群 : **再発67例**、無再発死亡29例 (計96イベント)

Secondary endpoint

【術後呼吸機能 (n=1054)】

- 術後1年時のFEV1.0 (1秒量) 中央値
区域切除 -8.5% vs 肺葉切除 -12.0% (p<0.001)
- 術後1年時のFVC 中央値
区域切除 -7.4% vs 肺葉切除 -10.7% (p<0.001)

p値は高度な差を示したが…事前規定した判断規準 (≥10%) は満たさず

- OSの優越性が検証されたため、(プロトコールに従うと)
区域切除を標準治療と結論するが…
- 区域切除を標準治療とするか、オプションの1つに留めるか、
現在も呼吸器外科領域にてディスカッション中

本講義のOutline

- 非劣性試験のデザイン
- **非劣性試験に関するtopic**
- システムティックレビュー

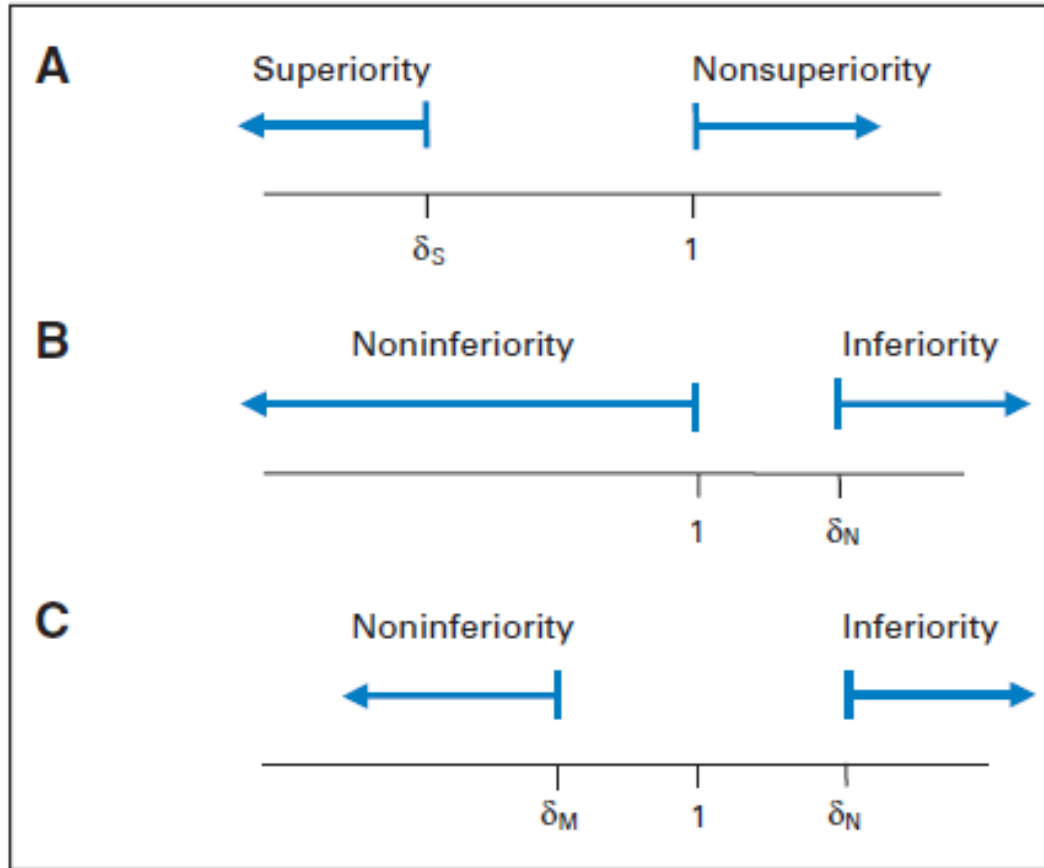
Hybrid design

Randomized Clinical Trial Design for Assessing Noninferiority When Superiority Is Expected

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

A B S T R A C T

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.



優越性試験

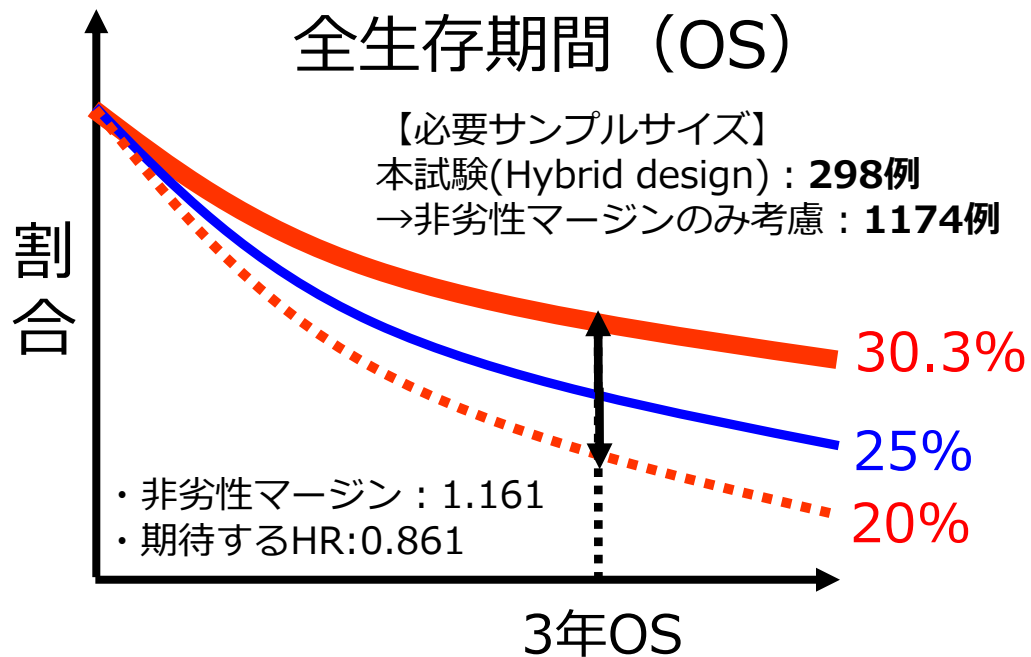
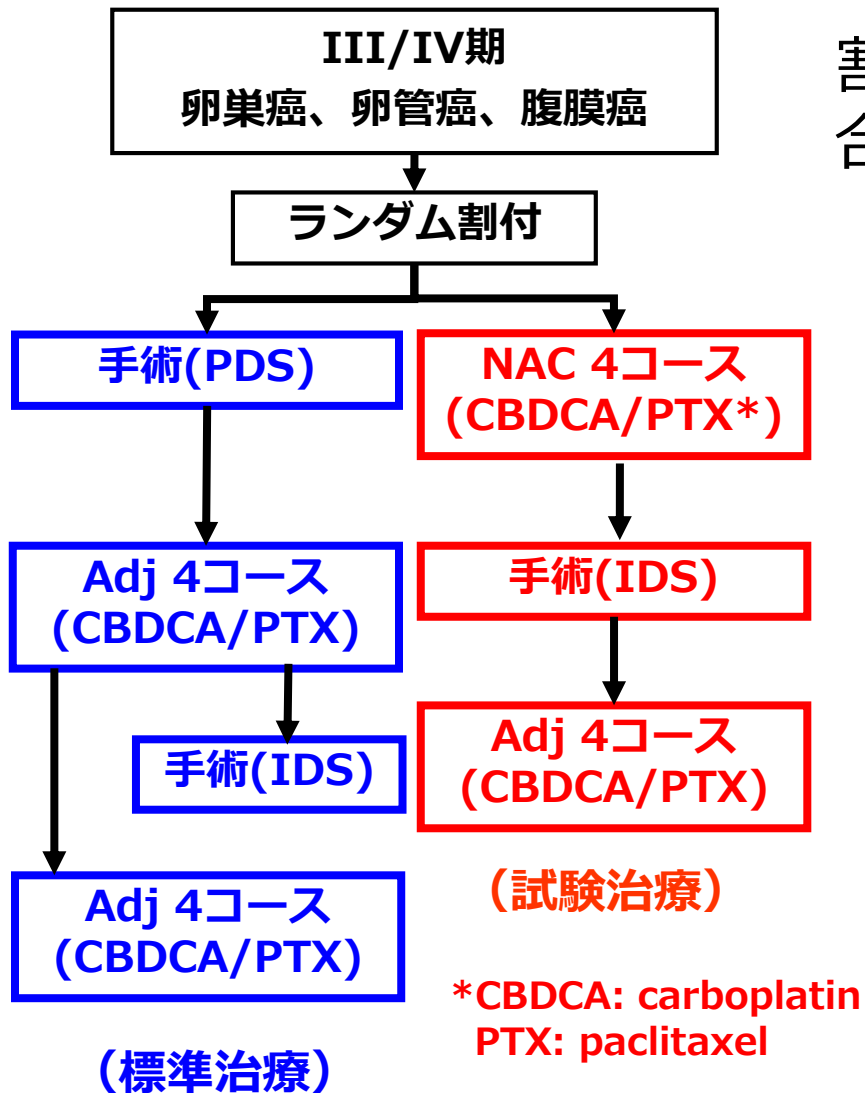
非劣性試験

**非劣性試験
(Hybrid design)**

Fig 1. Regions of true hazard ratio (standard over experimental) for which one would want (A) high probability of declaring superiority or not in a superiority trial, (B) noninferiority or not in a noninferiority trial, and (C) noninferiority or not in a hybrid design.

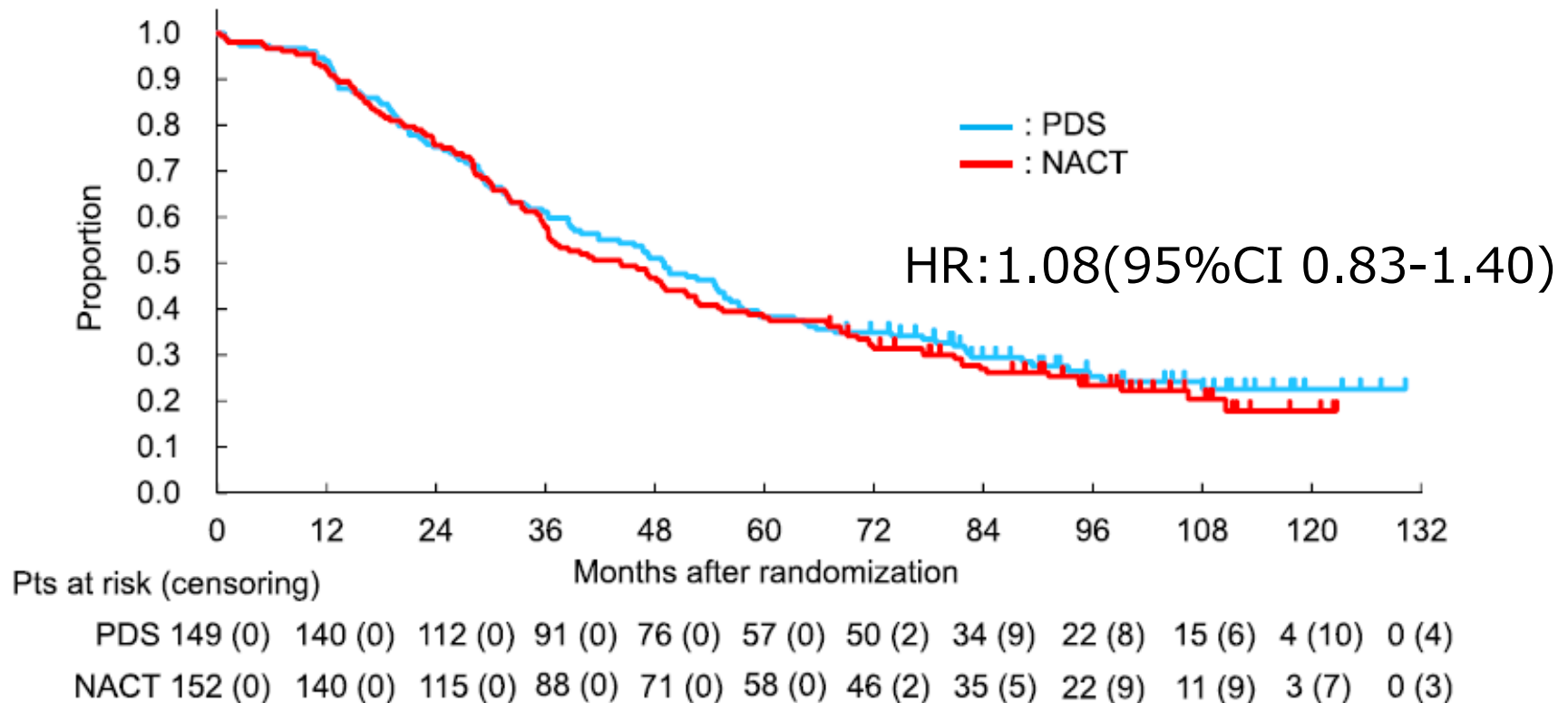
- 基本的な考え方および検定方法も含めて非劣性試験
 - 非劣性マージン (δ_N) に加えて上乗せ (δ_M) を期待することで必要サンプルサイズを小さく抑えることが可能

JCOG0602



- 試験治療群の手術回数が1回で済む可能性が高い(非劣性)
- 既報より、NAC導入前と比較して、NACにより3年OSが16%改善(優越性)
- バウンダリ : HR=0.953

JCOG0602の結果



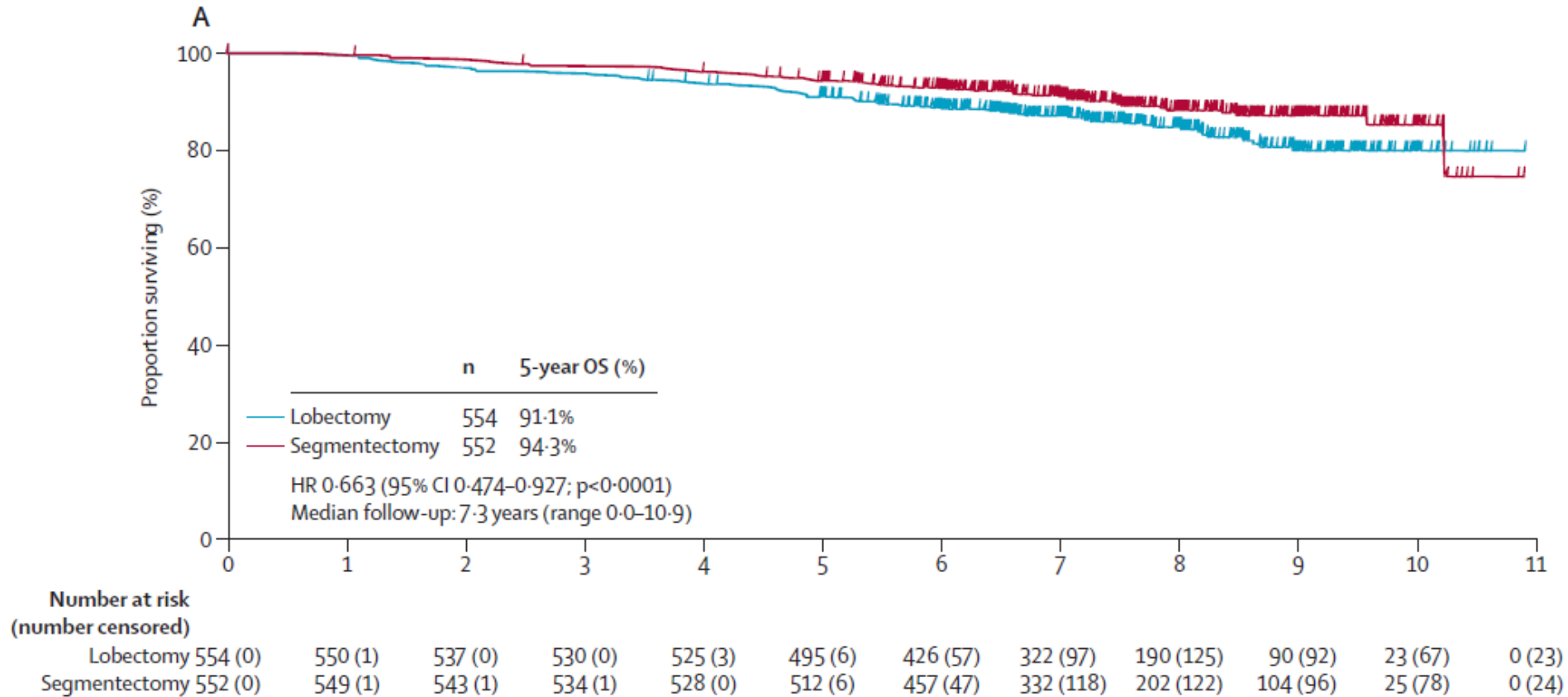
- OSの非劣性を検証できず

- 別論文 (Onda et al. (2016) Eur J Cancer) にて低侵襲性が示されたことも報告されており、日常臨床にてNAC→IDSは行われている

非劣性検定→優越性検定は妥当か？

- 非劣性が検証されなかった場合の優越性検定
 - 非劣性が検証された場合に限り統計学的に妥当
 - Gate keepingの考え方(帰無仮説が棄却された場合 α を消費しない)
 - 非劣性に加えて優越性まで検証された試験
 - JCOG0802/WJOG4607L, JASPAC-01 等
- 優越性が検証されなかった場合の非劣性検定
 - 統計学的/臨床的な問題あり
 - 多重性の問題で試験全体の α エラーが計画時を上回る
 - 優越性が検証目的の試験で非劣性を示すことの臨床的な意義

再掲 : JCOG0802/WJOG4607L



試験の意思決定&検定の手順

【意思決定&検定の手順】

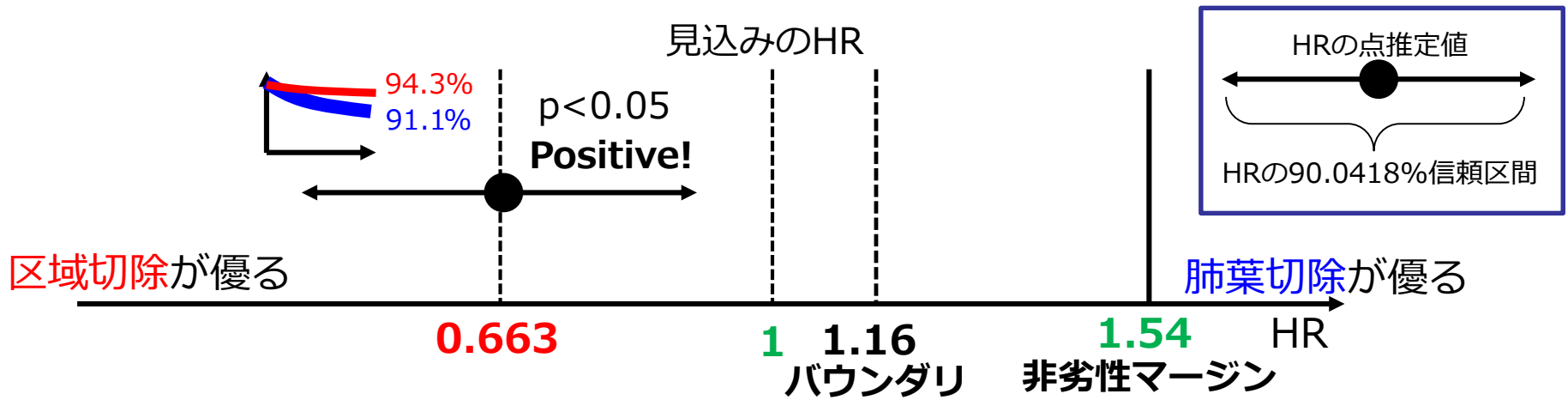
区域切除の肺葉切除に対するHR 0.663 (90.0418%*信頼区間 0.500-0.879)

①非劣性検定：信頼上限0.879 < 1.54のため区域切除の非劣性が検証された



②優越性検定：信頼上限0.879 < 1のため区域切除の優越性も検証された

*中間解析での優越性検定で α を消費したため、多重性を考慮した片側 $\alpha=4.9791\%$ を有意水準として主たる解析を実施



本講義のOutline

- 非劣性試験のデザイン
- 非劣性試験に関するtopic
- システムティックレビュー

Question

非劣性試験の解析や論文作成等を通じて

具体的な低侵襲の指標はどのようなものがあるのだろうか？

低侵襲に関する評価項目や具体的な達成基準は設けているのだろうか？検定は行われているのか？

低侵襲に関する結果が記述のみで図表として掲載されていないのはどのくらいの頻度であるのか？



非劣性が統計的に達成できなかった原因は？
検出力不足？マージンの見込み違い？

統計的な結論と臨床的な結論が食い違うのはどの程度あるのだろうか？

外科系と内科系で試験デザインや論文の書き方に違いがあるのだろうか？

システマティックレビューの目的

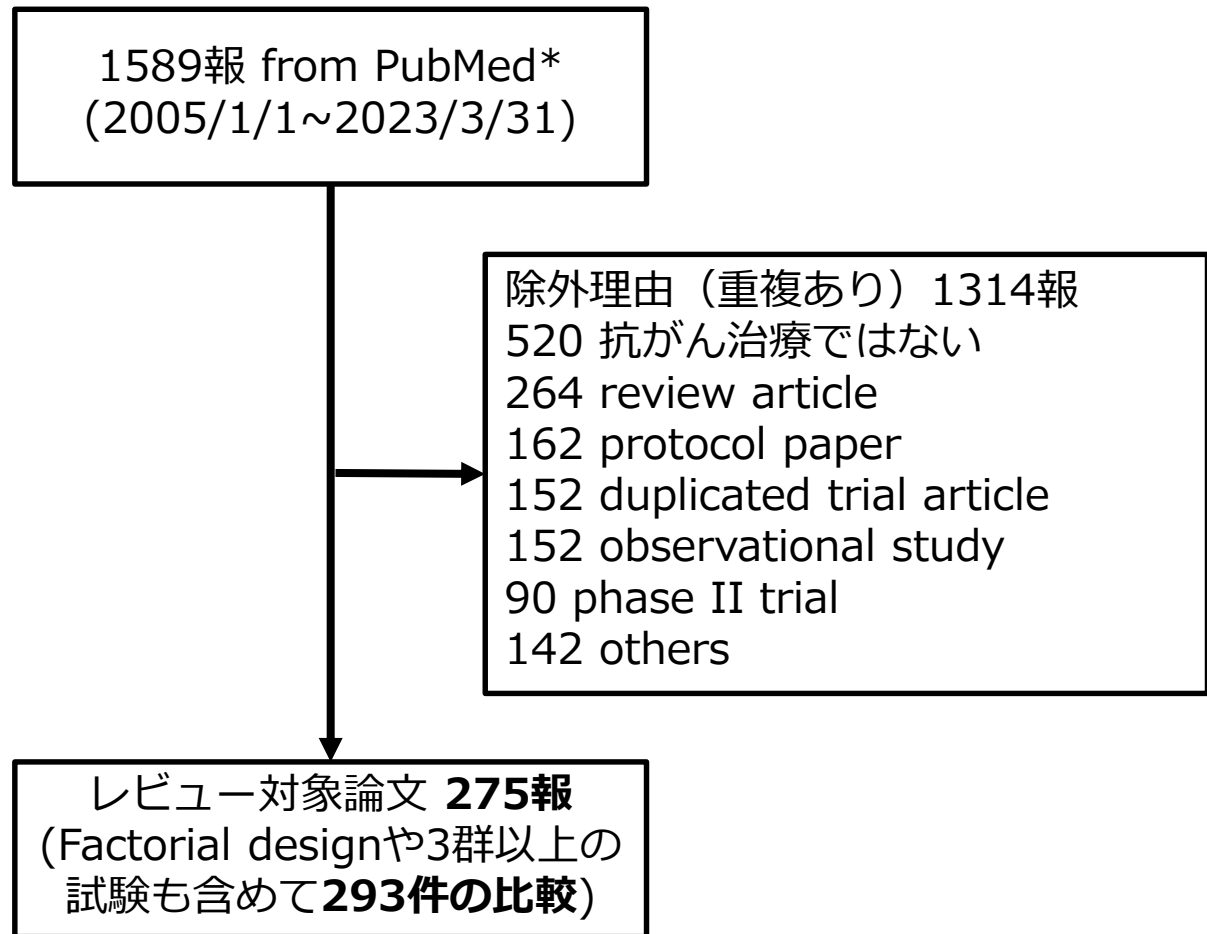
- 非劣性試験に関するレビュー論文の既報
 - 至適な非劣性マージンの検討
 - 主たる解析結果のreporting方法の妥当性
- 低侵襲性の内容や、論文中での記載方法を臨床・統計の両側面からレビューした文献は未だにない



対象：抗がん治療の検証的な非劣性試験

調査内容：試験デザインの概要、想定される低侵襲の内容、評価指標、統計学的・臨床的な意思決定

Flow chart

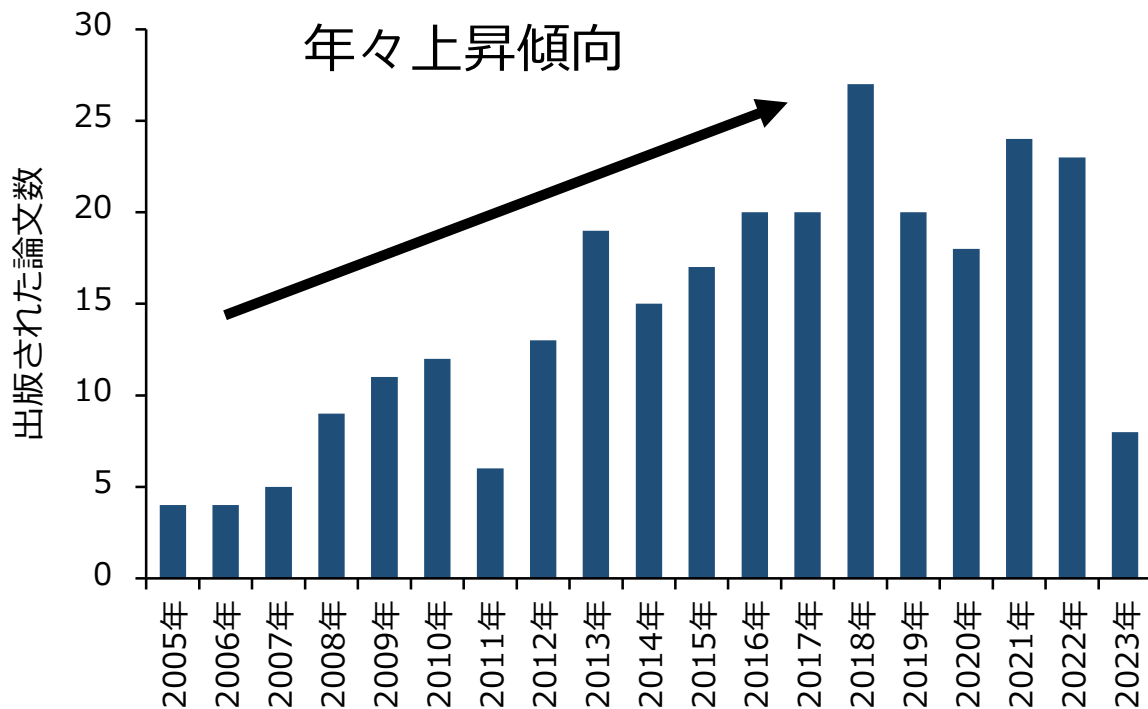


*PubMedによる候補試験の調査 (検索日: 2023年3月31日)

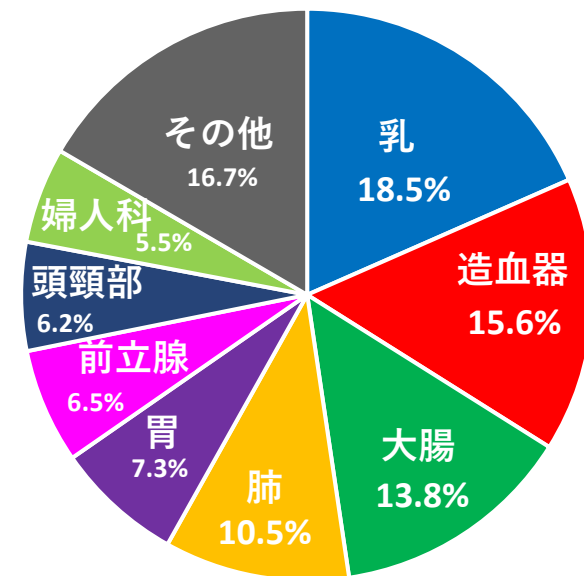
検索条件: "noninferiority*" [Text Word] AND ("cancer*" [Text Word] OR "tumor*" [Text Word] OR "tumour*" [Text Word] OR "lymphoma*" [Text Word] OR "myeloma*" [Text Word] OR "leukemia*" [Text Word] OR "carcinoma*" [Text Word] OR "sarcoma*" [Text Word] OR "oncology*" [Text Word] OR "malignancy*" [Text Word]) AND ("phase iii" [Text Word] OR "phase 3" [Text Word] OR "randomized*" [Text Word] OR "randomised*" [Text Word] OR "confirm*" [Text Word])

年別の出版された論文数とがん種

出版論文数(n=275)



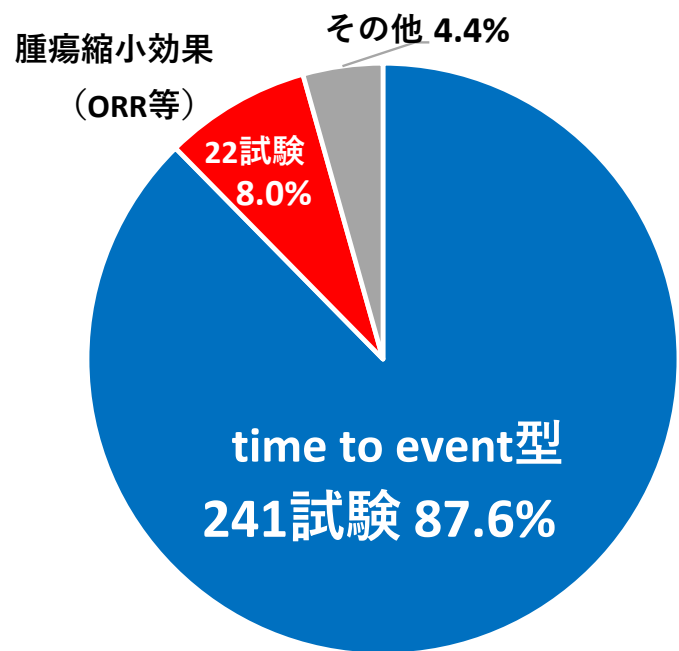
がん種(n=275)



- 乳がん、造血器腫瘍、大腸がん
で概ね半分を占める
- 乳がんに対する放射線治療（寡
分割照射，照射範囲の縮小）の
試験が多い

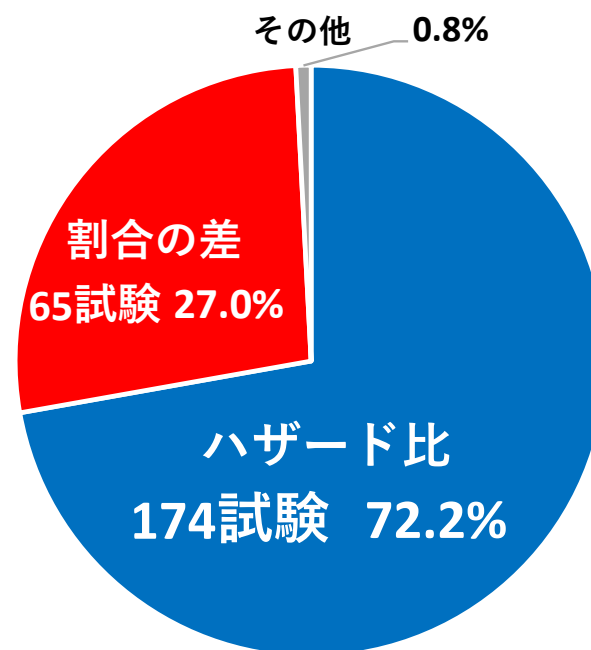
Primary endpoint

Primary endpoint(n=275)



- 概ねtime to event型
- 腫瘍縮小効果のうち、造血器腫瘍10試験、乳がん4試験 (いずれも2017年以前に出版)
- time to event型のうち、OSは77試験 (全体の28.0%)

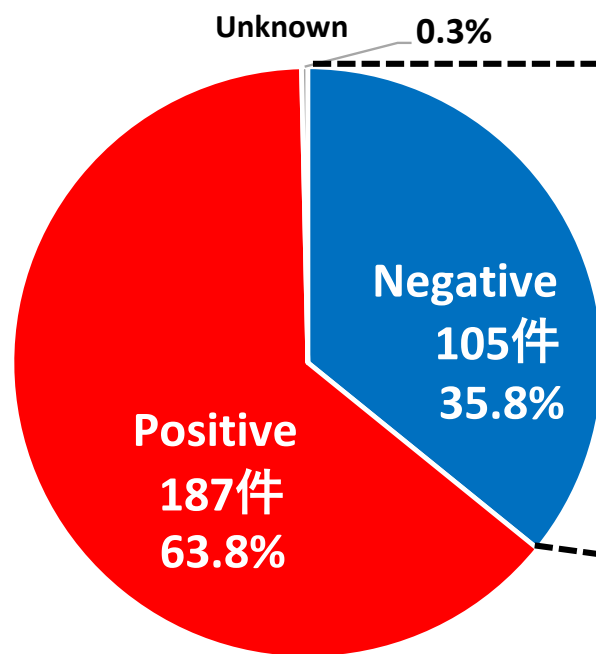
time to event型で用いた指標(n=241)



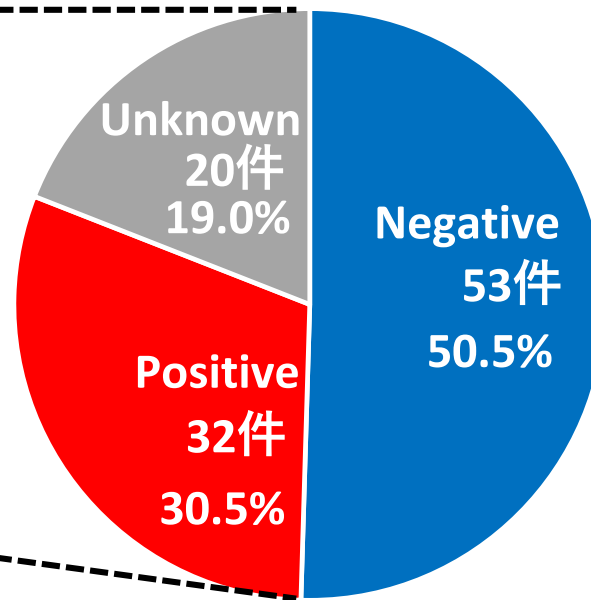
- 概ねハザード比であるが、割合の差も一定数用いられている

統計学的・臨床的な結論

統計学的結論(n=293)



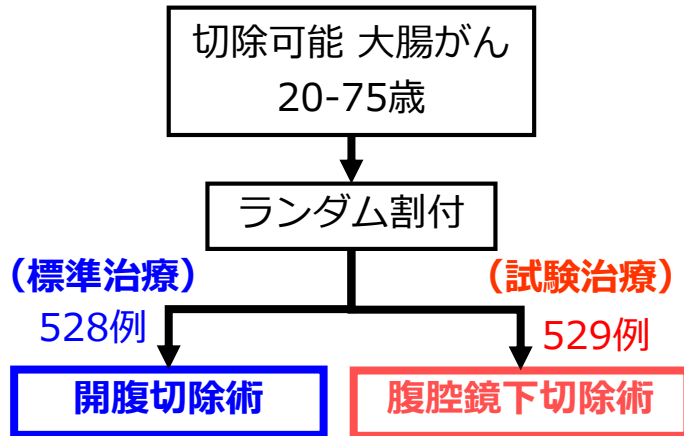
統計学的結論がNegativeのうち臨床的結論(n=105)



- 約3分の2がPositiveな結果
- 結論が明記されていない比較が1件

- 統計学的にNegativeであった結果に対して約30%でPositiveと結論
- Positiveと結論した理由として、ほぼ半分（15件）が検出力不足

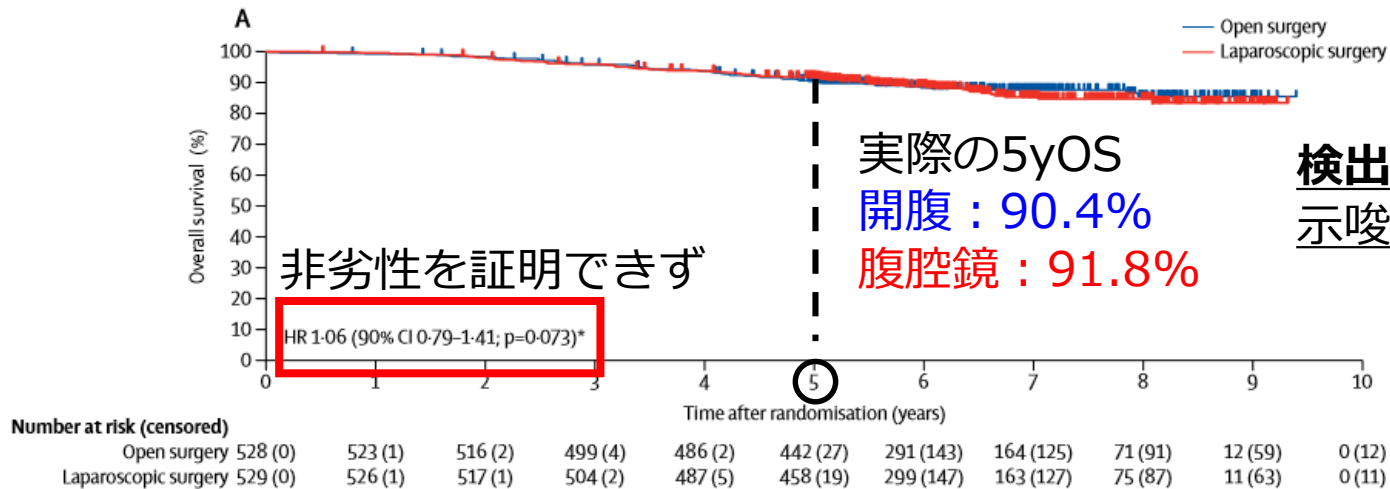
JCOG0404



【サンプルサイズ設定】

- 期待5年OS 75%, サンプルサイズ 818, 非劣性マージン 1.366
↓ 5年OSを上方修正
- 期待5年OS 82%, サンプルサイズ 1050, 非劣性マージン 1.366

観察イベント/必要イベント = **128/254**



統計学的に非劣性が検証されなかったが…

CQ 4 : 大腸癌に対して腹腔鏡下手術は推奨されるか？

腹腔鏡下手術は大腸癌手術の選択肢の1つとして行うことを弱く推奨する。**(推奨度2・エビデンスレベルB)** ただし、横行結腸癌および直腸癌に対する腹腔鏡下手術の有効性は十分に確立されていないことを患者に説明したうえで実施する。

局所進行癌、肥満や癒着症例は難度が高いため、個々の手術チームの習熟度を十分に考慮して適応を決定する。

海外の大規模ランダム化比較試験やコクランレビューにおいて、結腸癌およびRS癌に対する腹腔鏡下手術の有用性が開腹手術との比較で検討され、腹腔鏡下手術では手術時間が長い一方、出血量が少ない、腸管運動の回復が早い、在院期間が短いなどの短期成績が優れていること、合併症発生率および再発率・生存率は同等であることが報告されている。本邦で行われた大規模ランダム化比較試験であるJCOG0404試験では主要評価項目の全生存期間において、開腹手術群に対する腹腔鏡下手術群のハザード比は1.056 (90%信頼区間 0.790–1.413) , $p = 0.0732$ であり、腹腔鏡下手術の非劣性は示されなかったものの、5年生存率は開腹手術群で90.4% (95%信頼区間 87.5–92.6) , 腹腔鏡下手術群で91.8% (同 : 89.1–93.8) でどちらも良好であり、腹腔鏡手術も治療の選択肢として位置づけられる。ただし、本研究では、横行結腸癌、下行結腸癌や直腸癌は対象患者から除外されており、大腸癌全般の適応に関するエビデンスとしては一定のバイアスがかかっている可能性がある。また、同試験のサブグループ解析ではRS, cN2, 肥満例やT4にて腹腔鏡下手術の予後が悪い傾向があり、これらを考慮して慎重に適応を決定する必要がある。

統計学的結論＝臨床的結論とは限らない。

結果の解釈について臨床家と統計家のコラボが重要

まとめ

- 非劣性試験のデザイン
 - 非劣性マージンの決め方
 - リスク&ベネフィットのバランスに加えて、バイオクリープ（Bio creep）もケアする必要がある
 - ハザード比と生存割合
 - 同じ生存割合の差でも、予後が異なればハザード比と臨床的解釈は異なる
 - サンプルサイズ計算時は両指標のバランスを考慮する必要がある
 - 解析対象集団
 - 非劣性試験ではPPSが望ましいとされるが、PPSの定義および分析感度について考慮する必要がある
- Hybrid design
 - 優越性も見込めることを前提に、必要サンプルサイズを減らすことが可能なデザイン
 - 考え方も含めて非劣性試験そのもの