試験デザインに必要な
生物統計の基礎知識

研究支援センター 生物統計部医学統計室

若林 将史

令和6年度 医師研修

2024. 11.30 (土)

Ramucirumab with cisplatin and fluoropyrimidine as first-line therapy in patients with metastatic gastric or junctional adenocarcinoma (RAINFALL): a double-blind, randomised, placebo-controlled, phase 3 trial

Charles S Fuchs, Kohei Shitara, Maria Di Bartolomeo, Sara Lonardi, Salah-Eddin Al-Batran, Eric Van Cutsem, David H Ilson, Maria Alsina, Ian Chau, Jill Lacy, Michel Ducreux, Guillermo Ariel Mendez, Alejandro Molina Alavez, Daisuke Takahari, Wasat Mansoor, Peter C Enzinger, Vera Gorbunova, Zev A Wainberg, Susanna Hegewisch-Becker, David Ferry, Ji Lin, Roberto Carlesi, Mayukh Das, Manish A Shah, RAINFALL Study Group*

Lancet Oncol 2019; 20: 420–35
Published Online February 1, 2019
http://dx.doi.org/10.1016/S1470-2045(18)30791-5
This online publication has been corrected. The corrected version first appeared at thelancet.com/oncology on April 30, 2019
See Comment page 373

*RAINFALL Study Group principal investigators listed in the appendix
Yale Cancer Center, Smilow Cancer Hospital, New Haven, CT, USA (C S Fuchs MD, Prof J Lacy MD); National Cancer Center Hospital East, Kashiwa, Japan (K Shitara MD); Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy (M Di Bartolomeo MD); Istituto Oncologico Veneto-IRCCS, Padova, Italy (S Lonardi MD); Institute of Clinical Cancer Research at Krankenhaus Nordwest, UCT-University Cancer Center, Frankfurt, Germany (Prof S-E Al-Batran MD); University Hospitals Gasthuisberg, Leuven and KULeuven, Belgium (Prof E Van Cutsem MD); Memorial Sloan Kettering Cancer Center, New York, NY, USA (D H Ilson MD); Vall d'Hebron University Hospital and Institute of Oncology, Barcelona, Spain (M Alsina MD); Royal Marsden Hospital, Sutton, Surrey, United Kingdom (I Chau MD); Gustave Roussy Cancer Centre, Grand Paris, Villejuif, France (Prof M Ducreux MD); Université Paris-Saclay, France (Prof M Ducreux); Hospital Universitario Fundacion Favaloro, Buenos Aires, Argentina (G A Mendez MD);

Summary

Background VEGF and VEGF receptor 2 (VEGFR-2)-mediated signalling and angiogenesis can contribute to the pathogenesis and progression of gastric cancer. We aimed to assess whether the addition of ramucirumab, a VEGFR-2 antagonist monoclonal antibody, to first-line chemotherapy improves outcomes in patients with metastatic gastric or gastro-oesophageal junction adenocarcinoma.

Methods For this double-blind, randomised, placebo-controlled, phase 3 trial done at 126 centres in 20 countries, we recruited patients aged 18 years or older with metastatic, HER2-negative gastric or gastro-oesophageal junction adenocarcinoma, an Eastern Cooperative Oncology Group (ECOG) performance status of 0 or 1, and adequate organ function. Eligible patients were randomly assigned (1:1) with an interactive web response system to receive cisplatin (80 mg/m², on the first day) plus capecitabine (1000 mg/m², twice daily for 14 days), every 21 days, and either ramucirumab (8 mg/kg) or placebo on days 1 and 8, every 21 days. 5-Fluorouracil (800 mg/m² intravenous infusion on days 1–5) was permitted in patients unable to take capecitabine. The primary endpoint was investigator-assessed progression-free survival, analysed by intention to treat in the first 508 patients. We did a sensitivity analysis of the primary endpoint, including a central review of CT scans. Overall survival was a key secondary endpoint. This study is registered with ClinicalTrials.gov, number NCT02314117.

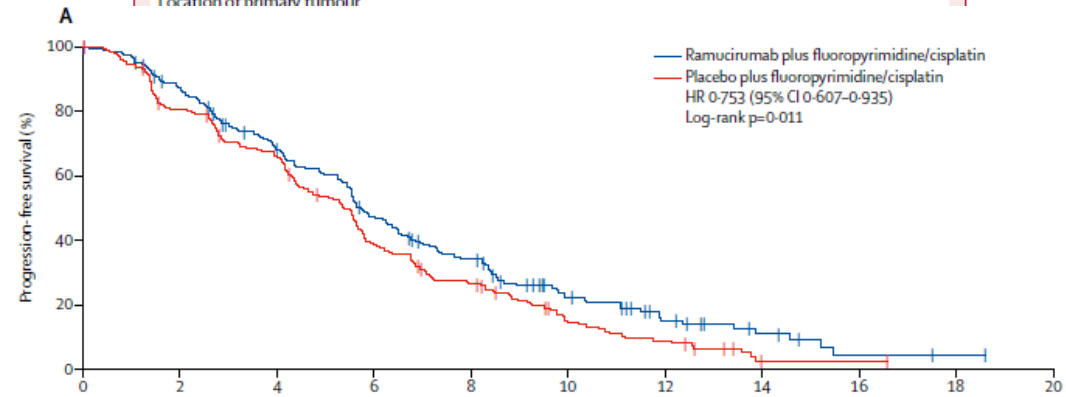
Findings Between Jan 28, 2015, and Sept 16, 2016, 645 patients were randomly assigned to receive ramu fluoropyrimidine and cisplatin (n=326) or placebo plus fluoropyrimidine and cisplatin (n=319). Investig progression-free survival was significantly longer in the ramucirumab group than the placebo group [HR] 0·753, 95% CI 0·607–0·935, p=0·0106; median progression-free survival 5·7 months [5·5–6·5] [4·5–5·7]). A sensitivity analysis based on central independent review of the radiological images did n the investigator-assessed difference in progression-free survival (HR 0·961, 95% CI 0·768–1·203, p=0·7 no difference in overall survival between groups (0·962, 0·801–1·156, p=0·6757; median overall survival [9·9–11·9] in the ramucirumab group vs 10·7 months [9·5–11·9] in the placebo group). The most comm adverse events were neutropenia (85 [26%] of 323 patients in the ramucirumab group vs 85 [27%] of 315 group), anaemia (39 [12%] vs 44 [14%]), and hypertension (32 [10%] vs 5 [2%]). The incidence of any- adverse events was 160 (50%) of 323 patients in the ramucirumab group and 149 (47%) of 315 patients group. The most common serious adverse events were vomiting (14 [4%] in the ramucirumab group vs placebo group) and diarrhoea (11 [3%] vs 19 [6%]). There were seven deaths in each group, either treatment or within 30 days of discontinuing study treatment, which were the result of treatment-re events. In the ramucirumab group, these adverse events were acute kidney injury, cardiac arrest, gastric peritonitis, pneumothorax, septic shock, and sudden death (n=1 of each). In the placebo group, these a were cerebrovascular accident (n=1), multiple organ dysfunction syndrome (n=2), pulmonary embolism (n=1), and small intestine perforation (n=1).

Interpretation Although the primary analysis for progression-free survival was statistically significant, was not confirmed in a sensitivity analysis of progression-free survival by central independent review improve overall survival. Therefore, the addition of ramucirumab to cisplatin plus fluoropyrimidine ch not recommended as first-line treatment for this patient population.

Funding Eli Lilly and Company.

Copyright © 2019 Elsevier Ltd. All rights reserved.

	Ramucirumab plus fluoropyrimidine and cisplatin group (n=326)	Placebo plus fluoropyrimidine and cisplatin group (n=319)
Median age, years	60 (51–68)	62 (54–68)
Sex		
Male	214 (66%)	215 (67%)
Female	112 (34%)	104 (33%)
Race*		
White	256 (79%)	264 (83%)
Asian	38 (12%)	31 (10%)
Black or African American	2 (1%)	3 (1%)
Other	13 (4%)	11 (3%)
Data missing	17 (5%)	10 (3%)
ECOG performance status		
0	141 (43%)	143 (45%)
1	185 (57%)	176 (55%)
Geographic region		
North America	52 (16%)	37 (12%)
Europe	194 (60%)	205 (64%)
Japan	32 (10%)	28 (9%)
Other	48 (15%)	49 (15%)
Location of primary tumour		



Number at risk (number censored)

Ramucirumab group	255	194	144	89	60	29	15	7	2	1	0
	(0)	(32)	(40)	(54)	(59)	(72)	(78)	(83)	(85)	(86)	(87)
Placebo group	253	179	141	78	51	21	12	1	1	0	0
	(0)	(30)	(36)	(42)	(45)	(55)	(56)	(61)	(61)	(62)	(62)

Charles S Fuchs *et al.* (2019) Lancet Oncol;20(3):420-435.

目次

- ランダム化とは？
  - 交絡とランダム化
- 臨床試験の結果の解釈
  - 仮説検定、検定p値
- サンプルサイズ設定の必要性
  - 多ければ多いほど良いのか？

目次

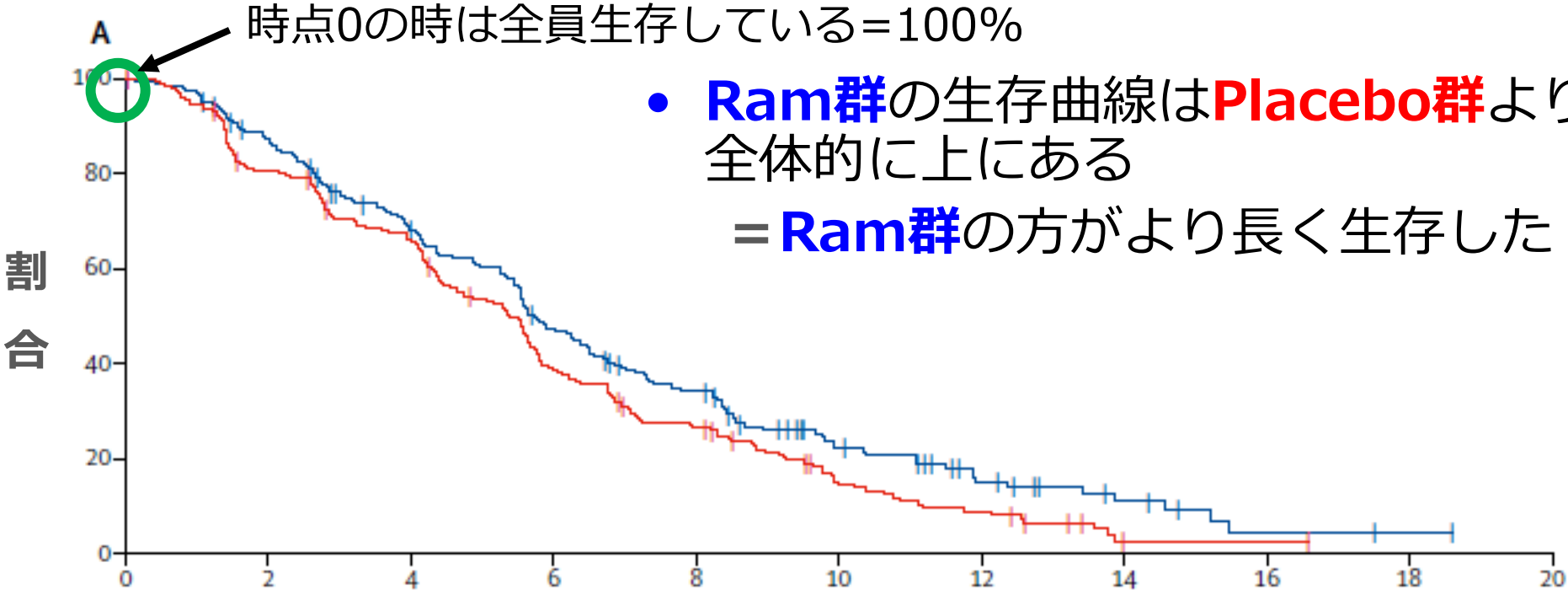
- ランダム化とは？
  - 交絡とランダム化

- 臨床試験の結果の解釈
  - 仮説検定、検定p値

- サンプルサイズ設定の必要性
  - 多ければ多いほど良いのか？

生存曲線

- 縦軸に割合、横軸に時間をとり、集団における各時点の生存割合をつないだもの
- イベントが発生するとその時点で割合が減少する

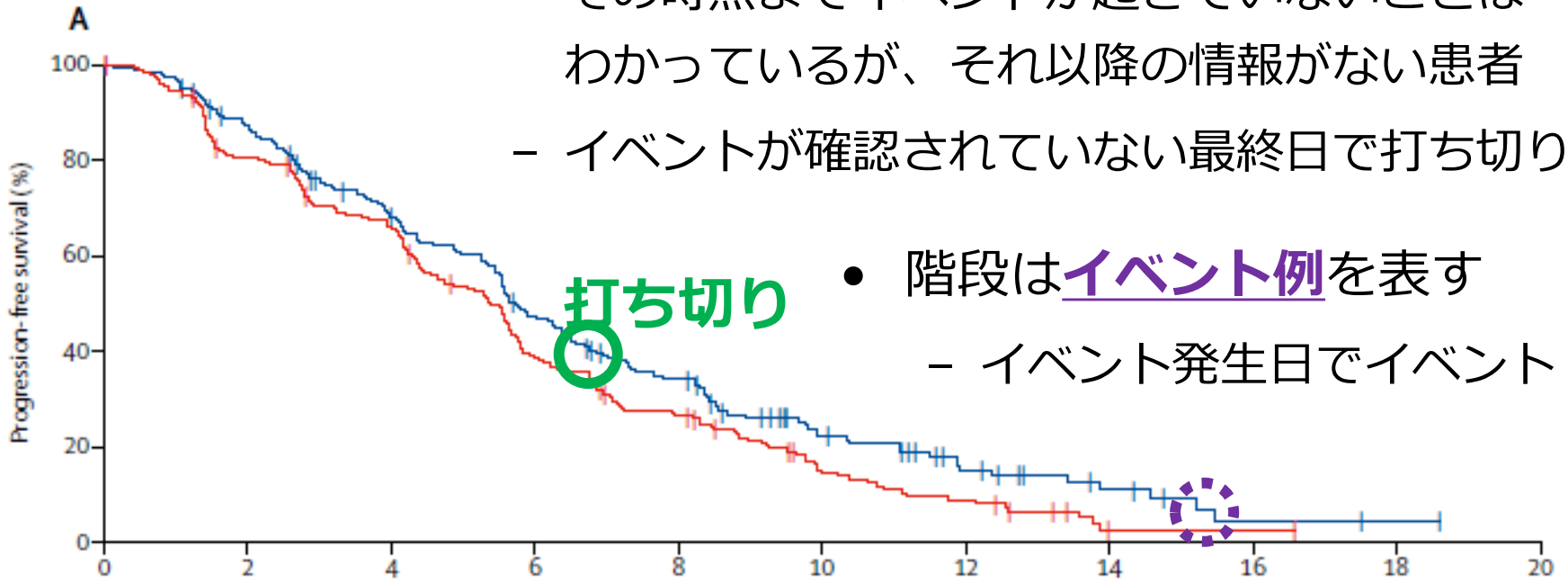


- **Ram群**の生存曲線は**Placebo群**よりも全体的に上にある
  ＝**Ram群**の方がより長く生存した

Charles S Fuchs *et al.* (2019) Lancet Oncol;20(3):420-435.

イベントと打ち切り

- ヒゲの印は**打ち切り例**を表す
  - その時点までイベントが起きていないことはわかっているが、それ以降の情報がない患者
  - イベントが確認されていない最終日で打ち切り



- 階段は**イベント例**を表す
  - イベント発生日でイベント

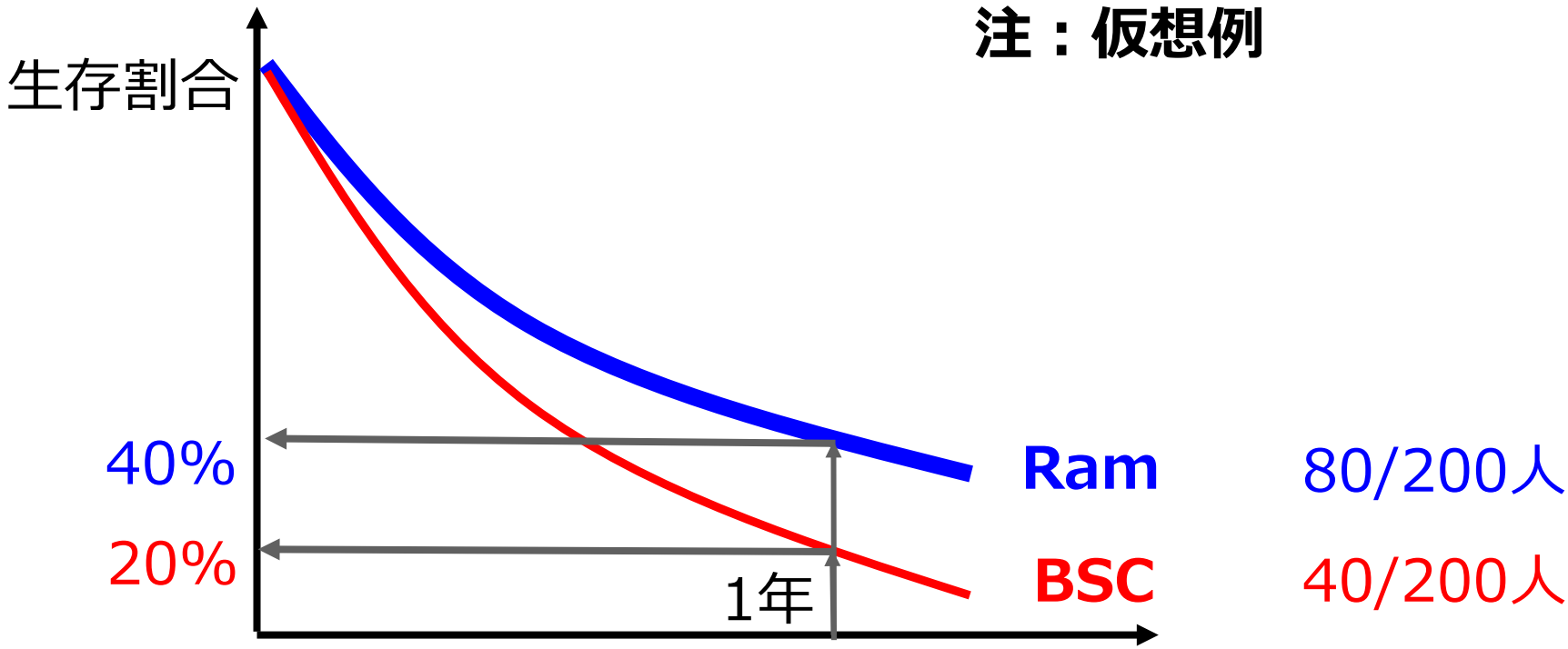
Number at risk (number censored)											
Ramucirumab group	255 (0)	194 (32)	144 (40)	89 (54)	60 (59)	29 (72)	15 (78)	7 (83)	2 (85)	1 (86)	0 (87)
Placebo group	253 (0)	179 (30)	141 (36)	78 (42)	51 (45)	21 (55)	12 (56)	1 (61)	1 (61)	0 (62)	0 (62)

リスク集合：ある時点までイベント/打ち切りが起きていない患者数

Charles S Fuchs *et al.* (2019) Lancet Oncol;20(3):420-435.

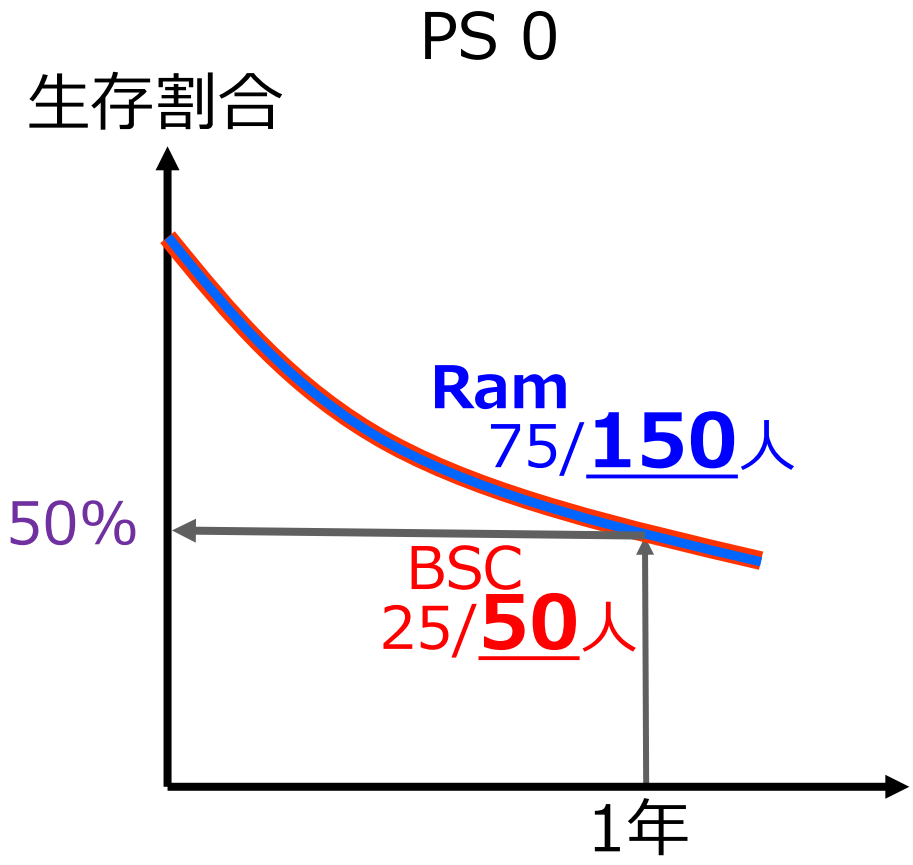
国内の学会で見かける発表

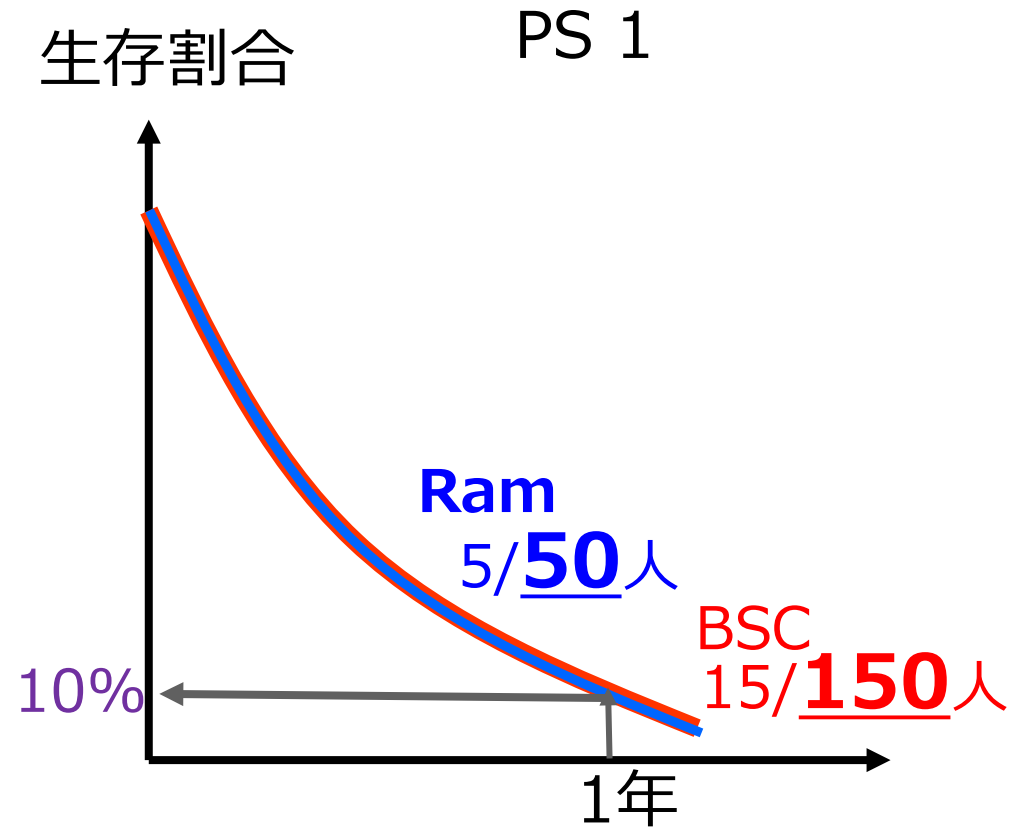
- ある病院の治療成績としてRamucirumab(200例)とBSC (Placebo)(200例) の予後について後方視的に解析を行った。
- Ram群はBSC群と比較し予後良好であったため、この対象にRamucirumabをすることが推奨される。



PSで分けた場合の予後

*PS(Performance Status) :全身状態を表し、
数値が大きいほど状態が悪いことを示す





PSによらずRamとBSCの予後は変わらない

比較したいのはRamとBSCの違いだから

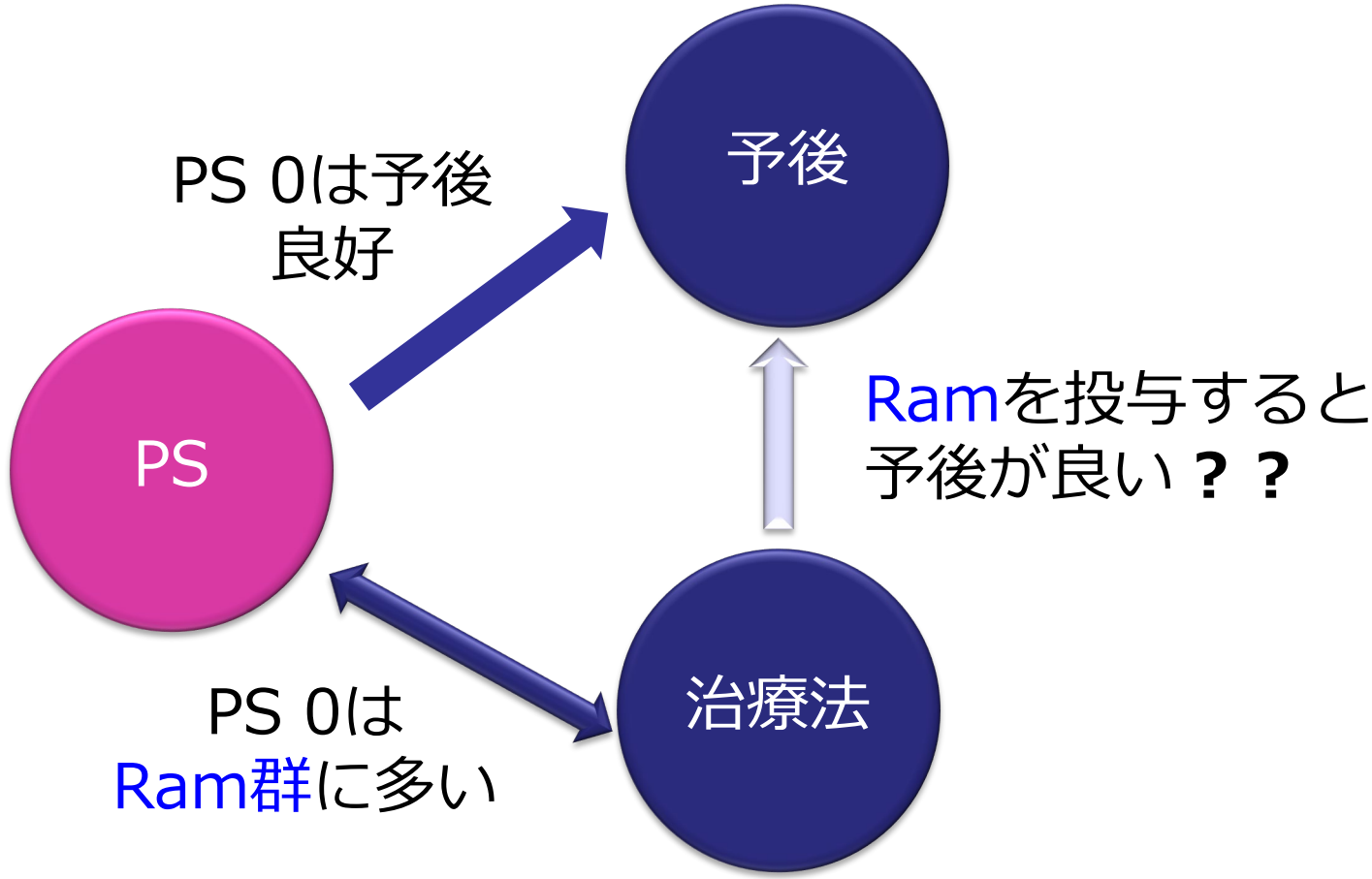
- 治療法以外の予後に影響する要因の条件が同じでなければ、”比較”にならない！！

治療法	PS 0		PS 1	合計
Ram	**150人** (75%)	＞＞	50人	200人
BSC	50人 (25%)	＜＜	**150人**	200人

- RamはBSCと比べ「PS 0」の割合が高い
- PSによって予後が異なる(PS 0は予後良)

交絡

- 治療法と予後に関連する第3の因子（PS）によって見かけ上の関連が生じてしまう現象のこと（**バイアス**の一種）
  - 交絡を引き起こす因子（=PS）のことを**交絡因子**という



交絡がないことを保証するには

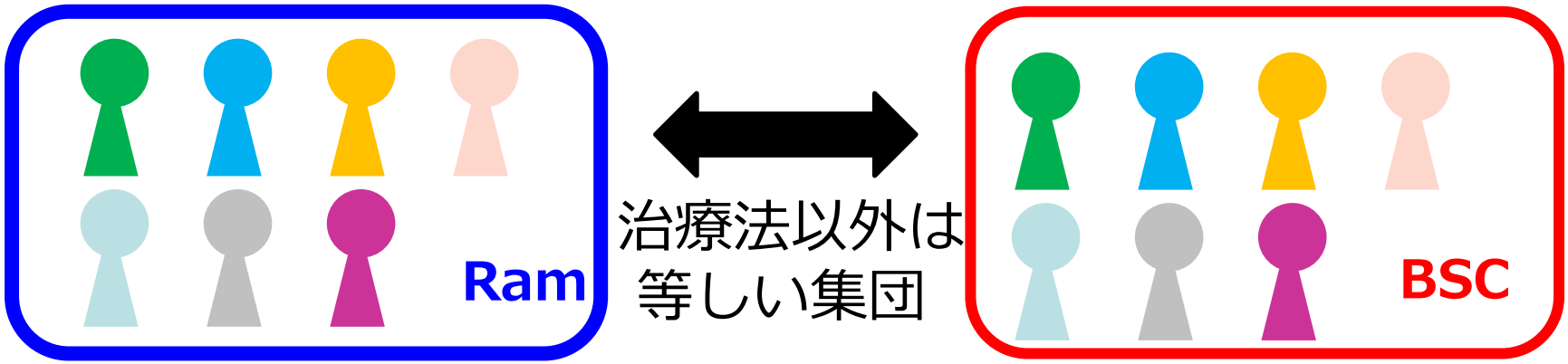
- 治療群間で予後に関係する背景因子を揃える
  - Performance Status（全身状態）
  - 原発部位（胃 / 胃食道接合部）
  - 原発巣有無
  - その他…

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

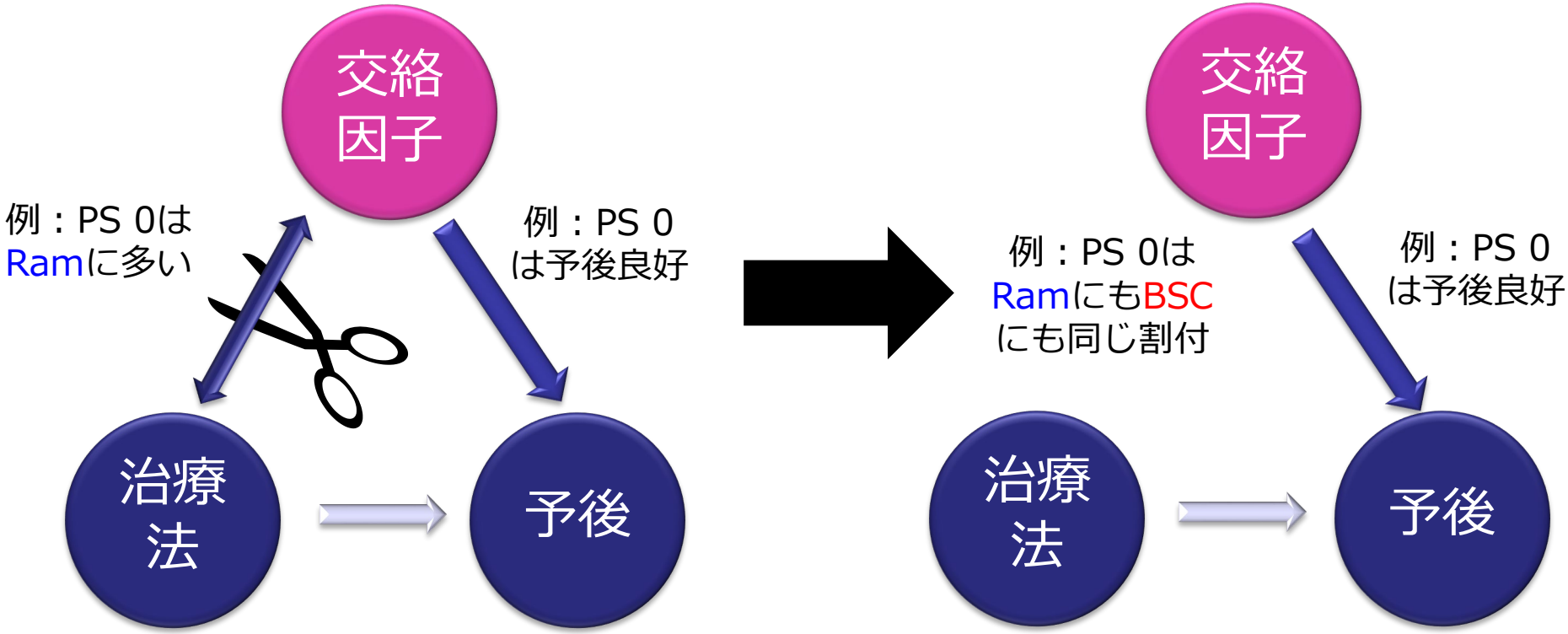
ランダム化 randomization

- 担当医や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
  - 例：候補患者に対して、背景因子がよりバランスされる群に高い確率で割付られる（最小化法）
- 予見による患者選択の偏りの防止
  - 状態の良い患者をRam群に割り付けやすくする、などを防ぐ



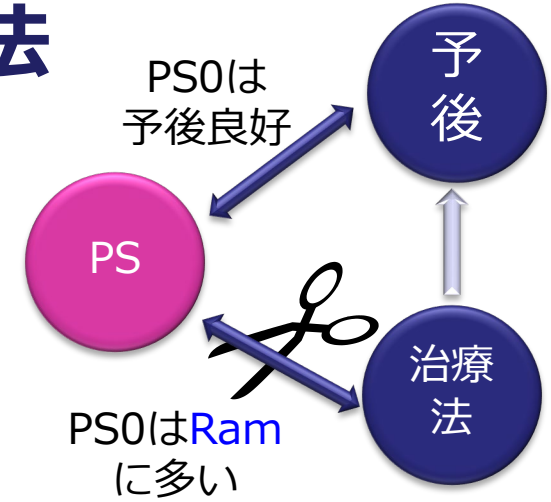
ランダム化の意義

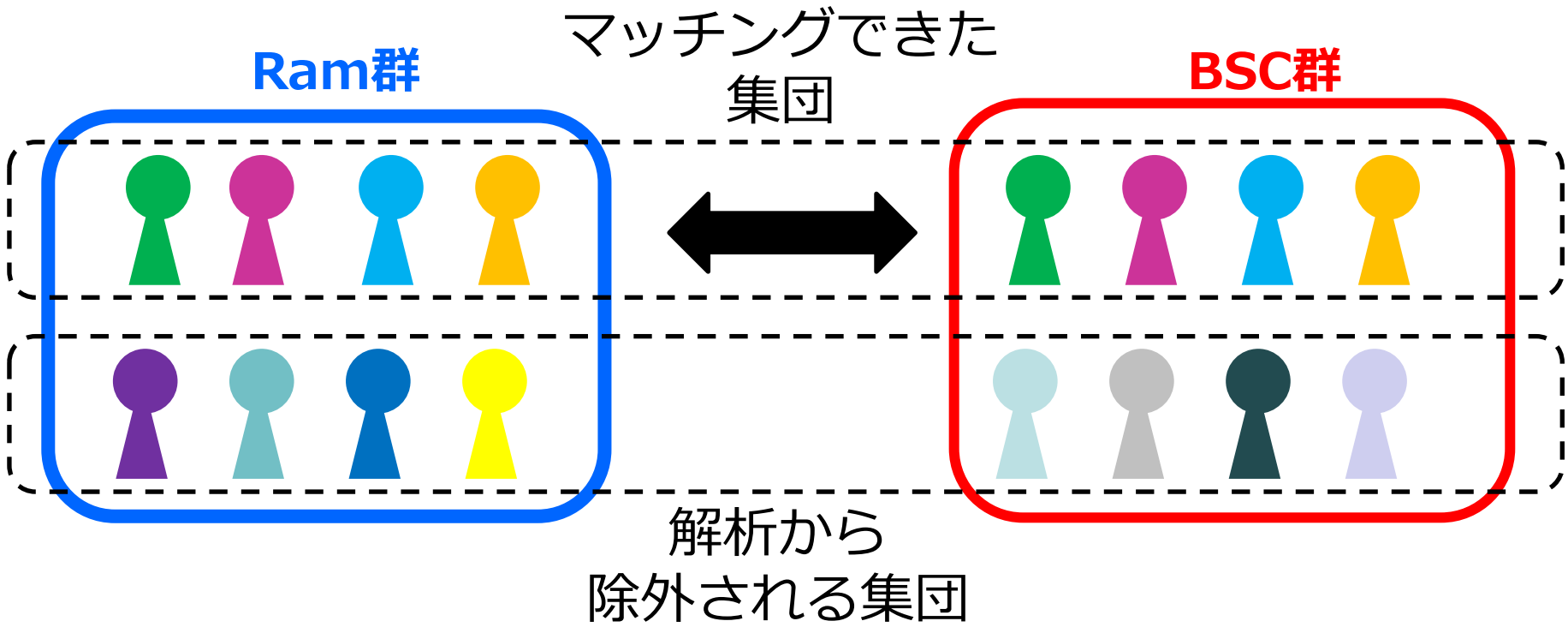
- **治療法**と(未知の因子も含めた)**交絡因子**の関連を除去できる
  - 治療群間で背景因子が等しくなるので、交絡が除去される
    - 治療法以外は等しい集団 → 効果に差があれば治療法の違い



【参考】デザイン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





【参考】解析段階で比較可能性を高めるための方法

- **サブグループ解析(subgroup analysis)**
  - サブセット解析、部分集団解析
  - 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)**
  - Cox回帰やロジスティック回帰を行う
- **傾向スコアを用いた解析(propensity score analysis)**
  - 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

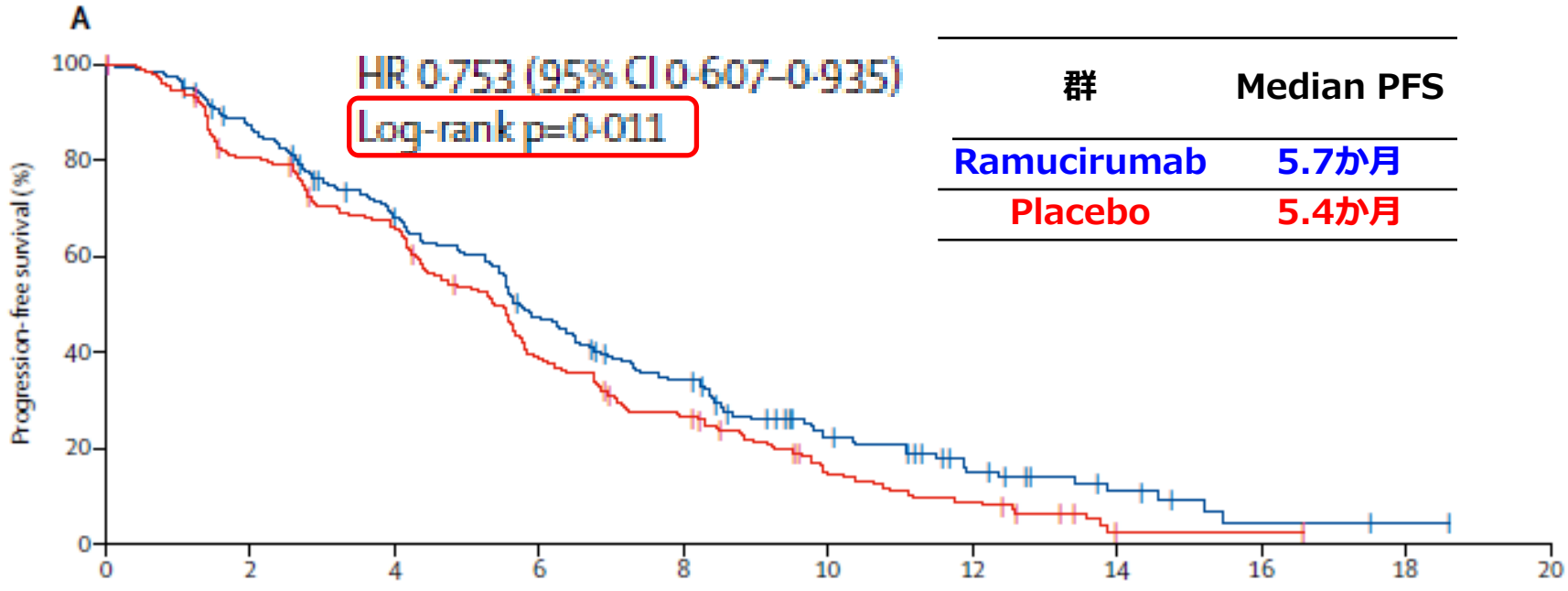
目次

- ランダム化とは？
  - 交絡とランダム化

- 臨床試験の結果の解釈
  - 仮説検定、検定p値

- サンプルサイズ設定の必要性
  - 多ければ多いほど良いのか？

検定p値とは？



群	Median PFS
Ramucirumab	5.7か月
Placebo	5.4か月

Number at risk (number censored)											
Ramucirumab group	255	194	144	89	60	29	15	7	2	1	0
	(0)	(32)	(40)	(54)	(59)	(72)	(78)	(83)	(85)	(86)	(87)
Placebo group	253	179	141	78	51	21	12	1	1	0	0
	(0)	(30)	(36)	(42)	(45)	(55)	(56)	(61)	(61)	(62)	(62)

「5%より大きい/小さい」、というのはよく聞くけど…

p値の正体は何か？

Charles S Fuchs *et al.* (2019) Lancet Oncol;20(3):420-435.

生存曲線が開いている時の解釈

- **2つの可能性がある。どちらが正しい？**

  - 本当に**Ram**と**Placebo**に差があるので、実際に差が出た
    - 正しい結論を得ている

  - 本当は**Ram**と**Placebo**に差がないのに、偶然差があるように見えた
    - 誤った結論をしてしまっている

どちらが正しいか、得られた結果から確認が必要

確かめる方法：仮説検定

- 証明したいことは、「RamとPlaceboに差がある」 ですが、

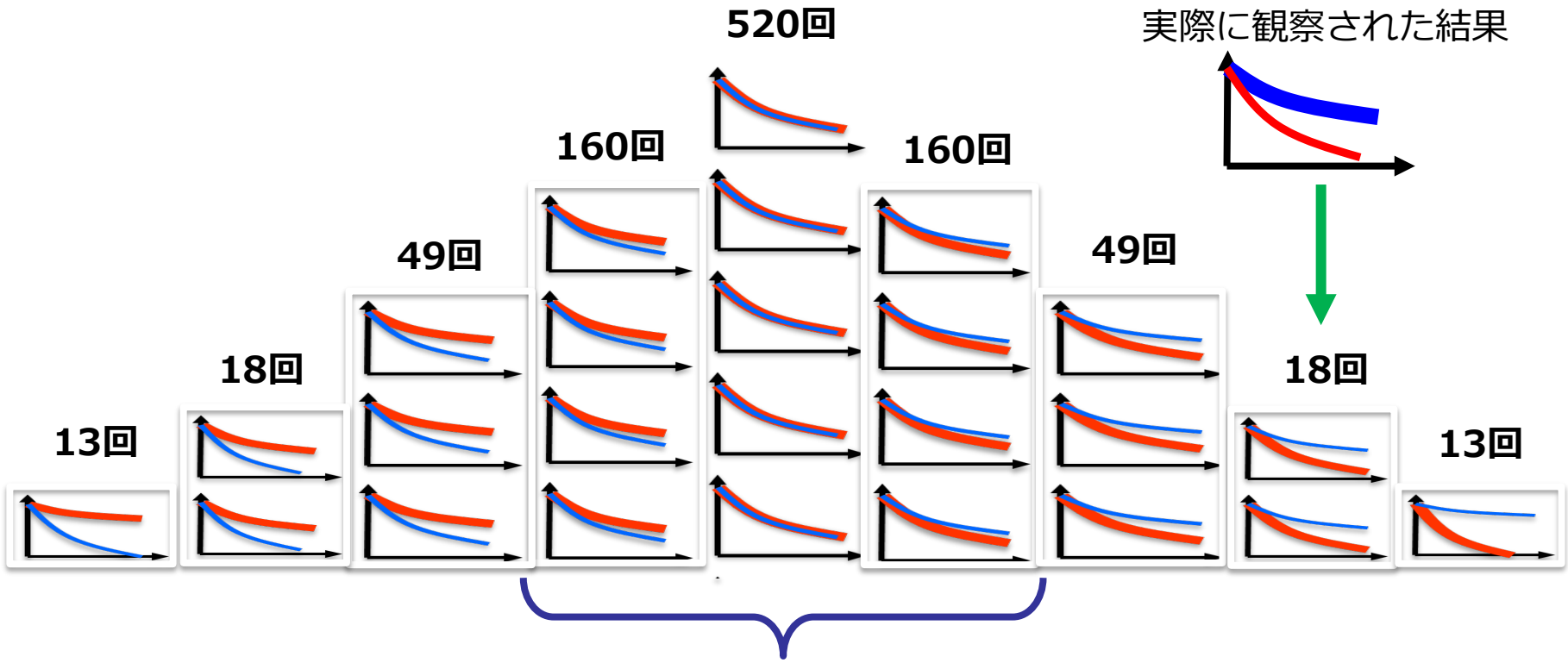
1. 「RamとPlaceboに差がない」という仮説を置く
   - この仮説を**帰無仮説**という
2. 「RamとPlaceboに差がない」という仮定の下で、何回も試験をした場合に得られる結果の分布を調べる
3. 実際に観察されたRamとPlaceboの差以上に大きな差になる確率を調べる
4. この確率が小さければ、そもそも「RamとPlaceboに差がない」という仮説（帰無仮説）が間違っていた、と判断する
5. 「RamとPlaceboに差がある」が正しいと判断する

生存曲線に【差がない】下での結果の分布

もし、RamとPlaceboの生存曲線に【差がない】が真実なら…

日本全国の切除不能胃癌患者500人選んで

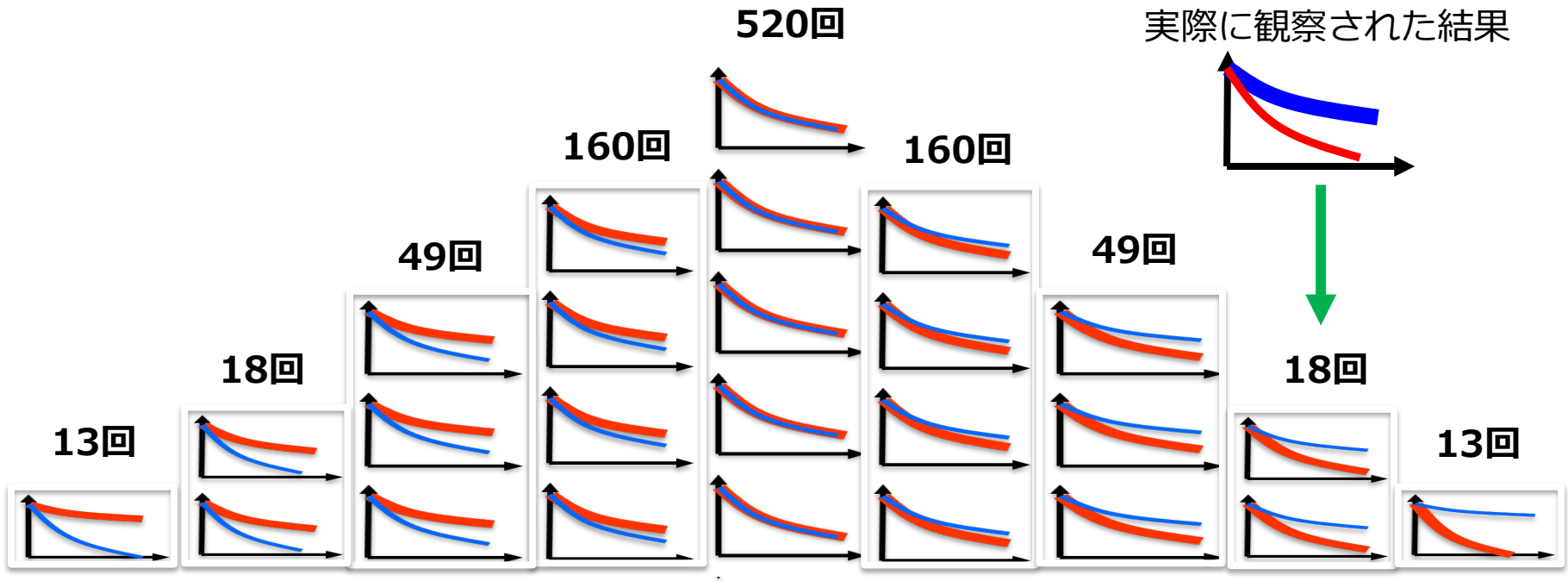
1000回試験すると、、、



【差がない】結果が最も多く観察される

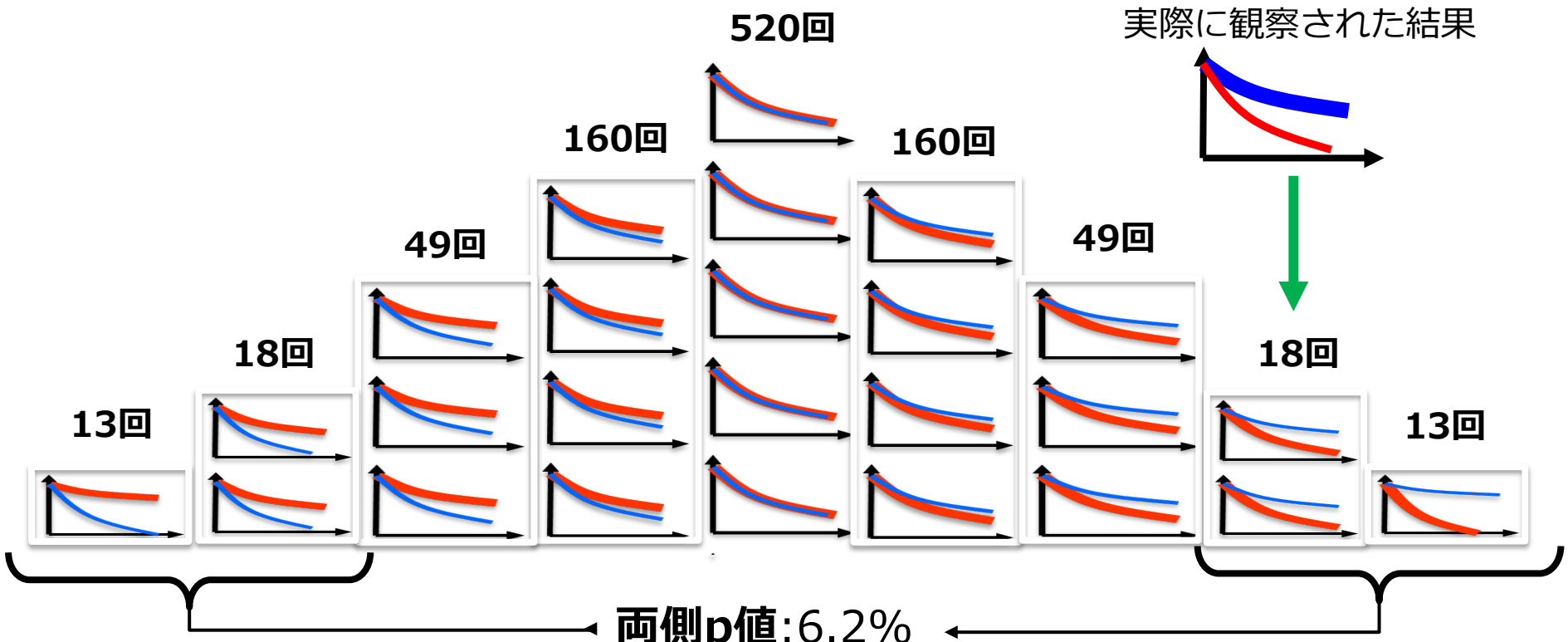
p値の計算

- 実際に観察された結果以上に大きな差になる確率(Probability)は、31/1000 = 3.1%
  - この確率のことを**p値**という
- 実際に観察された結果は【差がない】が真実だとしたら、1000回中31回くらいしか起こらないような**稀な結果**（?）



片側検定と両側検定

- 片側検定：得られた差よりも極端な値をRamが優る方向だけ考慮して求めた**片側p値**に基づいて行う検定
- 両側検定：得られた差よりも極端な値をどちらの治療法が優っているかによらずに求めた**両側p値**に基づいて行う検定

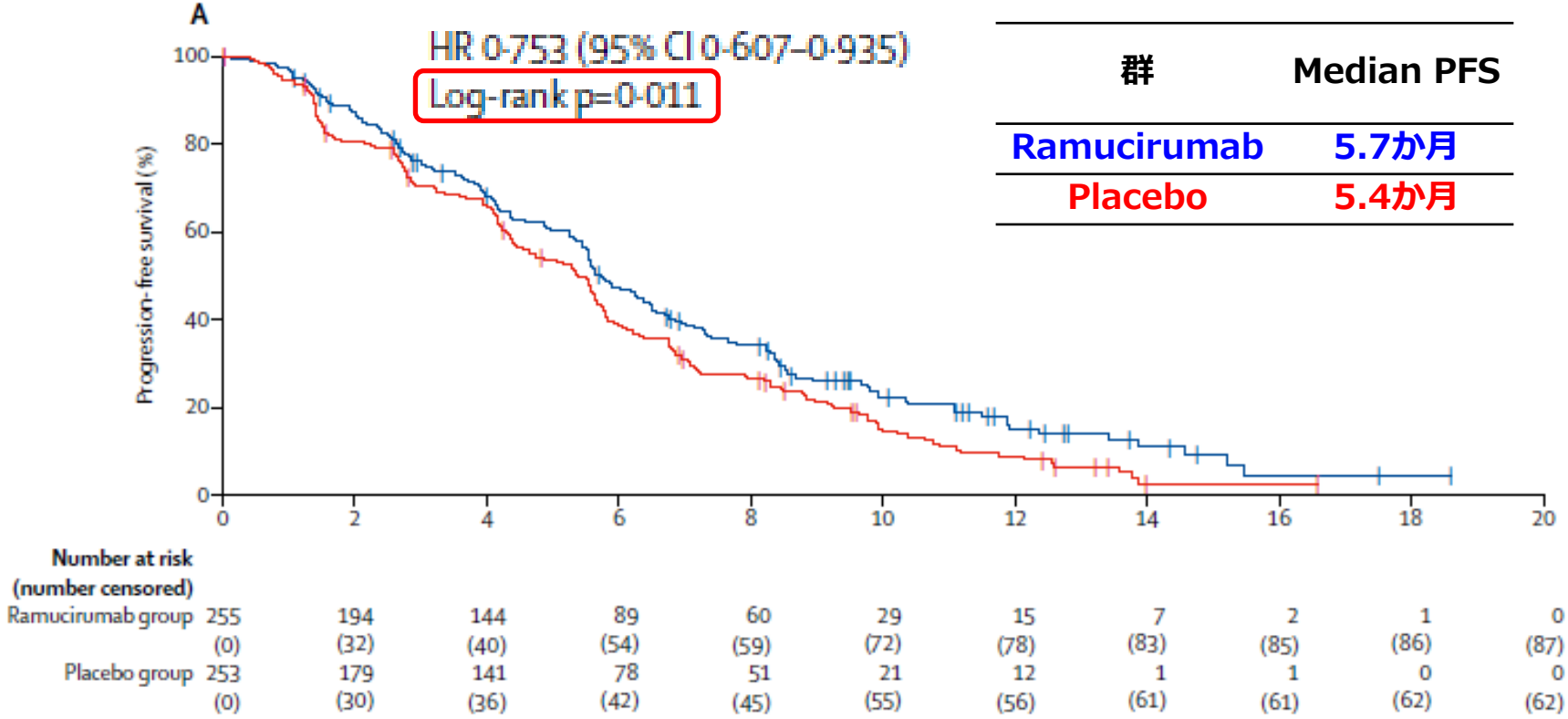


3.1%は稀な結果？

- 3.1%は**稀な結果と考える**場合
  - そもそも【差がない】という仮説が間違っていたと判断し、RamとPlaceboは**差があると結論する ＝ 【有意差あり】**
- 3.1%は**稀な結果とは考えない**場合
  - 【差がない】という仮説は間違っているとは言えないので、RamとPlaceboに**差があるとは言えないと結論する ＝【有意差なし】**
- 結果を見てから稀かどうかを判断すると後付けになってしまうので、事前に稀かどうかの規準を決めておく
  - この規準のことを**有意水準(α level)**という
  - p値が有意水準を下回ったら、【有意差あり】と結論する

RAINFALL試験の場合

- **<u>p=0.011</u>**：両群に差がないとしたら100回中1回くらいしか起こらない稀な事象
  - 事前に決めた規準(両側有意水準) α≦5%も満たす
  - RamucirumabはPlaceboと比較して優れていると判断



群	Median PFS
Ramucirumab	5.7か月
Placebo	5.4か月

Charles S Fuchs *et al.* (2019) Lancet Oncol;20(3):420-435.

目次

- ランダム化とは？
  - 交絡とランダム化

- 臨床試験の結果の解釈
  - 仮説検定、検定p値

- サンプルサイズ設定の必要性
  - 多ければ多いほど良いのか？

【質問】 臨床試験におけるサンプルサイズは、

1.多ければ多い方が良い？

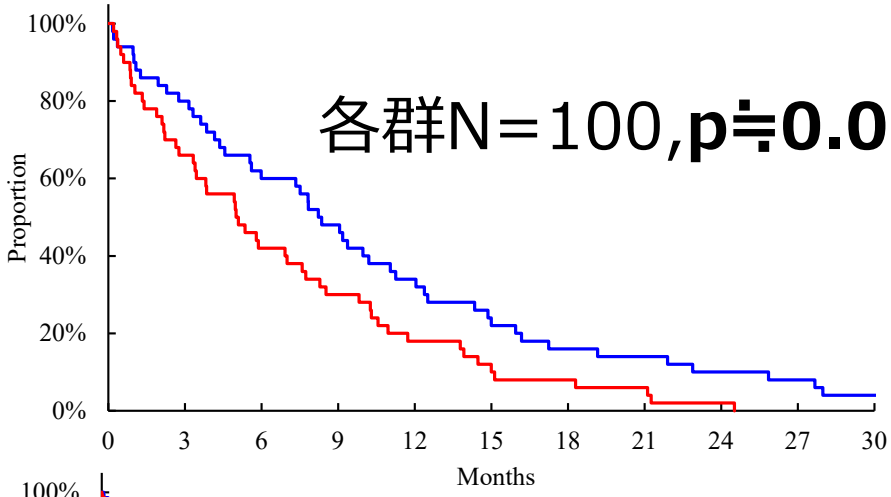
2.少なければ少ない方が良い？

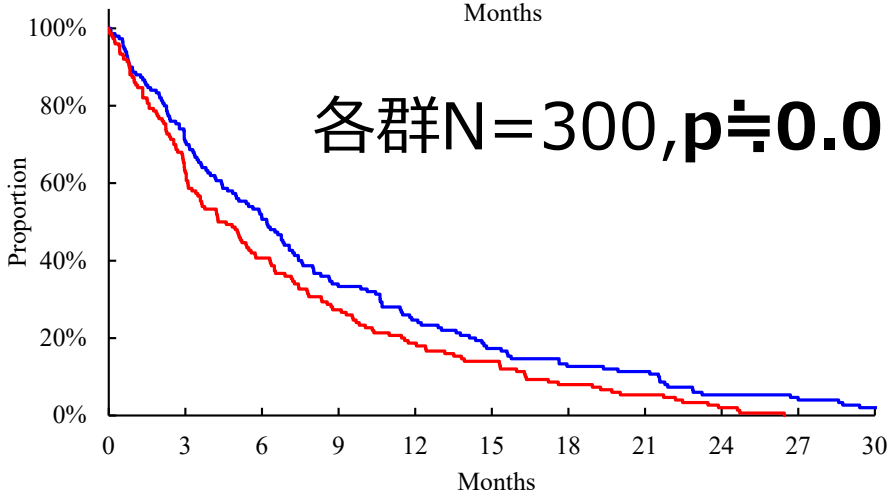
p値＝治療効果の大きさを表す指標？

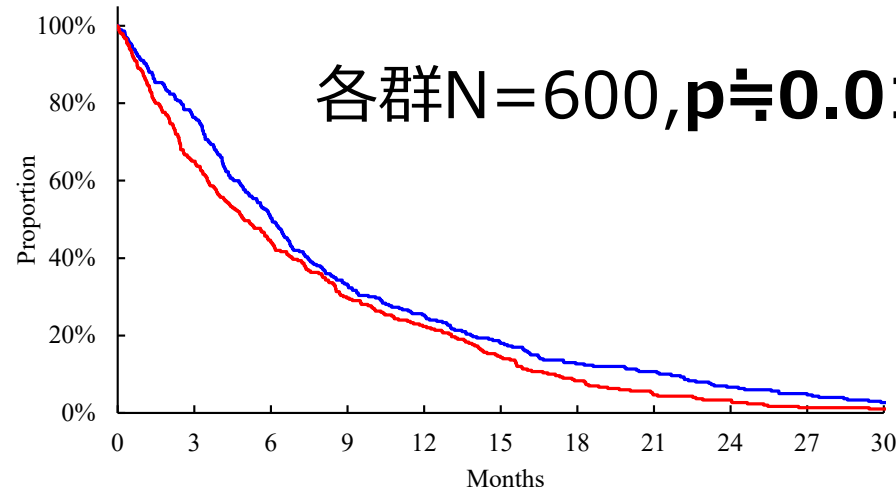
同じp≒0.01でも臨床的意味は異なる

p値は治療効果の大きさを表す指標ではない

注：仮想例

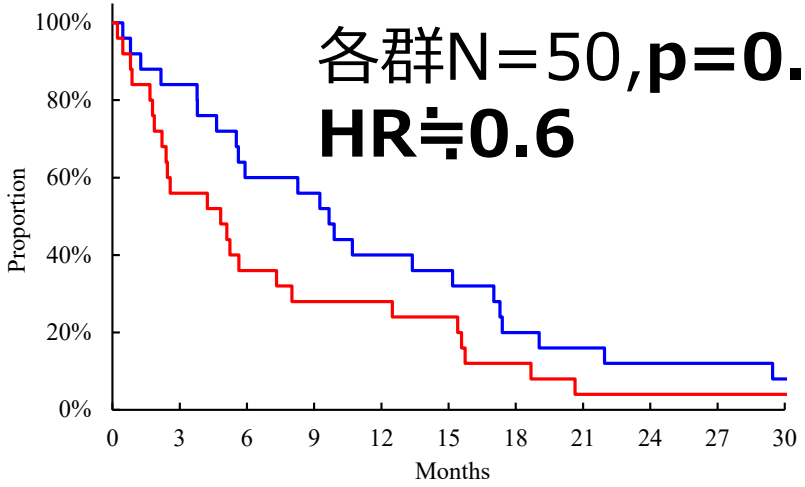




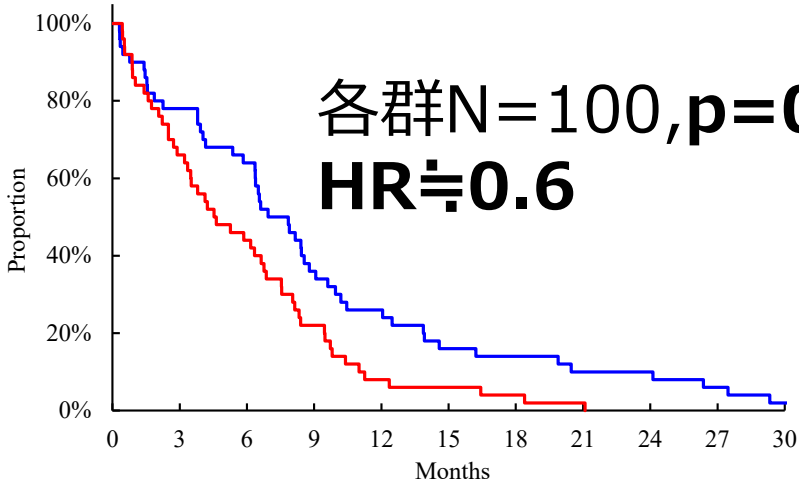


同じ治療効果でも…

サンプルサイズの大きさでp値の大きさは異なる

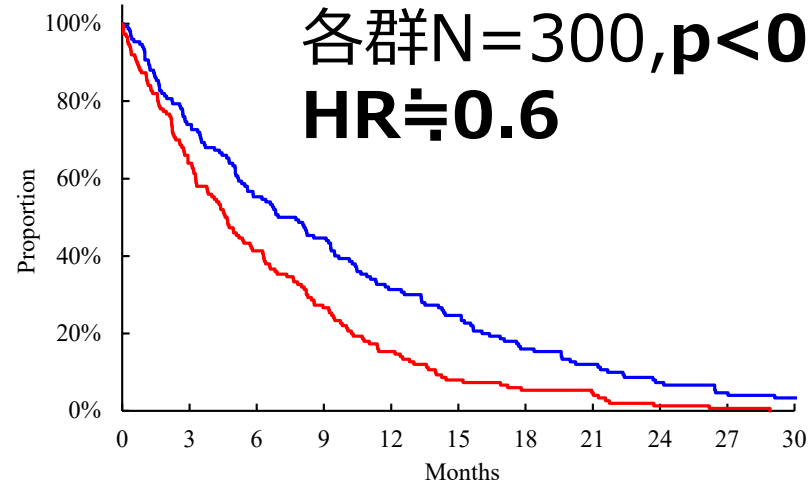


各群N=50,**p=0.10**
HR≒0.6



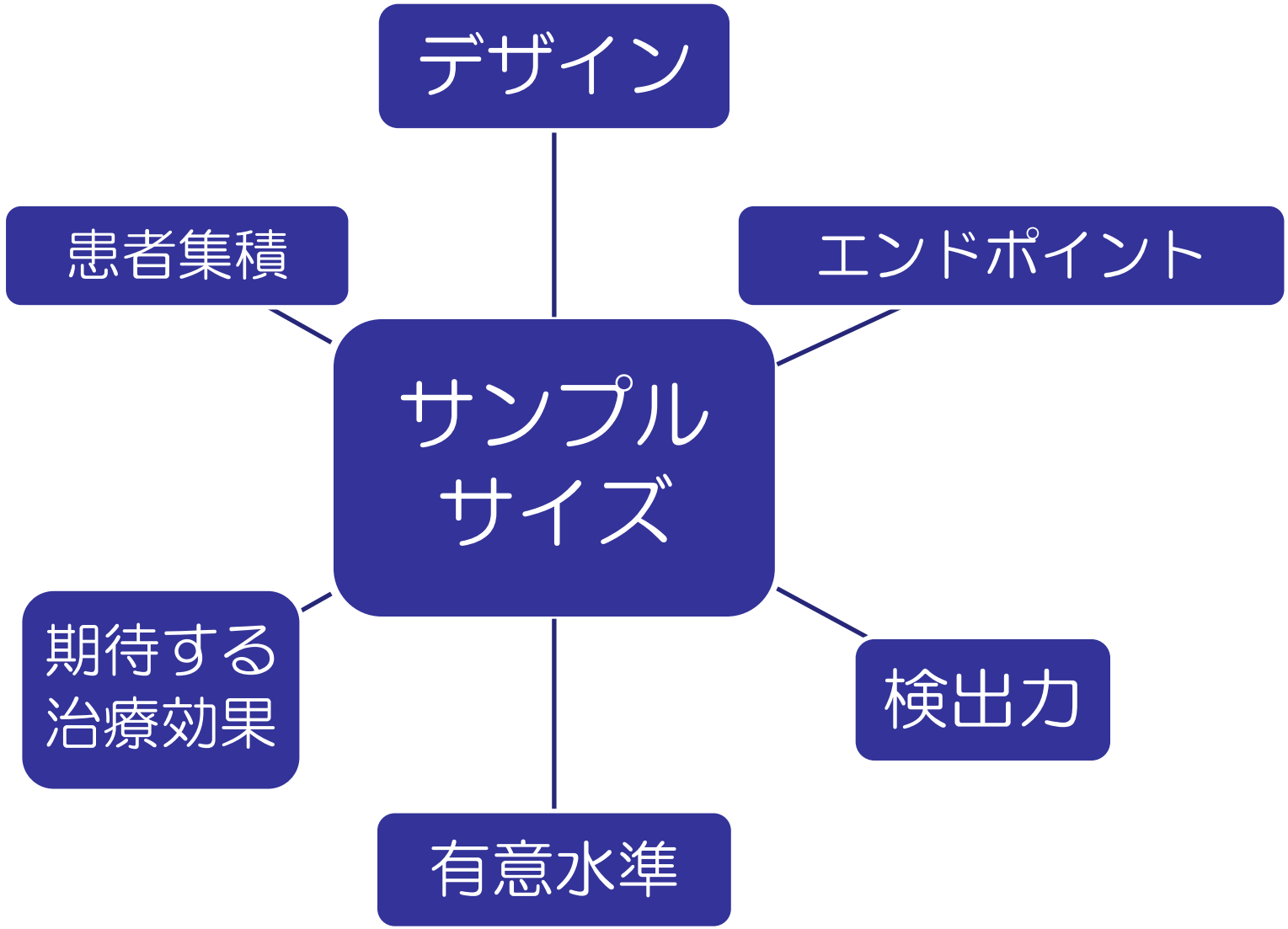
各群N=100,**p=0.01**
HR≒0.6

注：仮想例



各群N=300,**p<0.001**
HR≒0.6

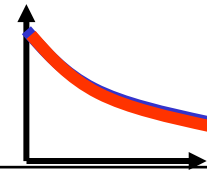
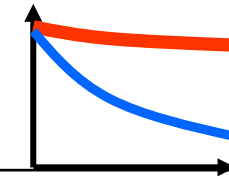
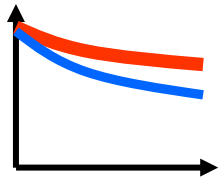
サンプルサイズに関係する要素



デザインによりサンプルサイズは変わる

- 単群試験 vs 2群比較試験
  - –通常は2群比較になった方が必要なN↑
- 優越性試験 vs 非劣性試験
  - –詳細は割愛. ICRwebを視聴してください

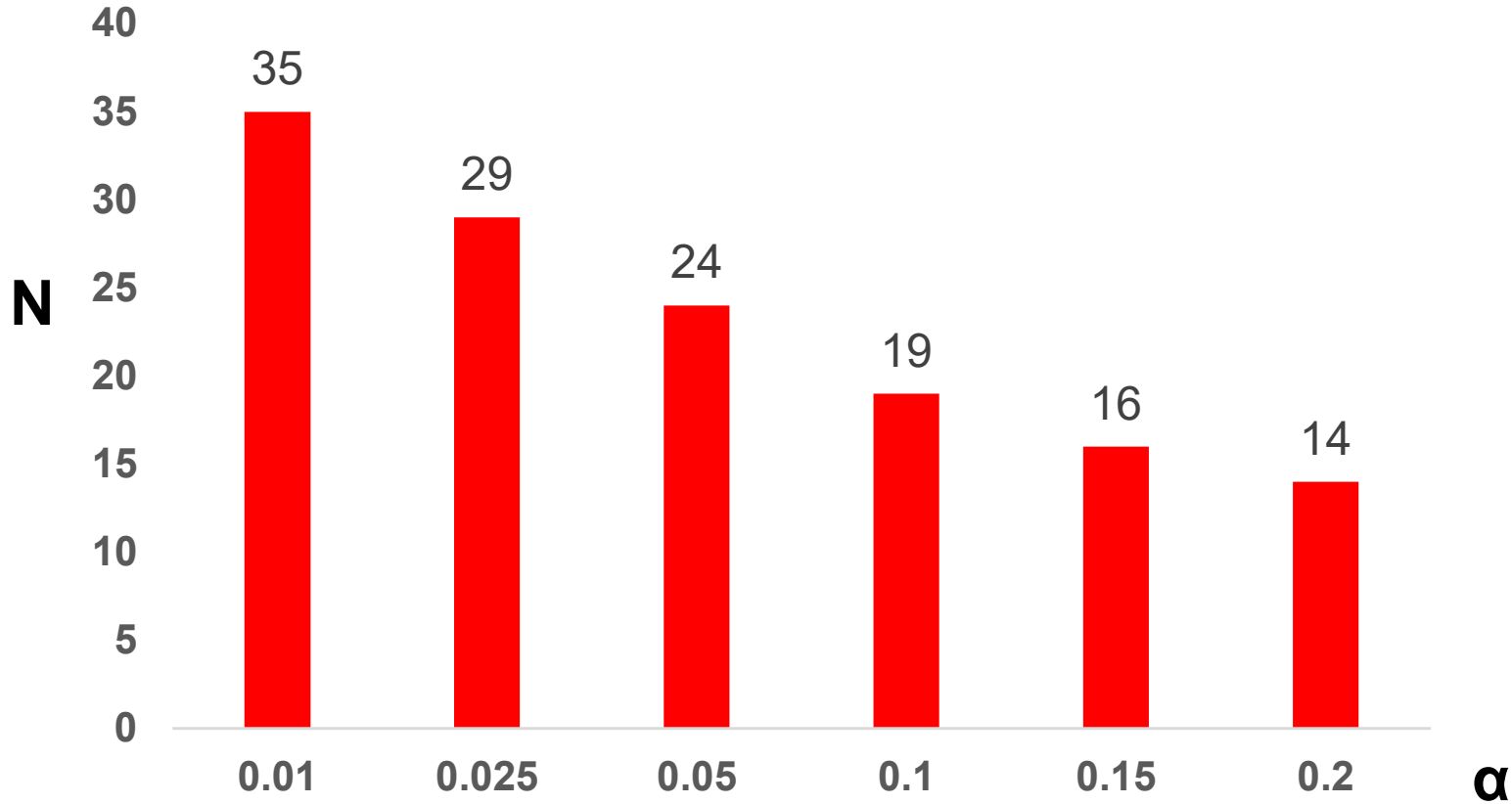
検定における2種類の誤り

		真実	
		帰無仮説（差がない） 	対立仮説（差がある） 
検定結果 	有意差なし	正しい	誤り （**βエラー**）
	有意差あり	誤り （**αエラー**）	正しい （**検出力.1-β**）

- 検定の結果が必ずしも真実を反映しているわけではない
- 検定の誤りを小さくするには？

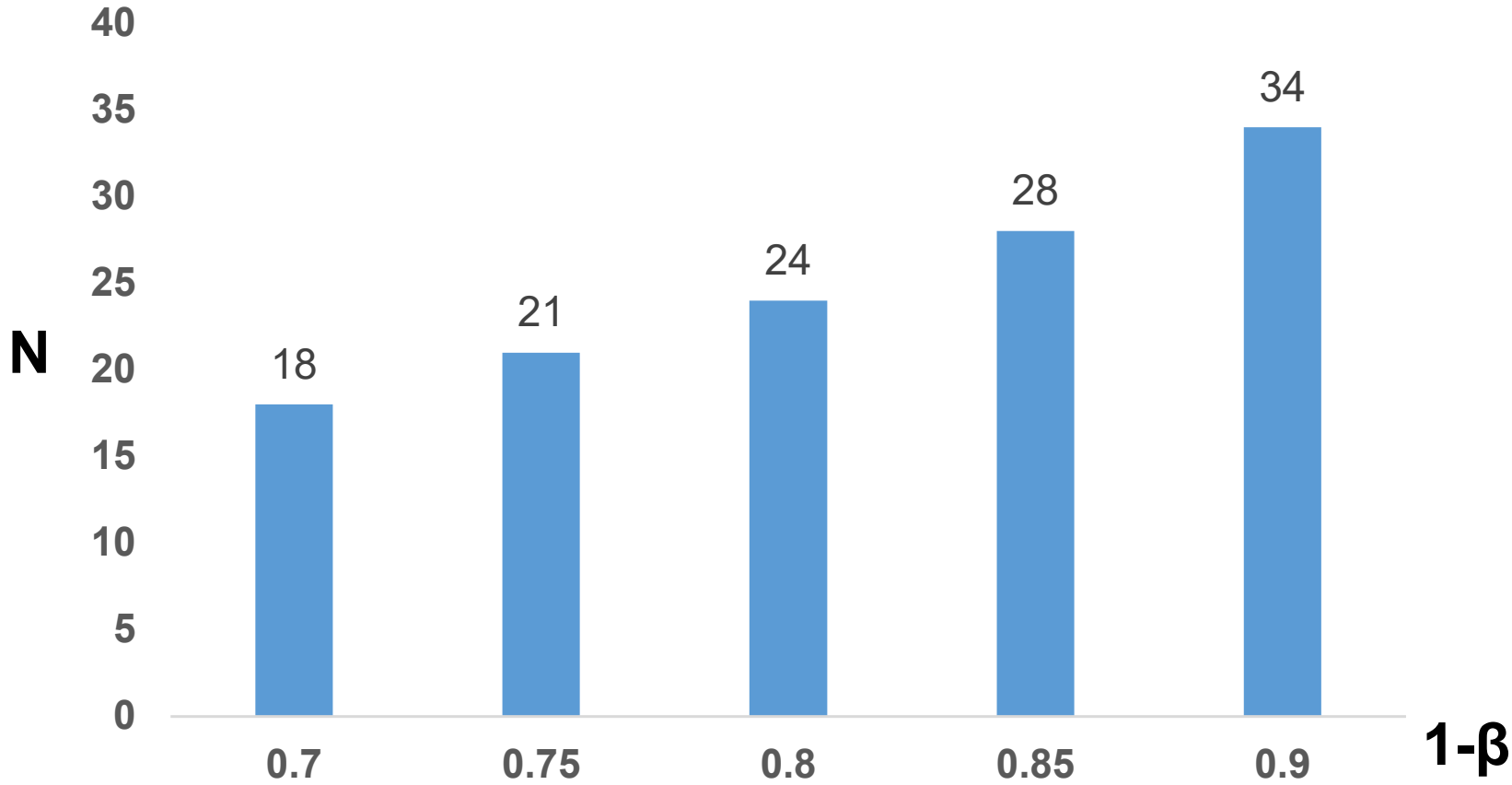
有意水準（α）とNの関係

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



検出力（1-β）とNの関係

- βエラー（効果のある薬を正しく効果ありと判断）を小さくするにはNを増やす
- 例：閾値=10%, 期待値=30%、α=5%



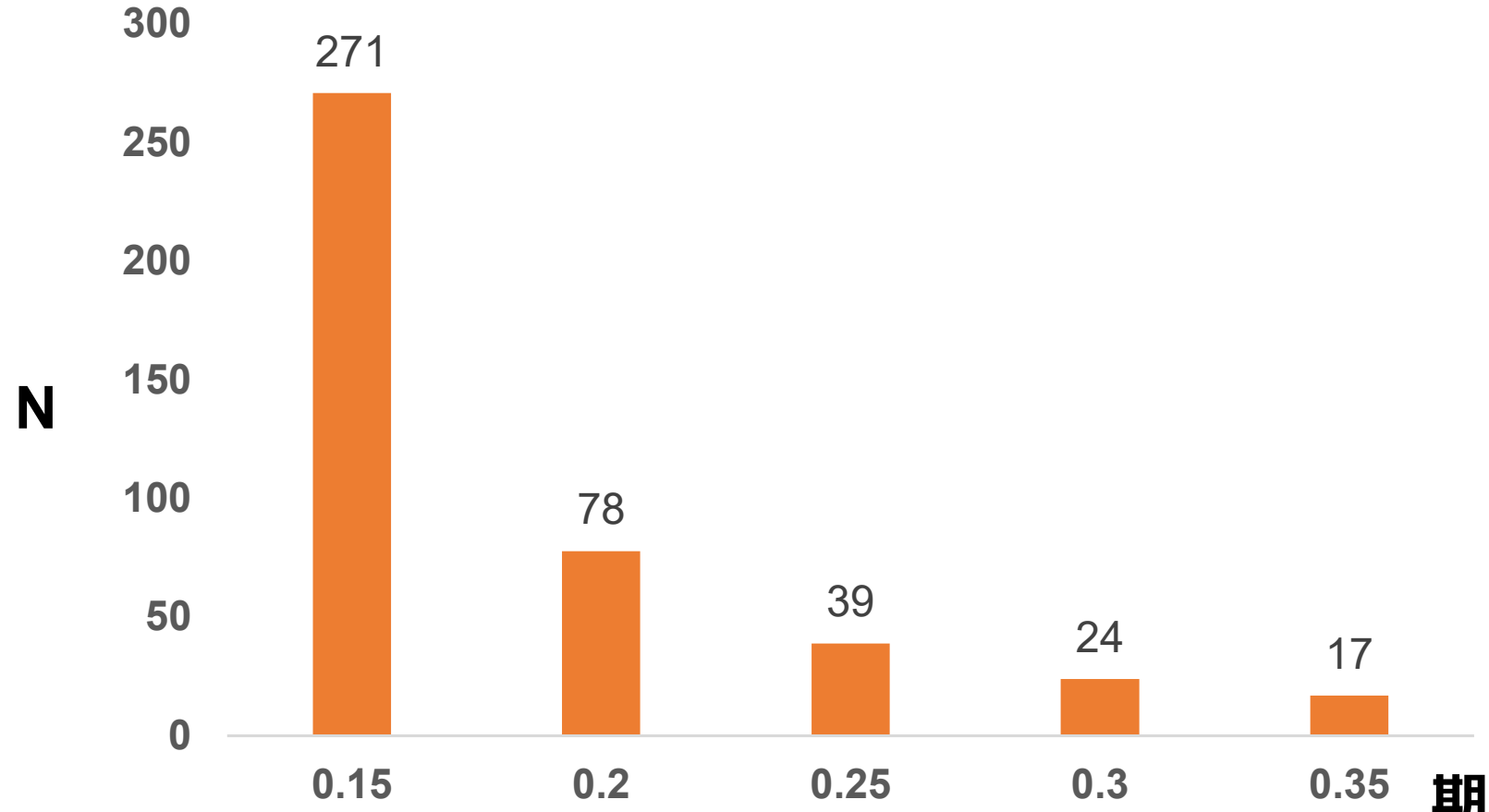
通常のαとβの設定値

- 検証的試験（Phase 3）の場合のα
  - 片側α=2.5% / 両側α=5%がデフォルト
  - ICH E9で決まっている（消費者が負うリスク）
- 検証的試験（Phase 3）の場合の検出力（1-β）
  - 80%以上がデフォルト（企業が負うリスク）
- 検証的試験（Phase 3）の場合、α < β がデフォルト
  - αエラーの方がβエラーよりも社会にとってリスク
  - Nを増やさない限り、αとβを同時に小さくはできない

閾値・期待値とNの関係

- 閾値に比べて大きな治療効果を期待するのであれば、少ないNでOK

  例：α=5%、検出力=80%、閾値=10%の場合



サンプルサイズ計算式の骨格

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

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

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

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

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

2群比較の各検定法のサンプルサイズ計算式の骨格

エンドポイントが
連続値の場合

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

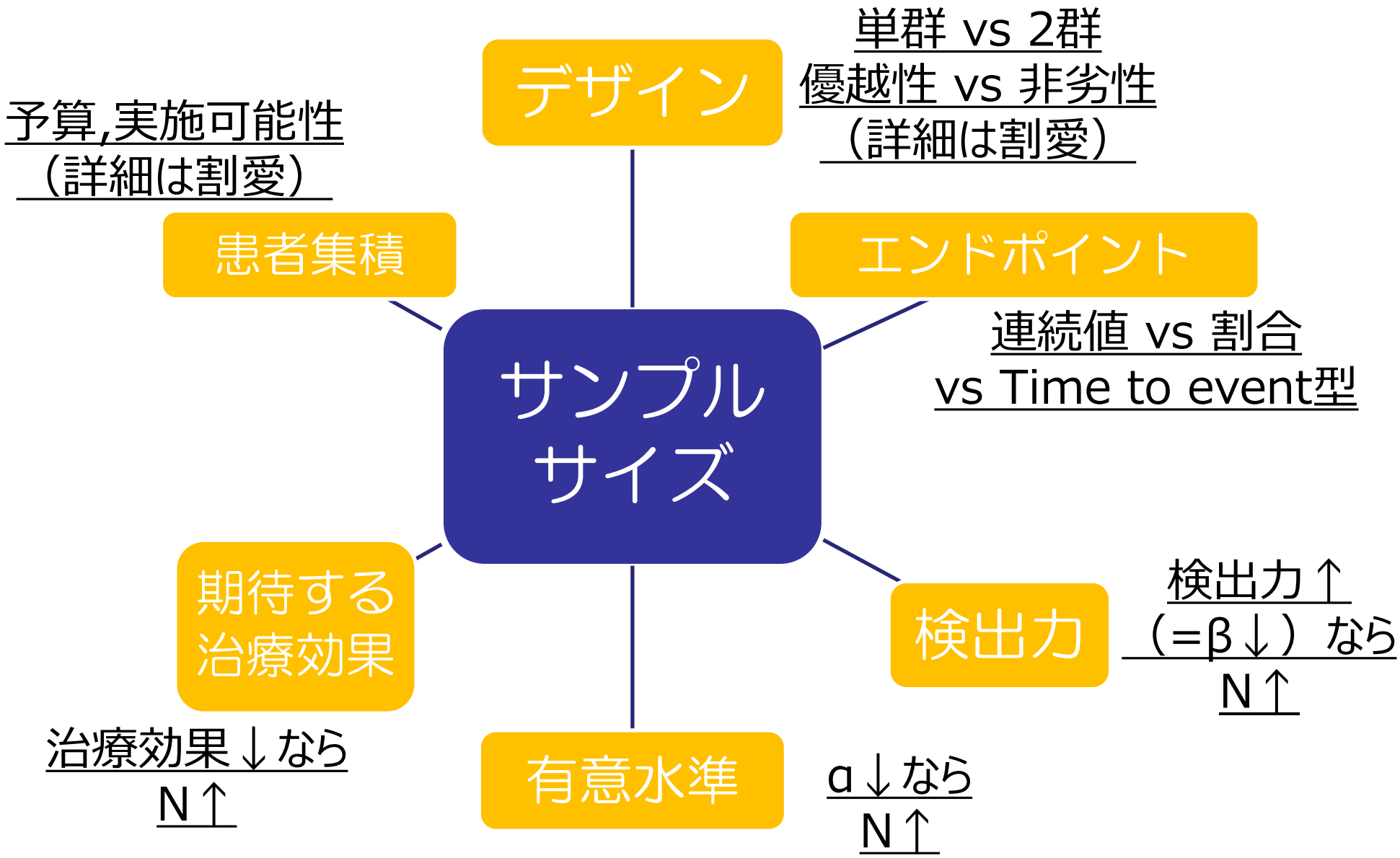
エンドポイントが
割合の場合

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

エンドポイントが
Time-to even型の場合

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

サンプルサイズに関係する要素



【再掲】臨床試験におけるサンプルサイズは、

1.多ければ多い方が良い？

2.少なければ少ない方が良い？

研究デザインに適したサンプルサイズ設定

- 後ろ向き観察研究（いわゆる”レトロな研究”）
  - 多ければ多いほどよい（例数を決めて絞り込むことで、かえってバイアスが入ることもある）
  - 既に収集済みのデータのため倫理的なハードルは低い
- 前向き観察研究
  - 多ければ多いほどよいが…リソース次第
  - 統計的な観点から結果を論じたい場合や侵襲を伴う治療の場合はサンプルサイズ計算によって決めるのが望ましい場合が多い
- 臨床試験（介入研究）
  - **臨床的に意義のある差を検出するのに必要な最少の例数**に設定すべき

まとめ

- ランダム化とは
  - 担当医や患者の意思によらず確率に基づき各治療群に患者を割り付ける行為
  - （未知の因子も含めて）交絡を排除し比較可能性を担保
- がん臨床試験の結果の解釈
  - 仮設検定、検定p値
  - 統計的な差と臨床的な差の違い
- サンプルサイズ設定の必要性
  - サンプルサイズに関係する要素
  - 研究デザインに適した設定が必要