



# 第1回：仮説検定1

本セミナーの内容はJCOG臨床試験セミナーの水澤純基先生の  
スライドをもとに作成させていただきました

生物統計部門 口羽文

生物統計基礎セミナー 2013/11/21(木)



# 仮説検定に関する講義

## 仮説検定1: 仮説検定の考え方

 理論, 解釈, 限界

## 仮説検定2: 仮説検定のいろいろな方法

 t検定,  $\chi^2$ 検定, などの紹介

# どんなときに検定が使われているか

## QOLスコアの平均値の比較

	A群	B群	C群	P値
QOLスコア	19.4	12.6	16.7	0.0015

Dabakuyo et al. (2009) 20, 1352-1361 Ann Oncol

## 家族歴の有無

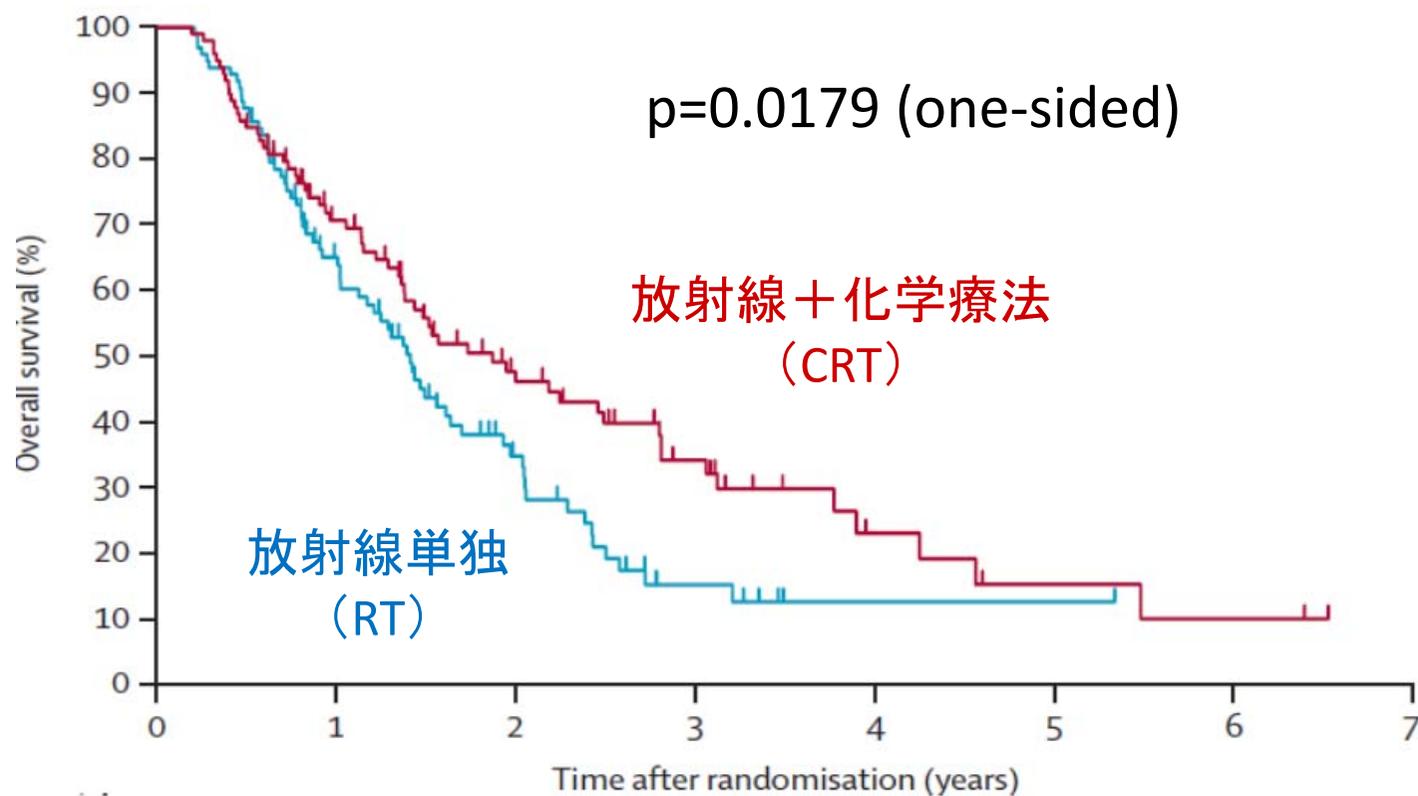
家族歴	ケース群	コントロール群	P値
あり	120人	156人	
なし	268人	232人	0.201

Ma et al. (2009) 61(4), 447-456 Nutr Cancer

## 生存曲線の比較

など

今日の目標：  
検定結果を正しく解釈できるようになること



Atagi *et al.* (2012) *Lancet Oncology* 13(7): 671-8.

生物統計基礎セミナー 2013/11/21

4

# 流れ

## 仮説検定の考え方

-  帰無仮説

-  対立仮説

-  P値

## 仮説検定に伴う誤り

-   $\alpha$ エラー

-   $\beta$ エラー

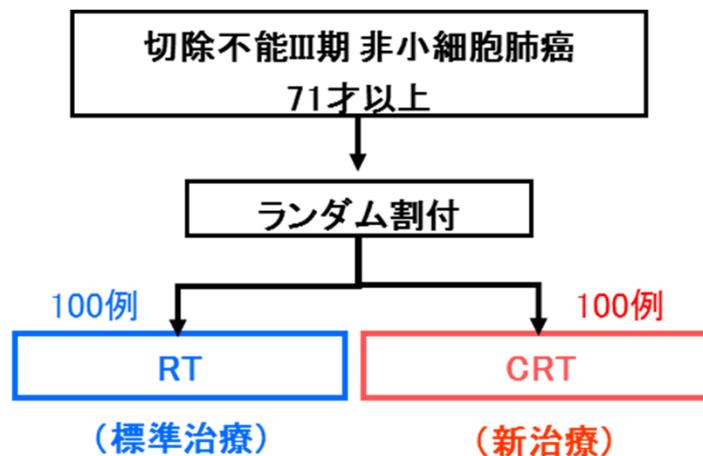
-  検出力 (power)

## サンプルサイズ設計

## 仮説検定の限界と補足

## JCOG0301研究の目的

- 👤 対象: 71才以上, 切除不能Ⅲ期非小細胞肺癌患者
- 👤 研究課題: 放射線+化学療法(CRT)は放射線単独(RT)に比べ死亡を減らすことができるか?

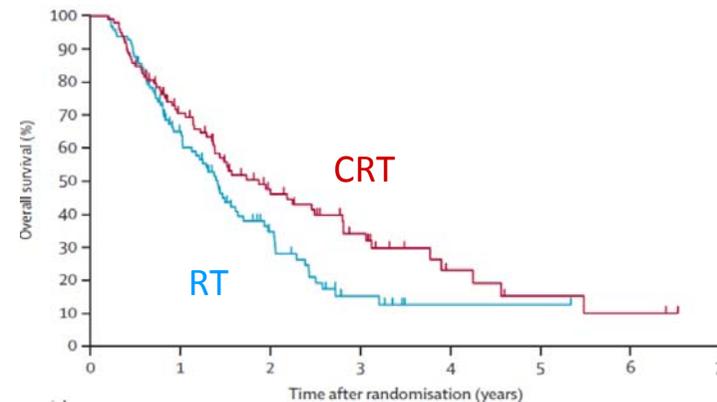


# 研究課題の評価

👤 研究課題：  
「CRTはRTよりも死亡を減らすことができるか？」

👤 2群間の生存曲線を比較

👤 「生存曲線間に差がある」といえるかどうかを評価



## 2群比較: 仮説検定

👤 「RTとCRTに差がある」かどうかを検討したい

👤 「RTとCRTに差がない」という真逆の仮定をおき  
データがこの仮定にどの程度矛盾するかを  
評価することによって結論を導く方法

🌿 背理法によって結論を導く

## 2群比較: 仮説検定

### 1. 帰無仮説の設定

- 🌸 「RTとCRTに差がない」

### 2. 帰無仮説の下での分布

- 🌸 「RTとCRTに差がない」と仮定した下で, 何回も同じ試験を繰り返した場合に得られる結果の分布

### 3. P値の計算

- 🌸 観測されたRTとCRTの差以上に大きな差になる確率を求める

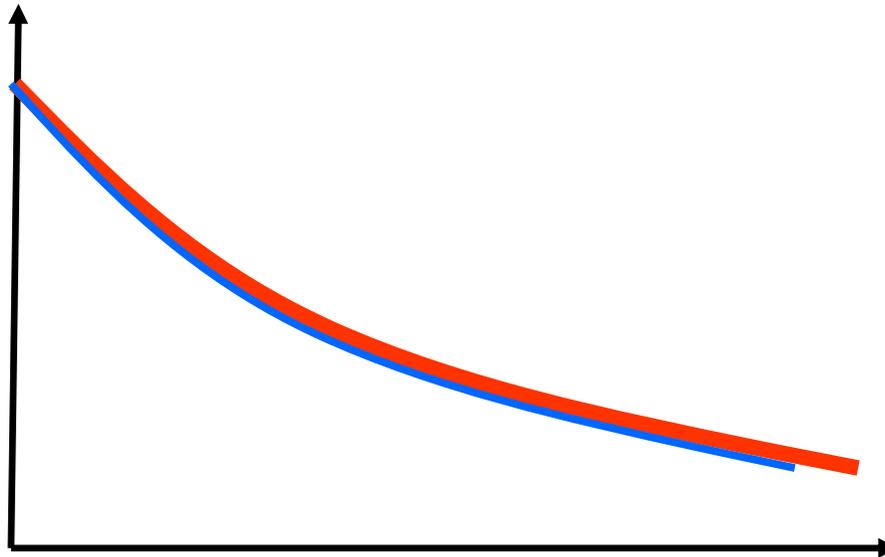
### 4. P値に基づく判断

- 🌸 P値が小さければ「RTとCRTに差がない」の仮定が間違っていたと判断し, 「RTとCRTに差がある」と判断する
- 🌸 「RTとCRTに差がある」を対立仮説という

# 1. 帰無仮説の設定

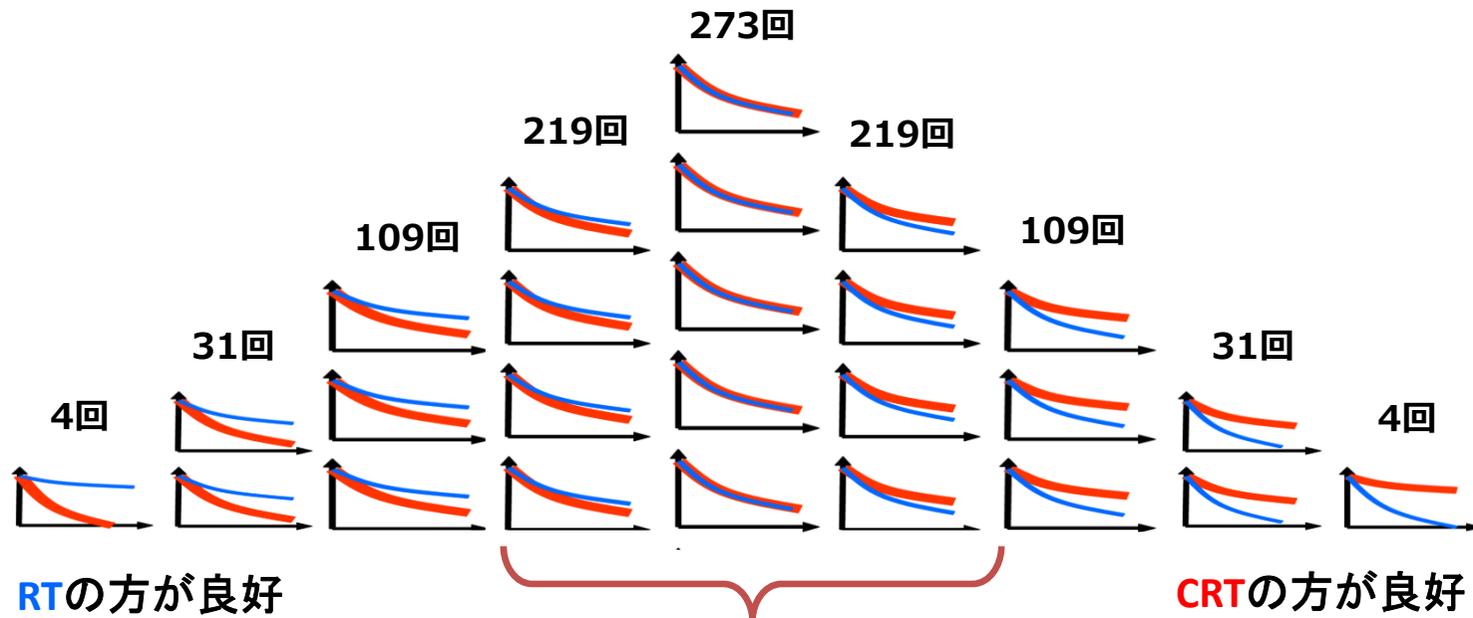
🎅 英語では, Null Hypothesis

🎅 「RTとCRTに差がない」



## 2. 帰無仮説の下での分布

👤 対象者数200人の試験を1000回繰り返すと・・・



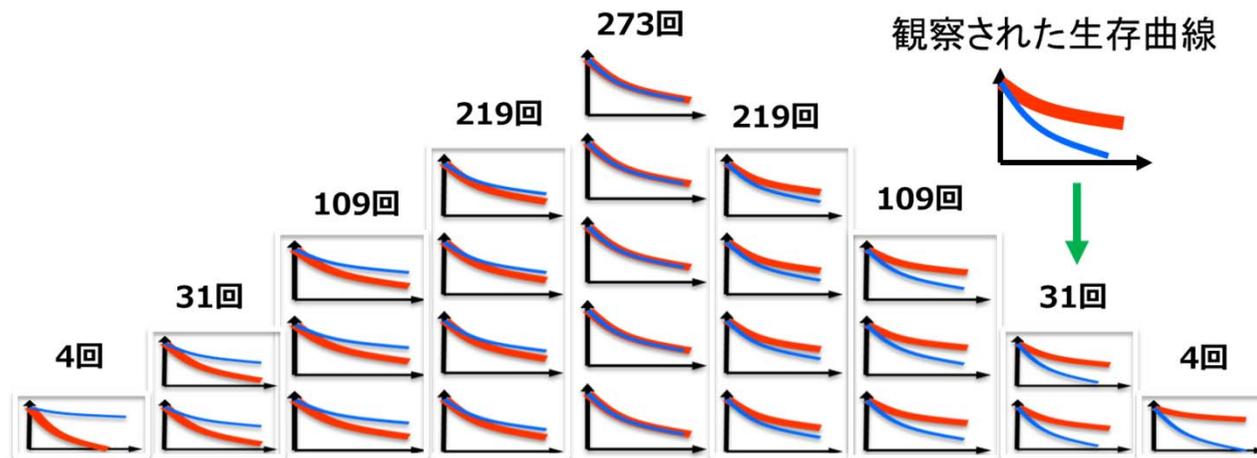
【差がない】結果が最も多く観察される

生物統計基礎セミナー 2013/11/21

11 

### 3. P値の計算

- 🎅 帰無仮説の下での分布と観察データとの比較
- 🎅 観察されたRTとCRTの差以上に大きな差になる確率
- 🎅 P値 =  $35/1000 = 0.035$ 
  - 🌿 【差がない】が真実だとしたら、1000回中35回くらい起きそうな結果



生物統計基礎セミナー 2013/11/21

12

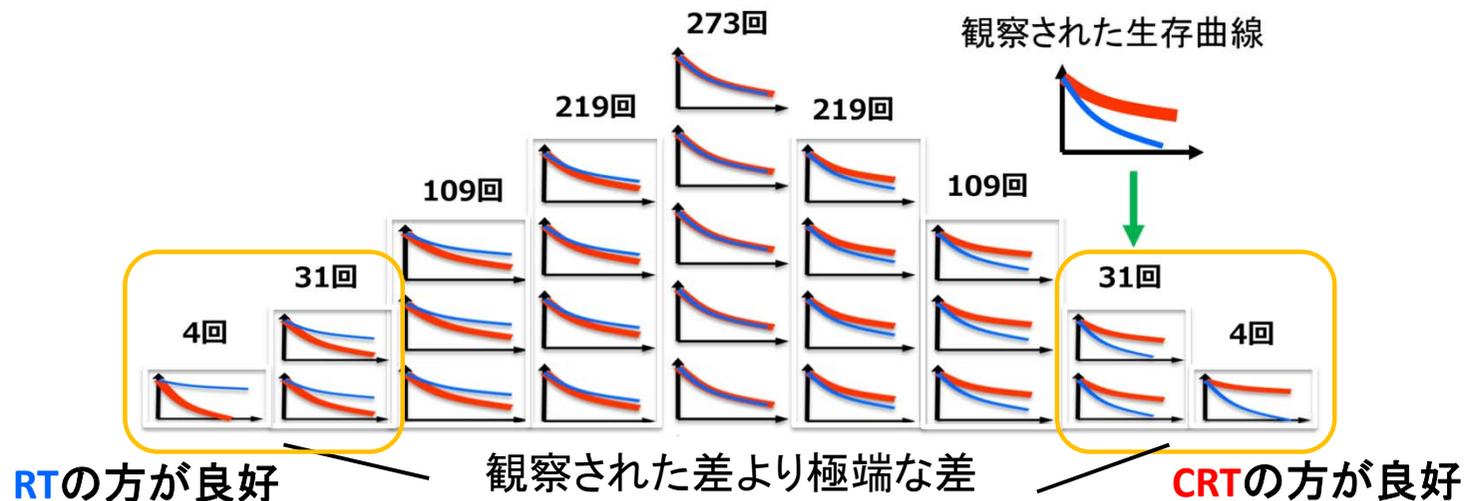
### 3(注). P値の計算 一片側と両側一

🎅 片側P値 =  $35/1000 = 0.035$

🎄 対立仮説「 $RT < CRT$ 」

🎅 両側P値 =  $70/1000 = 0.07$

🎄 対立仮説「 $RT \neq CRT$ 」



生物統計基礎セミナー 2013/11/21

13 🎄

### 3. P値の計算 — 解釈 —

👤 P値とは,  
帰無仮説「RTとCRTに差がない」の下で  
データよりも極端なことが起きる確率

👤 P値が大きい

- ・・・当たり前なことが起きた
- ・・・帰無仮説の仮定は合っていたのだろう

👤 P値が小さい

- ・・・稀なことが起きた
- ・・・帰無仮説の仮定が間違っていたのかも

## 4. P値に基づく判断

👤 P値 =  $35/1000 = 0.035$

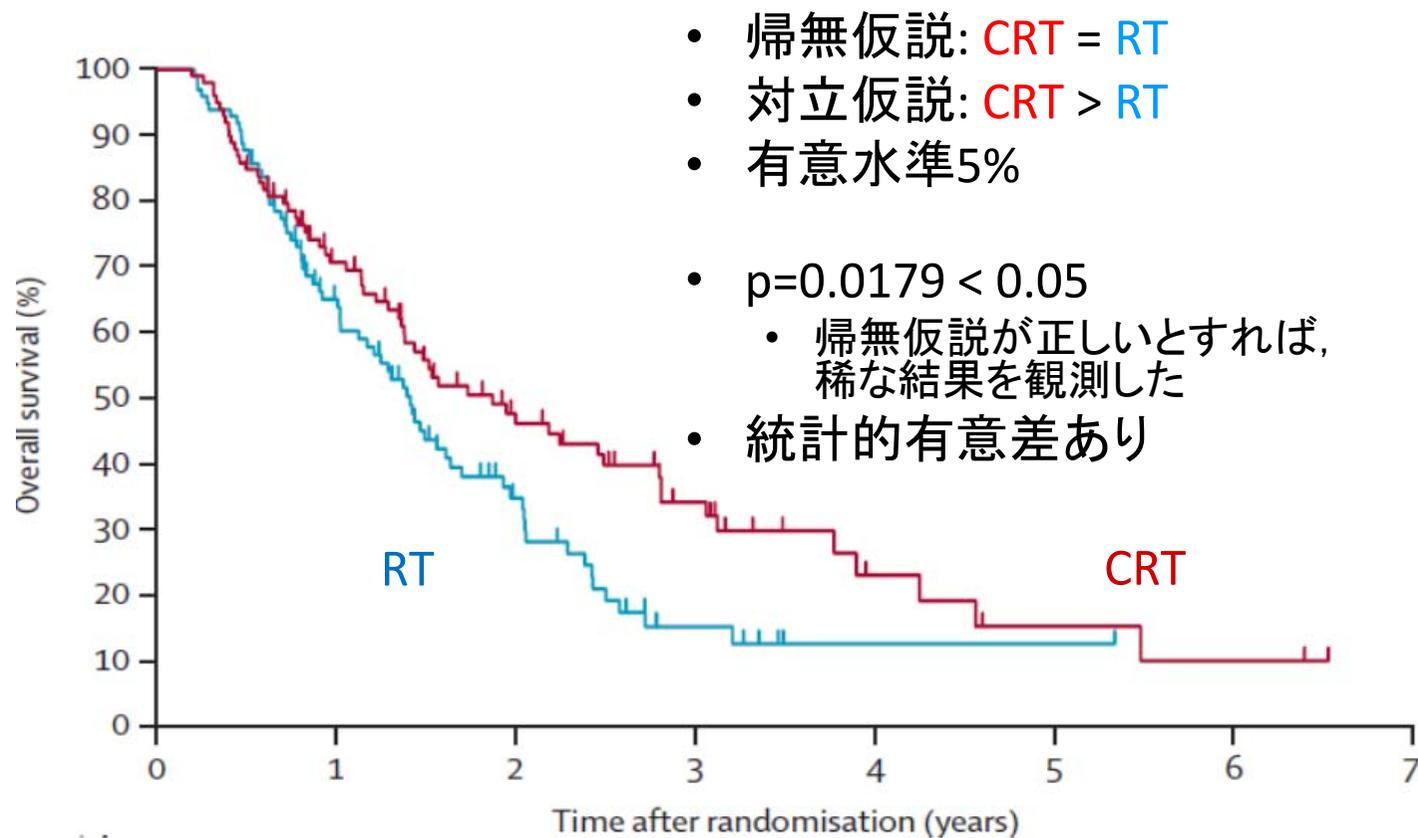
🌿 【差がない】が真実だとしたら, かなり稀な結果を観測した

🌿 【差がない】という前提が間違っているのかも

👤 **十分に稀な結果**といえる場合,  
帰無仮説【差がない】を棄却し,  
対立仮説【差がある】を採択する

👤 十分に稀かどうかの基準・・・**有意水準**  
🌿 後付けにならないように事前に決め, 明記する

## JCOG0301の結果



Atagi *et al.* (2012) *Lancet Oncology* 13(7): 671-8.

生物統計基礎セミナー 2013/11/21

16

## 2群比較: 仮説検定

### 1. 帰無仮説の設定

- 🌸 「RTとCRTに差がない」

### 2. 帰無仮説の下での分布

- 🌸 「RTとCRTに差がない」仮定した下で, 何回も同じ試験を繰り返した場合に得られる結果の分布

### 3. P値の計算

- 🌸 観測されたRTとCRTの差以上に大きな差になる確率を求める

### 4. P値に基づく判断

- 🌸 P値が小さければ「RTとCRTに差がない」の仮定が間違っていたと判断し, 「RTとCRTに差がある」と判断する
- 🌸 対立仮説「RTとCRTに差がある」

# 流れ

## 仮説検定の考え方

-  帰無仮説

-  対立仮説

-  P値

## 仮説検定に伴う誤り

-   $\alpha$ エラー

-   $\beta$ エラー

-  検出力 (power)

## サンプルサイズ設計

## 仮説検定の限界と補足

## $\alpha$ エラー

- 👤 真に【差がない】ときに【差がある】と判断してしまう誤り
- 👤 偽陽性 (false positive)
- 👤 第一種の過誤 (type I error) ともいわれる

## βエラー

- 👤 真に【差がある】ときに【差がない】と判断してしまう誤り
- 👤 偽陰性 (false negative)
- 👤 第二種の過誤 (type II error) ともいわれる
- 👤 検出力 (power) =  $1 - \beta$ 
  - 🌿 真に【差がある】ときに【差がある】といえる確率

## 仮説検定による判断のまとめ

		真 実	
		帰無仮説が 成り立っている (差がない)	帰無仮説が 成り立っていない (差がある)
判 断	帰無仮説を保留	正しい	誤り $\beta$ エラー
	帰無仮説を棄却	誤り $\alpha$ エラー	正しい

 検定結果はいつも正しいわけではない

## $\alpha$ エラーと $\beta$ エラーの関係

🧑  $\alpha$ エラーを小さくしたい



🧑 常に帰無仮説を保留



🧑  $\alpha$ エラー = 0%

🧑  $\beta$ エラー = 100%

🧑  $\beta$ エラーを小さくしたい



🧑 常に帰無仮説を棄却



🧑  $\alpha$ エラー = 100%

🧑  $\beta$ エラー = 0%

同時に小さくはできない

# 流れ

## 仮説検定の考え方

-  帰無仮説

-  対立仮説

-  P値

## 仮説検定に伴う誤り

-   $\alpha$ エラー

-   $\beta$ エラー

-  検出力 (power)

## サンプルサイズ設計

## 仮説検定の限界と補足

# サンプルサイズ設計

 研究課題に対して十分な判断を行うのに必要な対象者数は？

-  少なすぎれば, 研究課題に答えが出せない (power小)
-  多すぎれば, 無駄なコスト, 非倫理的

# サンプルサイズ設計に必要な3つの情報

## 治療効果の大きさ: $\Delta$ (デルタ)

-  新治療が, 標準治療にとって変わるのに必要な臨床的意味のある最小の治療効果の差

## 有意水準: $\alpha$

-  許容する $\alpha$ エラー

## 検出力 (power): $1-\beta$

-  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

## JCOG0301のサンプルサイズ設計

👤 治療効果の大きさ:

RT群での生存時間の中央値10か月

CRT群での生存時間の中央値15か月

治療効果の大きさ  $\Delta = 5$  か月

👤 有意水準 = 5%

👤 検出力 = 80%

👤 サンプルサイズ = 200

## 単群, 奏効割合に対するサンプルサイズ設計 (JCOG1002)

🧑‍🎄 効果の大きさ:

閾値奏効割合 = 65%

期待奏効割合 = 80%

🧑‍🎄 有意水準 = 10%

🧑‍🎄 検出力 = 80%

🧑‍🎄 サンプルサイズ = 50

Katayama *et al.* (2012) *Jpn J Clin Oncol* 42(6): 556-9.

生物統計基礎セミナー 2013/11/21

27

# サンプルサイズ設計ツール

## SWOG Statistical Center: Statools

 <http://www.swogstat.org/stat/public/>



### One Sample Binomial

#### Select Calculation and Test Type

<input checked="" type="radio"/> Sample Size	<input checked="" type="radio"/> 1 Sided
<input type="radio"/> Power	<input type="radio"/> 2 Sided

#### Select Hypothesis Test Parameters

Null Proportion	Alternative Proportion	Alpha
0.65	0.80	.1

Calculate Power/Sample Size

Power	Sample Size	Approx Lower Count Critical Value -1	Approx Upper Count Critical Value 30
0.80	40		

## $\alpha$ , $1-\beta$ の設定

 有意水準 $\alpha$  …… 普通は5~10%

 (被験者保護の観点から)効果のない治療を良いと判断してしまう誤りはできるだけ小さく

 検出力 $1-\beta$  …… 普通は80%~90%

 (主に試験実施側の観点から)効果のある治療は正しく効果があると言えるように

 被験者保護の観点より

  $\alpha$ を小さくすることを優先すべき

  $\alpha$ を小さくした上で検出力を80%以上にするように必要対象者数を決める

# 流れ

## 仮説検定の考え方

-  帰無仮説

-  対立仮説

-  P値

## 仮説検定に伴う誤り

-   $\alpha$ エラー

-   $\beta$ エラー

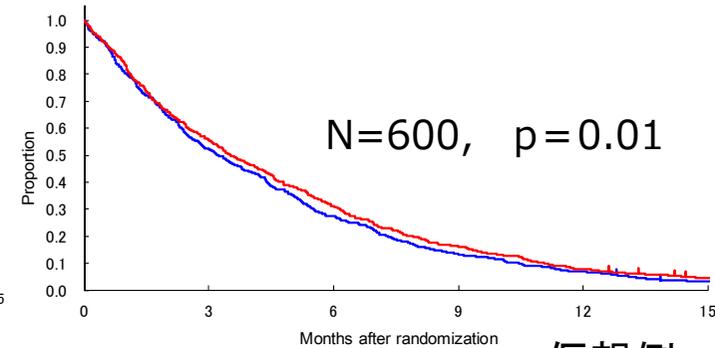
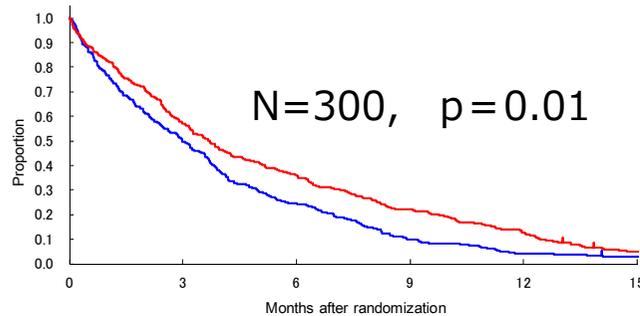
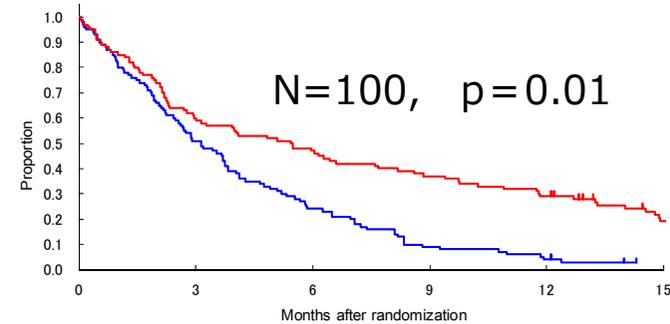
-  検出力 (power)

## サンプルサイズ設計

## 仮説検定の限界と補足

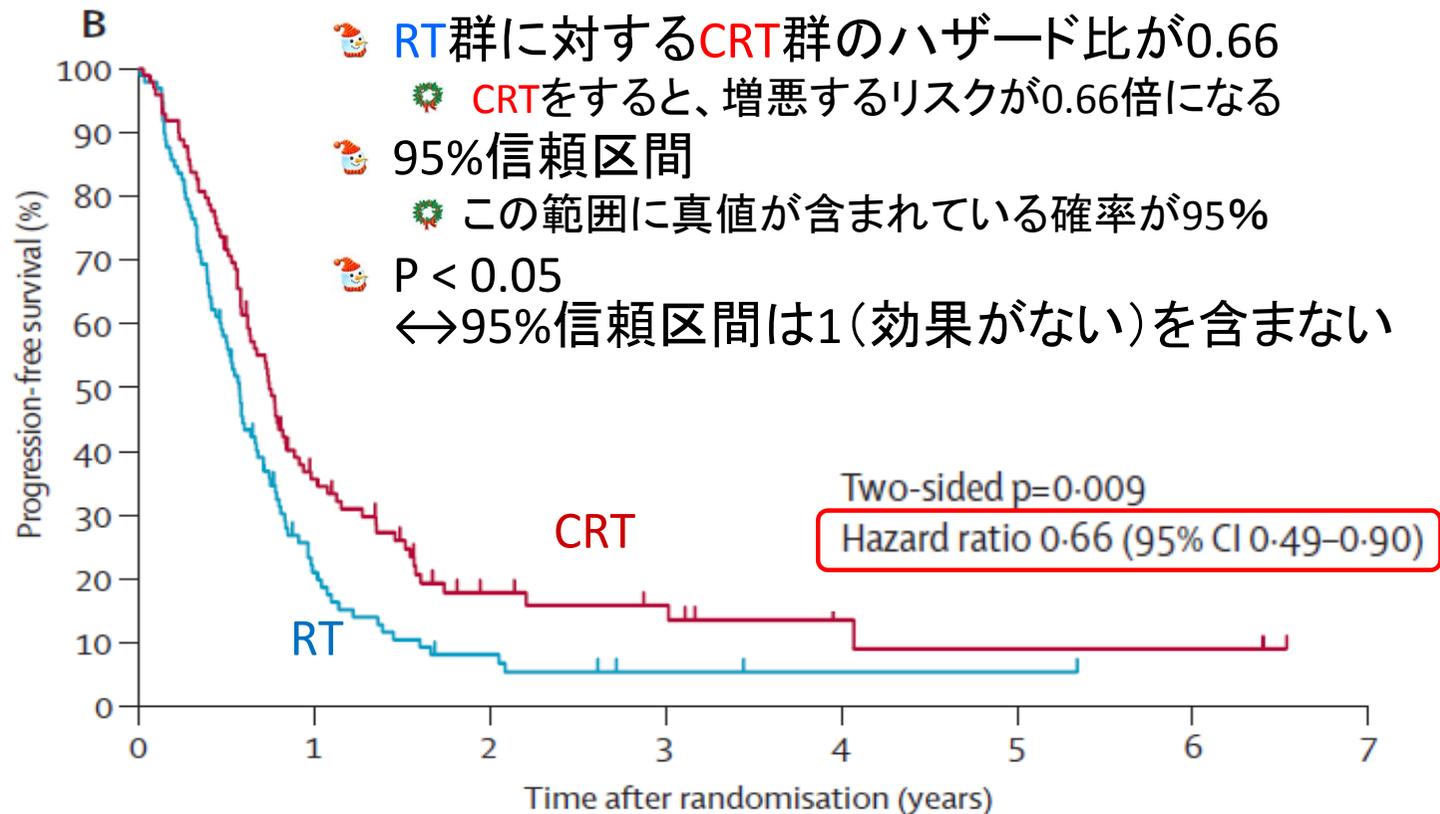
# 限界：P値では効果の大きさは分からない

- 🧑‍🎓 P値が小さい研究の方が治療効果が大きいというわけではない
- 🧑‍🎓 治療効果の大きさは「推定」する必要がある



仮想例

## 補足1: 効果の推定



Atagi *et al.* (2012) *Lancet Oncology* 13(7): 671-8.

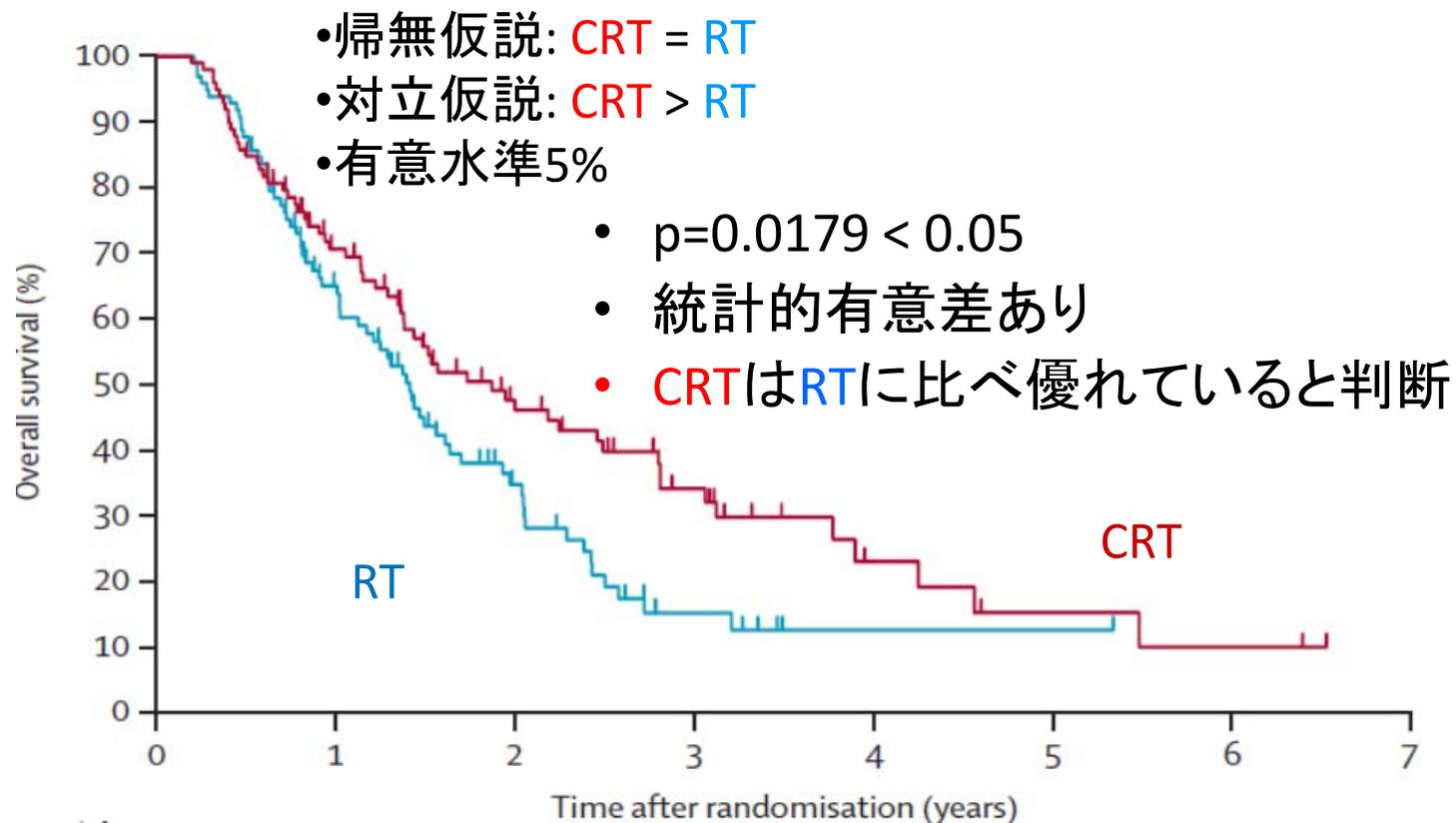
生物統計基礎セミナー 2013/11/21

32 🌿

## 補足2: 検定の多重性

- 👤 検定を繰り返すことにより、  
少なくとも1つの帰無仮説が誤って棄却される確率が大きくなること
  - 🌿 研究全体でみると、誤って効果ありとされる治療がたくさん生じる
- 👤 A, B, Cの3つの治療法を比較したい
- 👤 A対B, A対C, B対C, と検定(有意水準5%)を3回行う
- 👤 たまたまどこかで群間差がみられる確率 > 5%
- 👤 多重性の調整: 個々の検定で許容する $\alpha$ エラーを小さくすることで試験全体に対する $\alpha$ エラーを制御すること
  - 🌿 E.g., Bonferroniの方法: 各検定の有意水準 =  $0.05 / (\text{検定の数})$
  - 🌿 ある群間比較のp値が0.03と得られても,  
3回の検定を行っているので  $0.03 > 0.05/3 = 0.017 \dots$  有意差なし

## 今日のまとめ： 検定結果を正しく解釈する



Atagi *et al.* (2012) *Lancet Oncology* 13(7): 671-8.

生物統計基礎セミナー 2013/11/21

34