JCOG
Japan Clinical Oncology Group

がん臨床試験で必要な最低限の統計知識

~第3相ランダム化比較試験の結果を解釈するためのポイント~

JCOGデータセンター

町田 龍之介

第25回JCOG臨床試験セミナー（入門編）

2022年10月8日 (土)

ランダム化比較試験の結果を解釈したい！

肺がん内科グループ
JCOG0301

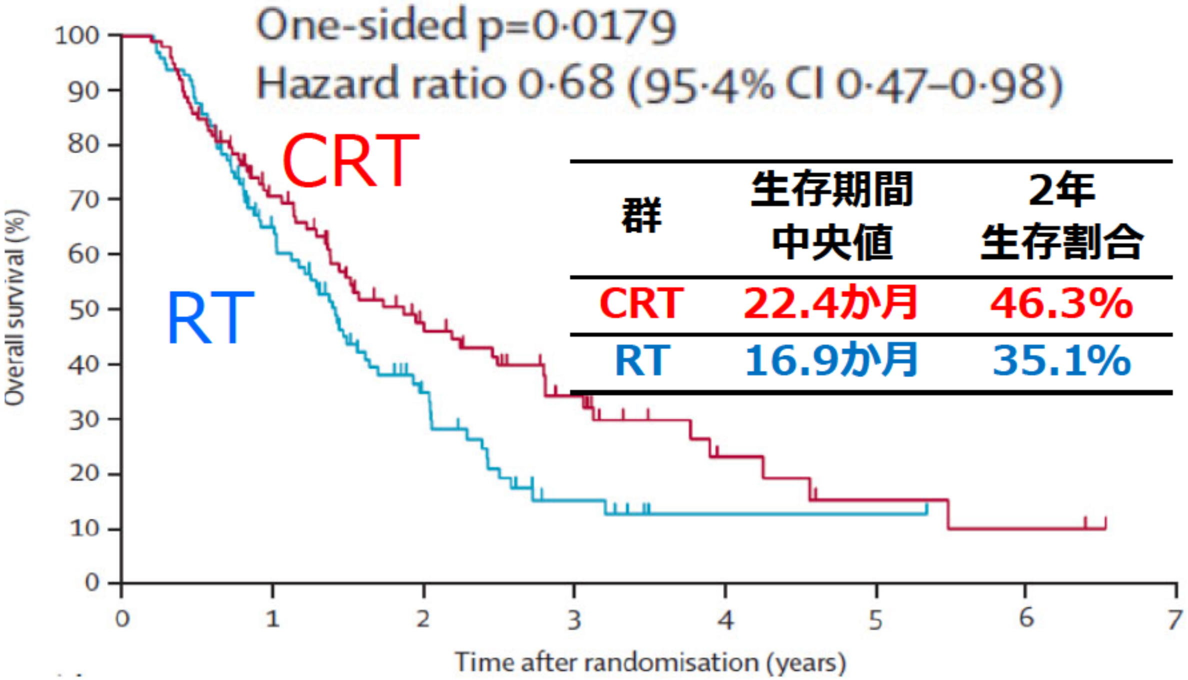
切除不能III期 非小細胞肺癌
71才以上

↓

ランダム割付

（標準治療） 100例
放射線単独
(RT)
↓
RT施行例 98例
RT完遂 93例
RT非完遂 5例
未治療例 2例

（新治療） 100例
放射線+化学療法
(CRT)
↓
CRT施行例 96例
CRT完遂 88例
CRT非完遂 8例
未治療例 4例



群	生存期間中央値	2年生存割合
CRT	22.4か月	46.3%
RT	16.9か月	35.1%

結論：CRTはRTと比較して臨床的に意味のあるベネフィットがあり、CRTはこの対象に対して標準治療として考慮されるべき治療法である

何故このような結論になるの？

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

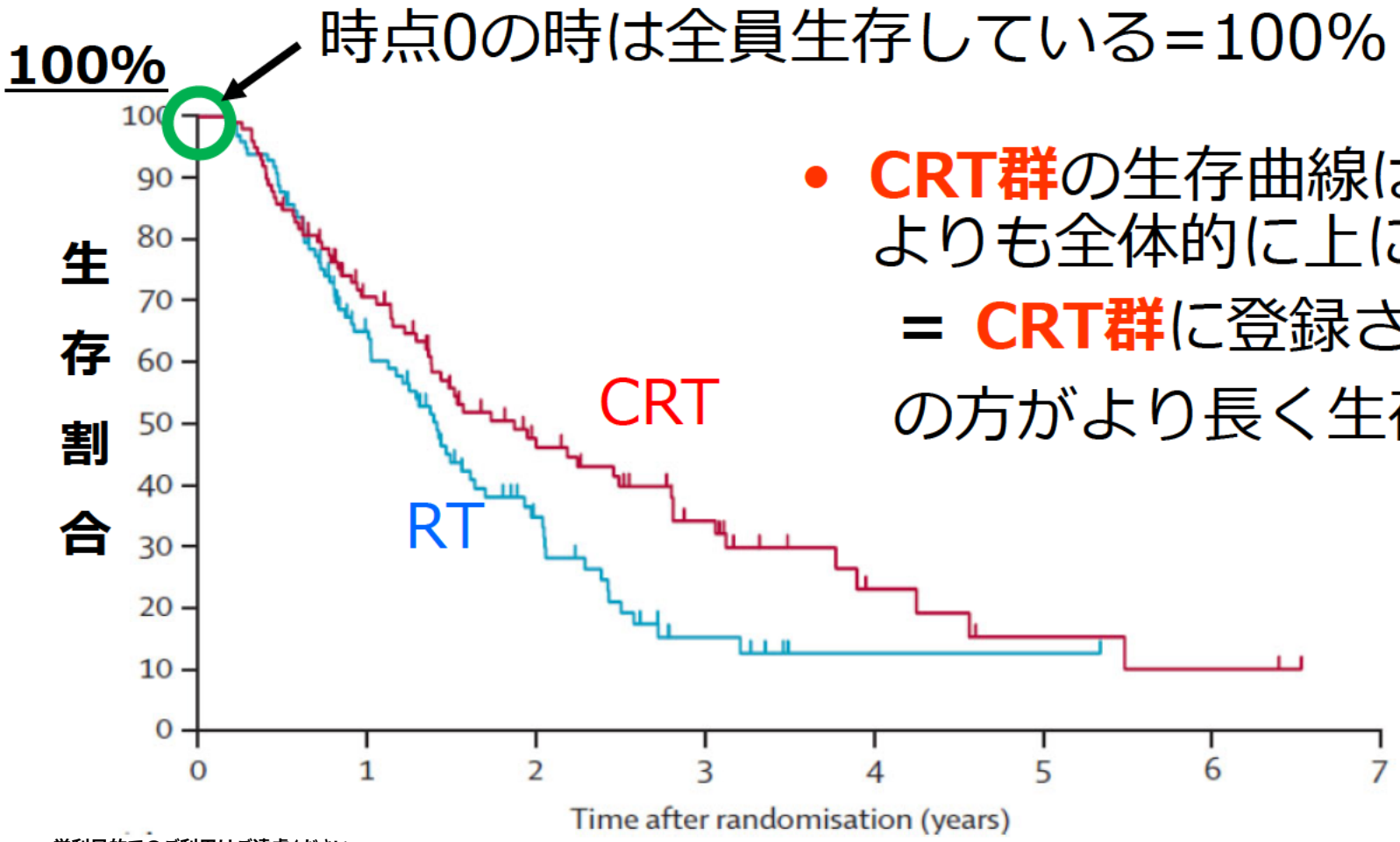
Outline ~ランダム化試験を解釈するために知っておくべきこと~

- **生存曲線**で結果を判断
  - 年次生存割合、生存期間中央値
- **ランダム化**はなぜ必要か？
  - **交絡**と**ランダム化**
- 結果の検証方法
  - **仮説検定**の考え方と**p値**の意味
  - **αエラー**、 **βエラー**、 **検出力**の理解
- 治療効果の大きさの見方
  - **ハザード比**の意味
- 解析対象は？
  - **Intention-To-Treat解析** (ITT解析)

生存曲線

生存曲線とは

- 横軸に時間、縦軸に生存割合をとり、集団における各時点の生存割合をつないだもの
- 死亡が発生するとその時点で生存割合が減少する



- **CRT群**の生存曲線は**RT群**よりも全体的に上にある
  ＝ **CRT群**に登録された集団の方がより長く生存した

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

JCOG 0301　肺がん内科/ NSCLC-高齢者-カルボ放射線 PhaseIII　調査

追跡調査用紙

までにデータセンターに郵送

施設名　○○○○○がんセンター　担当医　○× ○×

患者イニシャル　姓 A 名 A　性別 男　生年月日 昭和30年10月10日

カルテ番号　12345-6789　症例番号

転帰

0□ 生存　最終生存確認日　年 月 日

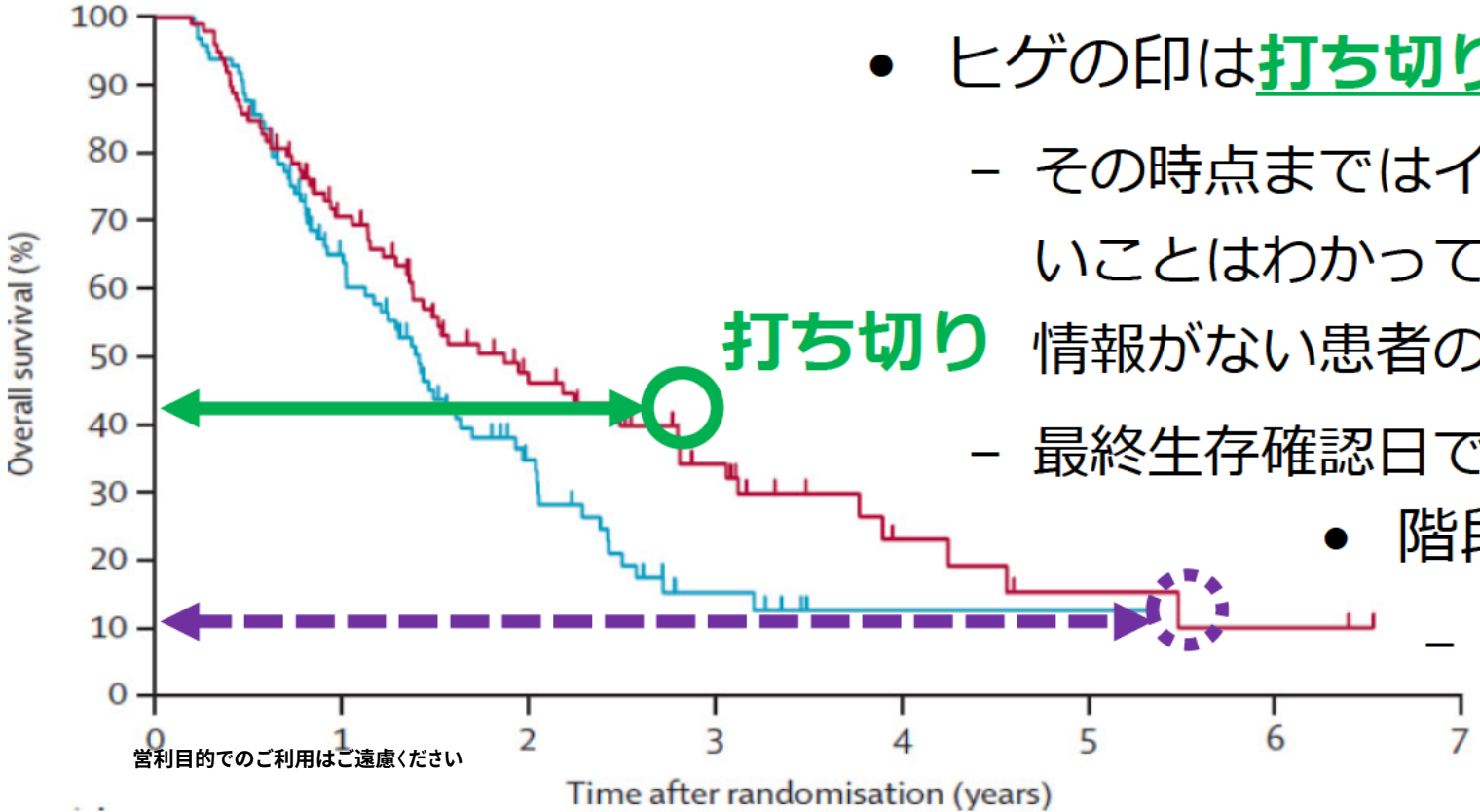
1□ 死亡　死亡日　年 月 日

前回調査の報告

死因　1□ 原病死　2□ 他病死　3□ 治療関連死　8□ その他　9□ 不明

死亡の状況

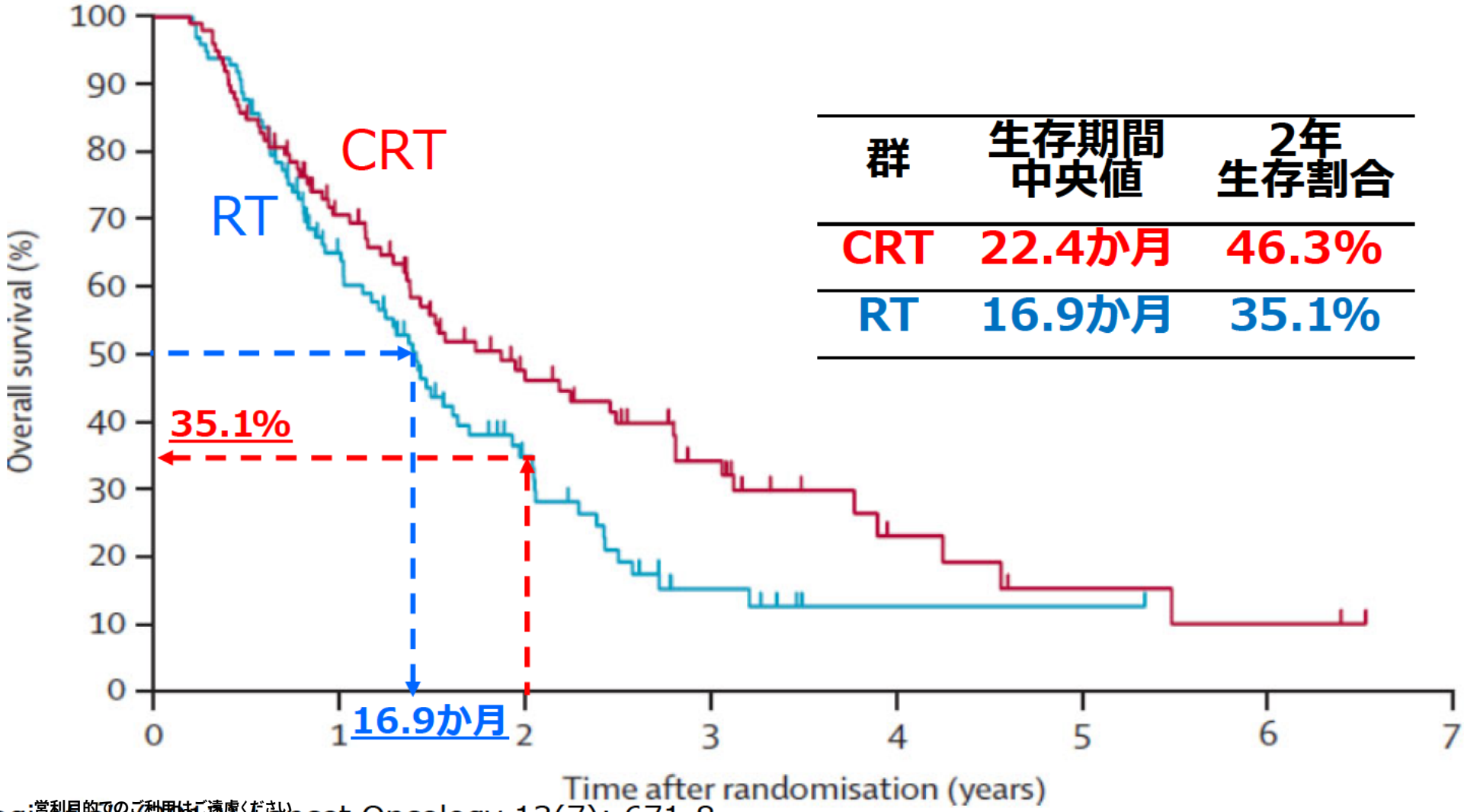
いずれの死因の場合も死亡時の状況を記入



- ヒゲの印は**打ち切り例**を表す
  - その時点までは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のこと
  - 最終生存確認日で打ち切り
- 階段は**死亡例**を表す
  - 死亡日でイベント

生存曲線から得られる要約値

- 生存期間中央値(MST; Median survival time)、年次生存割合
  - RT群の生存期間中央値は**16.9か月**、2年生存割合は**35.1%**



群	生存期間中央値	2年生存割合
CRT	22.4か月	46.3%
RT	16.9か月	35.1%

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

交絡とランダム化

ランダム化？？

ランダム化(ランダム割付)って
何のためにやっているの？
医師や患者が好きな治療をすれば
良いのでは？

肺がん内科グループ
JCOG0301

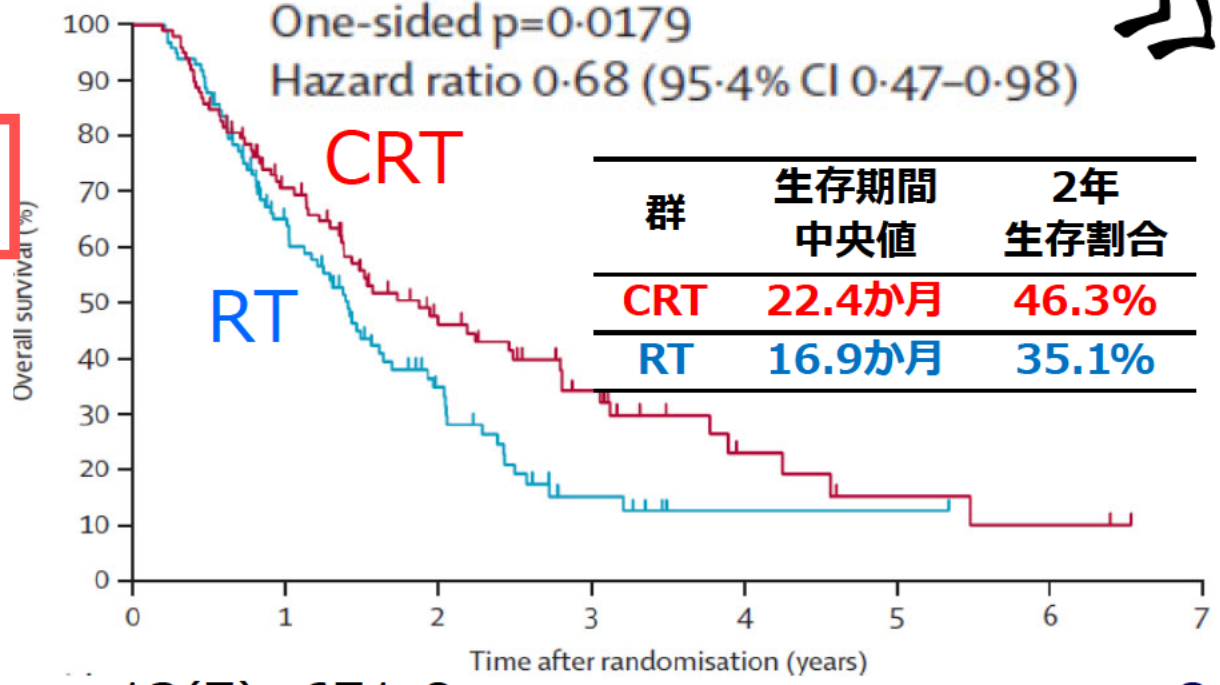
切除不能III期 非小細胞肺癌
71才以上

↓

ランダム割付

100例 → **放射線単独（RT）**（標準治療）

100例 → **放射線+化学療法（CRT）**（新治療）



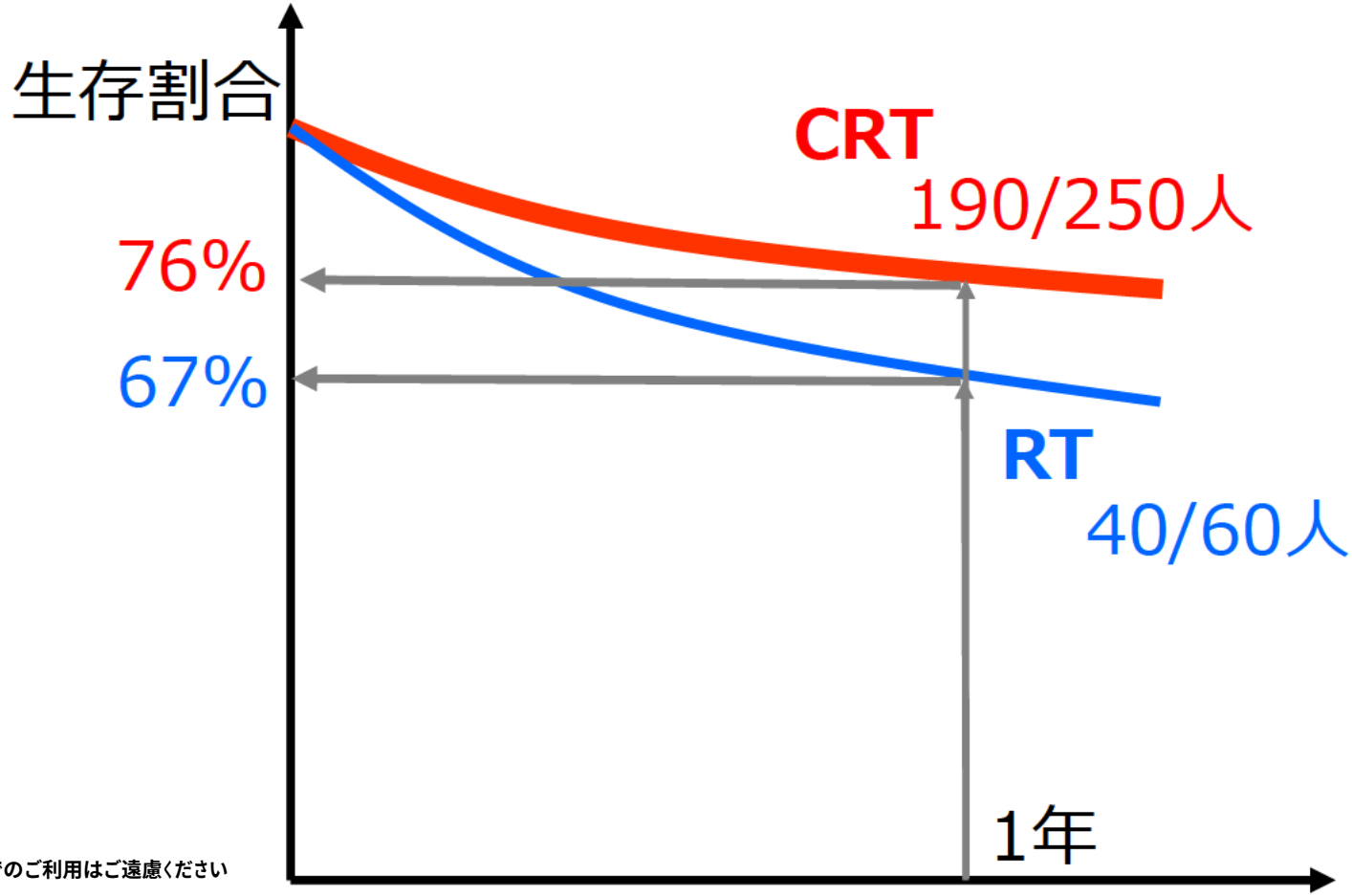
群	生存期間 中央値	2年 生存割合
CRT	22.4か月	46.3%
RT	16.9か月	35.1%

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

注：仮想例

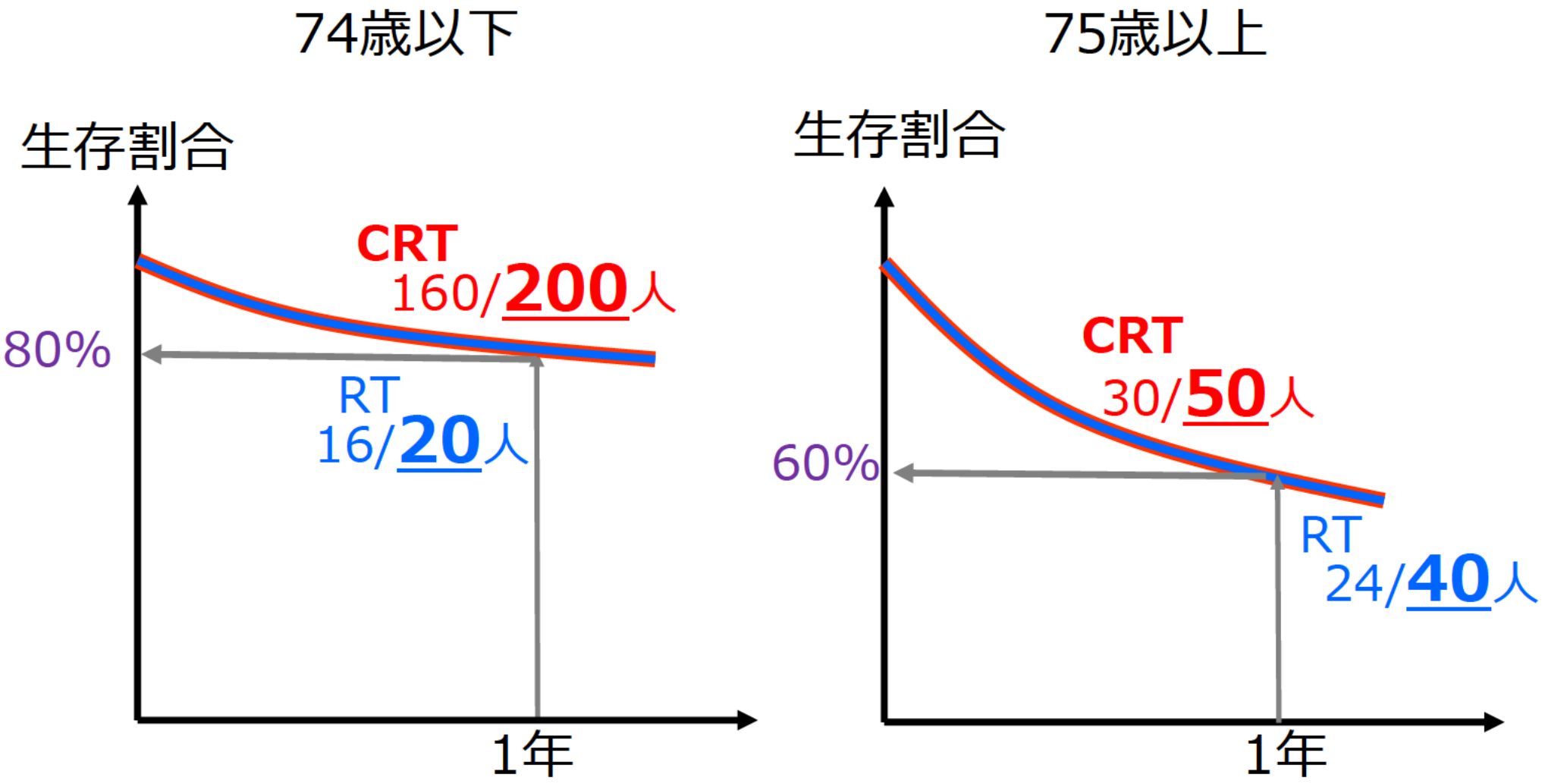
学会で見かける発表

- 当院における適格規準XXを満たす患者を**CRT群**(250例)と**RT群**(60例)に分けてレトロスペクティブに検討した。
- **CRT群**は**RT群**と比較し予後良好であった。
- この対象に**CRT**をすることが推奨される。



注：仮想例

仮に年齢で分けた場合の予後



年齢で分けるとCRTとRTの予後は変わらない

問題① 正しい解釈は？

どれか一つを選択してください。

① 全体の結果が正しい（CRTとRTの効果に差がある）

② 年齢別の結果が正しい（CRTとRTの効果に差がない）

③ 全体と年齢別の結果がどちらも正しい
（全体では効果に差があるが、年齢別では差がない）

④ 全体と年齢別の結果がどちらも誤り

注：仮想例

比較したいのはCRTとRTの違いだから

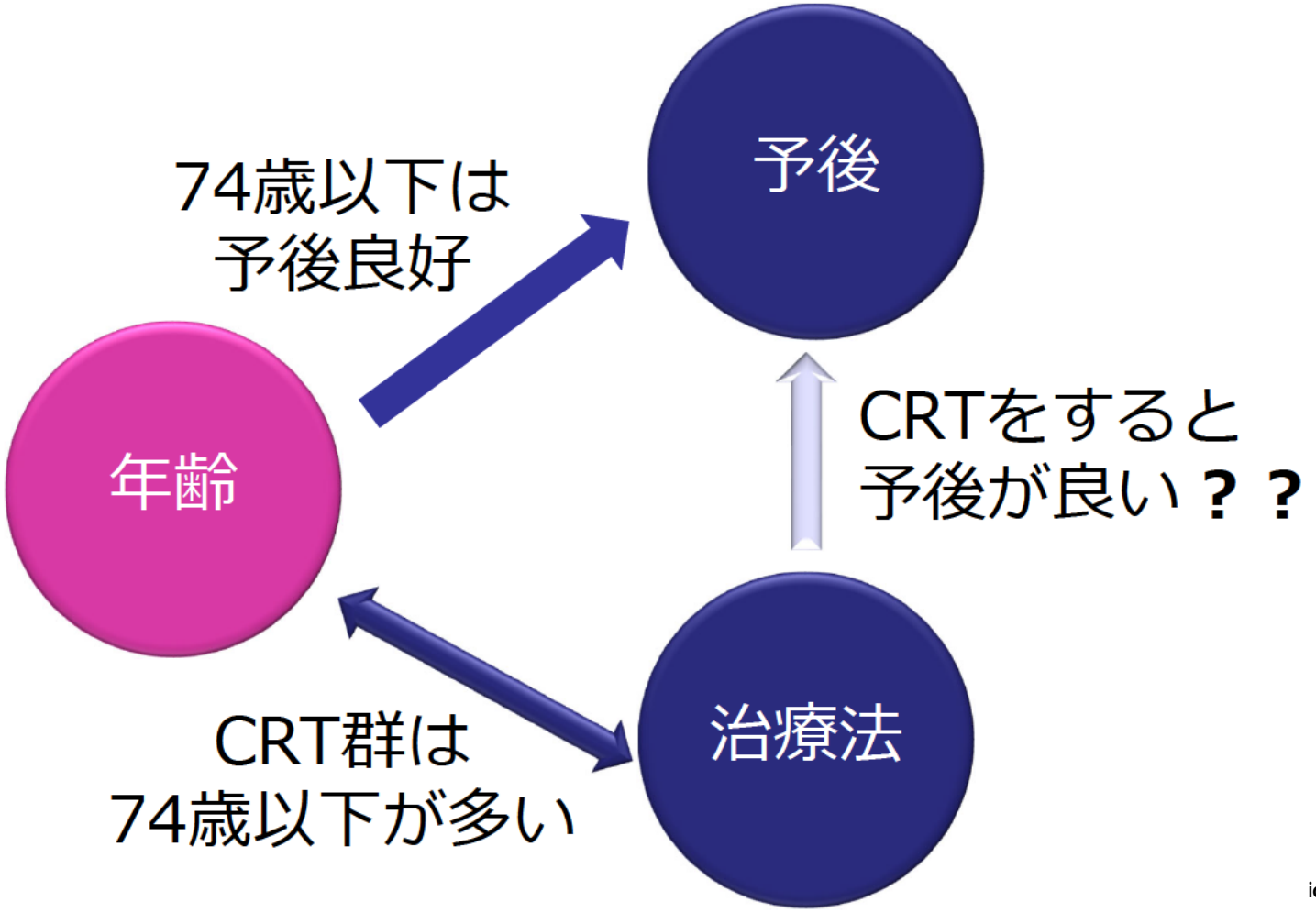
- 治療法以外の予後に影響する要因の条件が同じでなければ、”比較”にならない！！

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

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

交絡についてのまとめ

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



正しい解釈は？（交絡がないのは？）

① 全体の結果が正しい（CRTとRTの効果に差がある）

- 年齢による交絡があるため誤り

② 年齢別の結果が正しい（CRTとRTの効果に差がない）

- 交絡因子（年齢）が群間で揃っているので正しい

③ 全体と年齢別の結果がどちらも正しい
（全体では効果に差があるが年齢別では差がない）

④ 全体と年齢別の結果がどちらも誤り

交絡がないことを保証するには

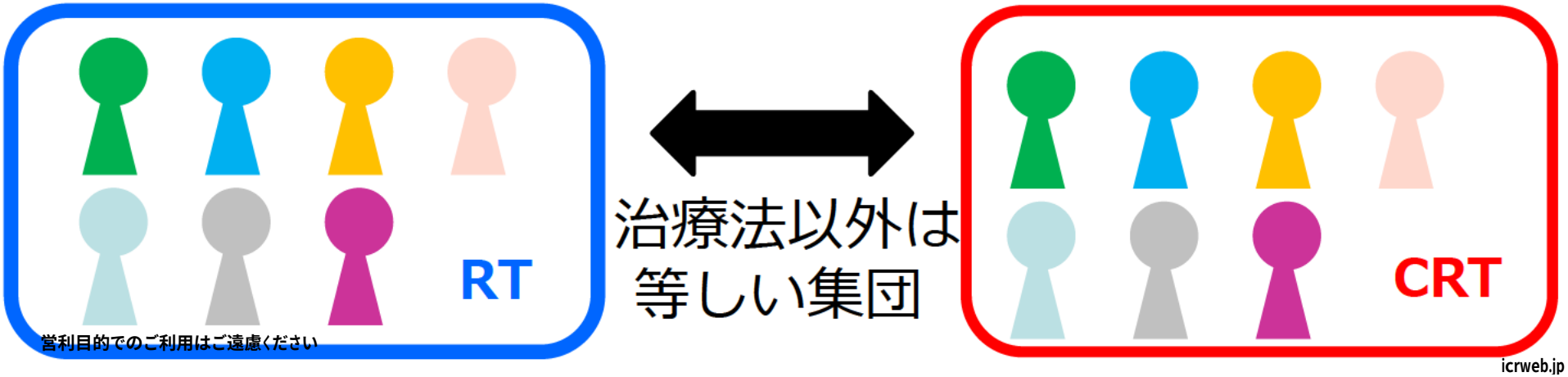
- 治療群間で予後に関係する背景因子を揃える
  - 年齢
  - Stage（がんの進行度）
  - Performance Status（全身状態）
  - その他（未知の因子も含めて）

因子がたくさんある・未知の因子があるために
全てを考慮できない

治療法を**ランダム**に決める

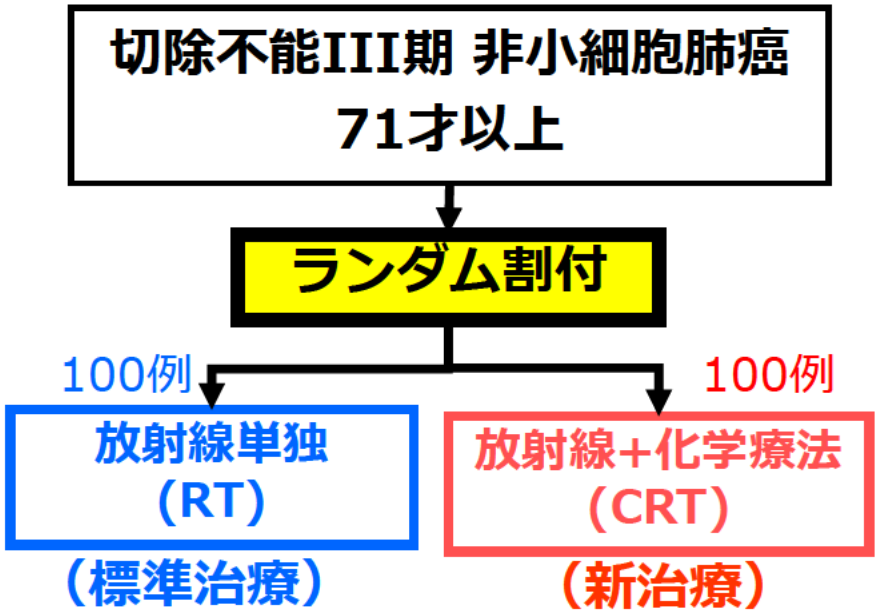
ランダム化 randomization

- 医師あるいは患者の意思によらず、確率に基づいて各治療群に患者を割り付ける
- 予見による患者選択の偏りの防止
  - 状態の良い患者はCRTに割り付けられやすくなる、などを防ぐ
- **比較可能性（内的妥当性）が担保される**
  - 治療法以外は等しい集団 → 効果に差があれば治療法の違い

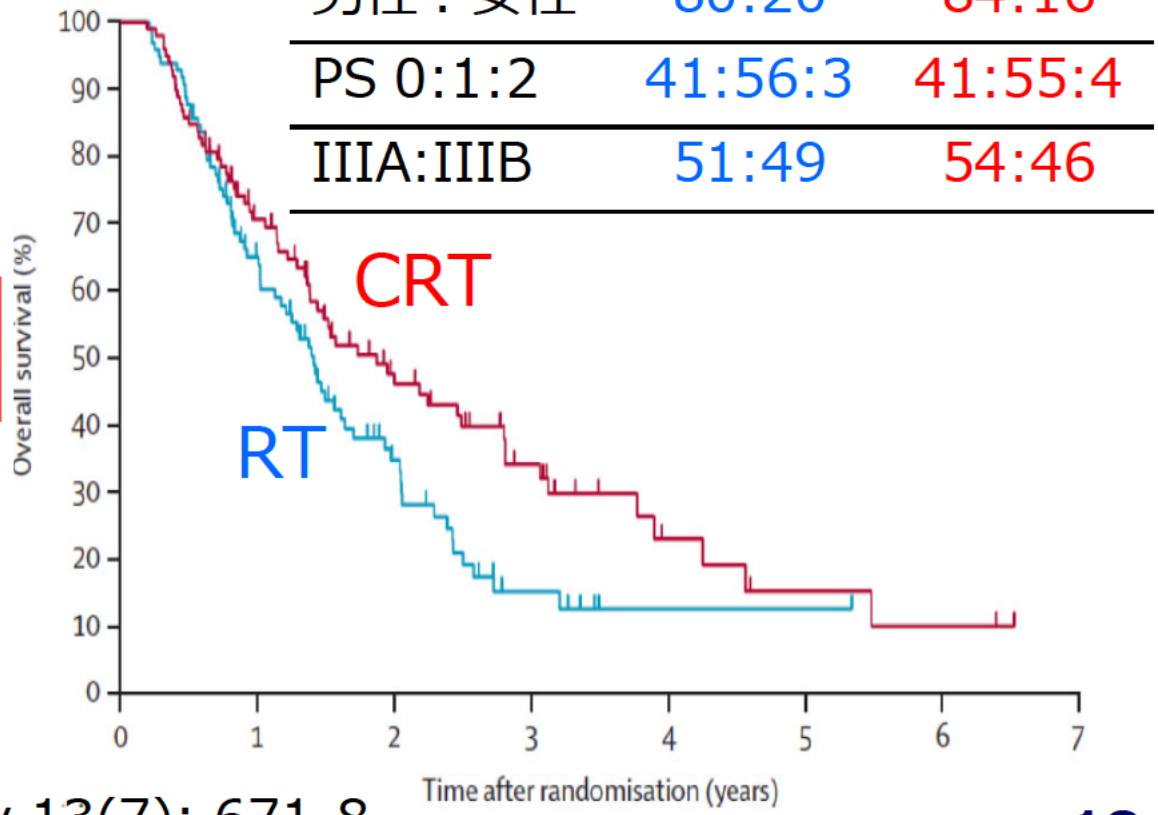


JCOG0301の場合

- **RTとCRTを比較するためにランダムに割り付けた**
  - 治療法以外の背景因子は平均的に治療群間で同じ
  - 生存曲線の違いは治療法による違いであると期待できる



	RT	CRT
年齢中央値	77歳	77歳
男性：女性	80:20	84:16
PS 0:1:2	41:56:3	41:55:4
IIIA:IIIB	51:49	54:46



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

仮説検定

CRT群は勝ったの？

肺がん内科グループ
JCOG0301

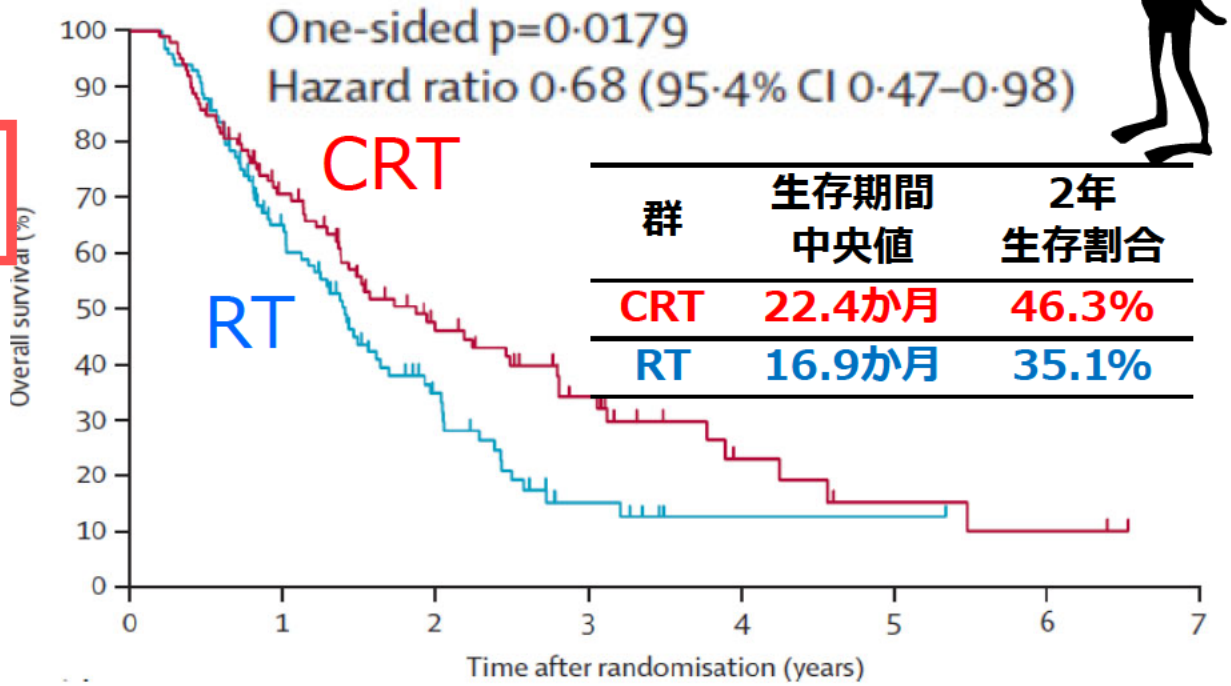
切除不能III期 非小細胞肺癌
71才以上

ランダム割付

100例
放射線単独
（RT）
（標準治療）

100例
放射線+化学療法
（CRT）
（新治療）

ランダム化しているから比較可能性があることはわかった。確かに、CRT群の生存曲線がRT群よりも上にあるけど、ランダム化して生存曲線が上にあればCRT群が勝ったと言って良いの？

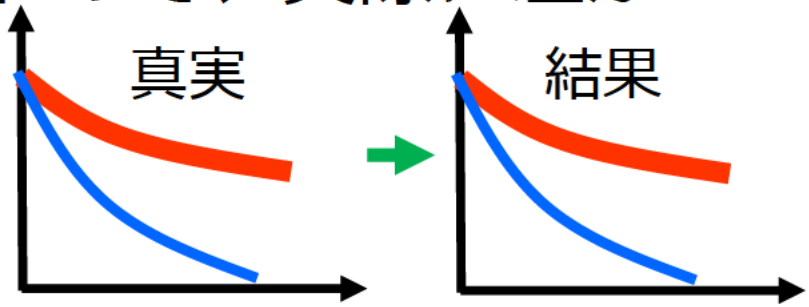


群	生存期間 中央値	2年 生存割合
CRT	22.4か月	46.3%
RT	16.9か月	35.1%

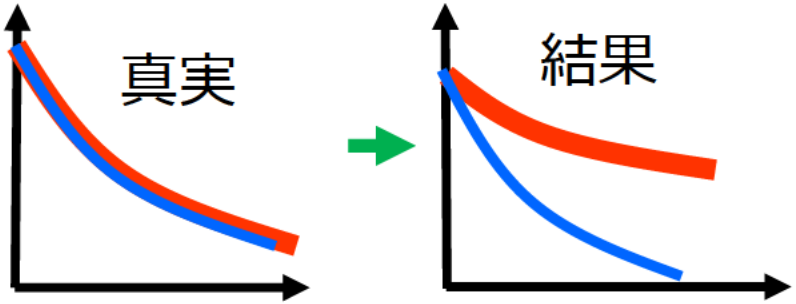
Atagi *et al.* (2012) Lancet Oncology 13(7): 671-8.

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「**RT**と**CRT**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「**RT**と**CRT**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かめたい！

確かめる方法：仮説検定

- 証明したいことは、「RTとCRTに差がある」 ですが、

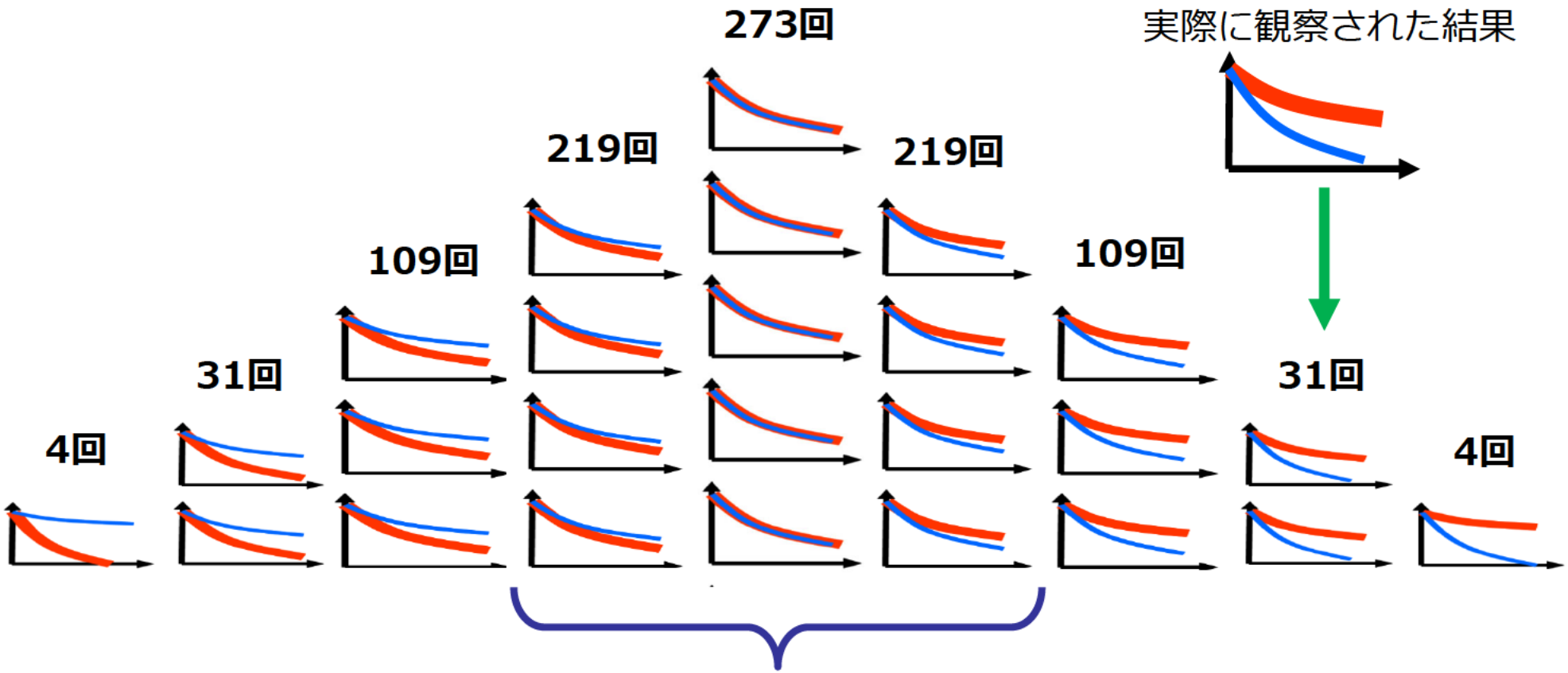
1. 「RTとCRTに差がない」という仮説を置く
   - この仮説を**帰無仮説**という
2. 「RTとCRTに差がない」という仮説の下で、何回も試験をした場合に得られる結果の分布を調べる
3. 「RTとCRTに差がない」という仮説の下で、実際に観察されたRTとCRTの差以上に大きな差になる確率を調べる
4. この確率が小さければ、そもそも「RTとCRTに差がない」という仮説（帰無仮説）が間違っていた、と判断する
5. 「RTとCRTに差がある」が正しいと判断する

RTとCRTの生存曲線に【差がない】下での結果の分布

もし、RTとCRTの生存曲線に【差がない】が真実なら…

日本全国の切除不能III期 非小細胞肺癌

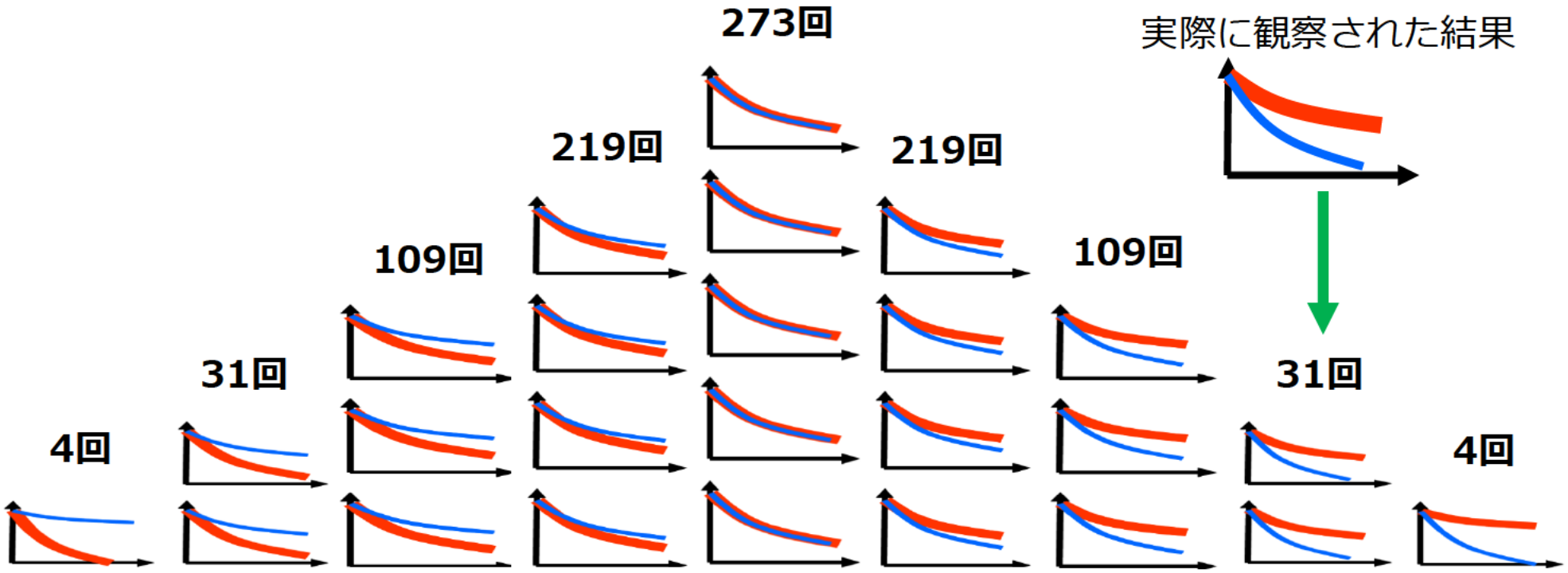
71才以上の患者から200人選んで1000回試験すると、、、



【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(**P**robability)は、35/1000 = **3.5%**
  - 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35回くらいしか起こらないような**稀な結果**（?）



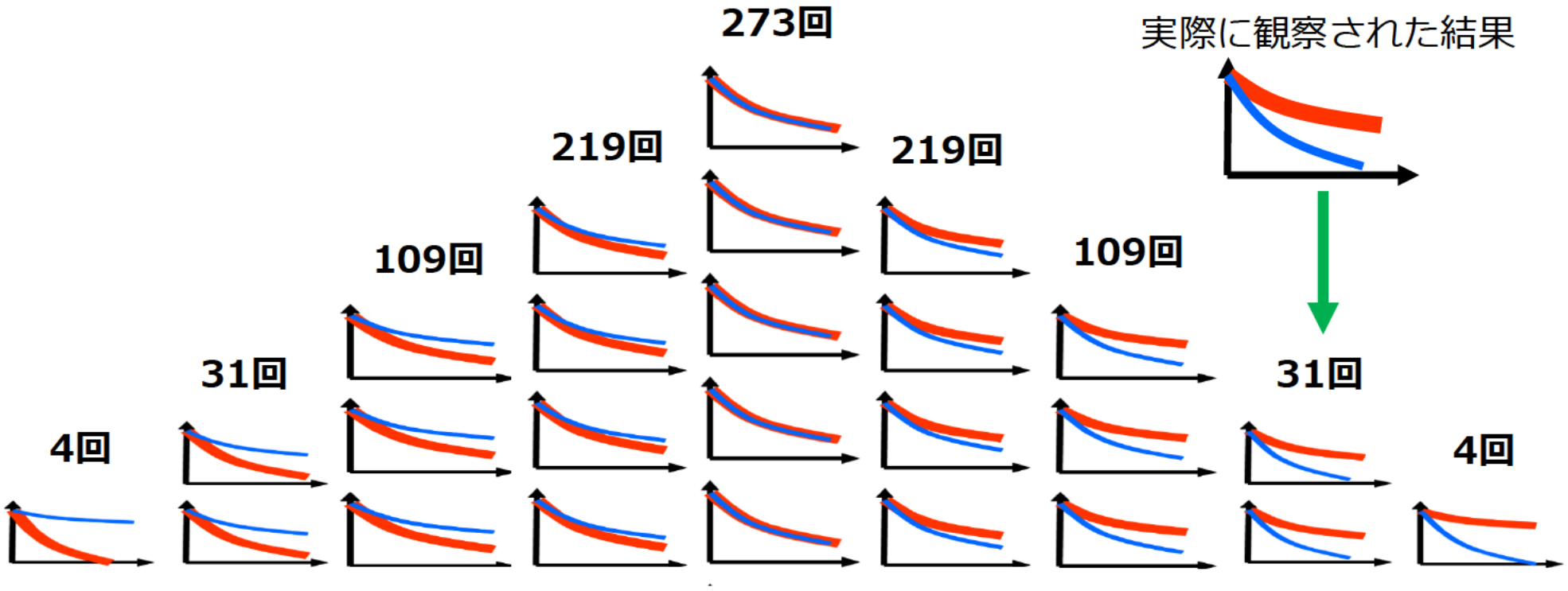
3.5%は稀な結果？

- 3.5%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、RTとCRTは**差があると結論する** ＝ **【有意差あり】**
- 3.5%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、RTとCRT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

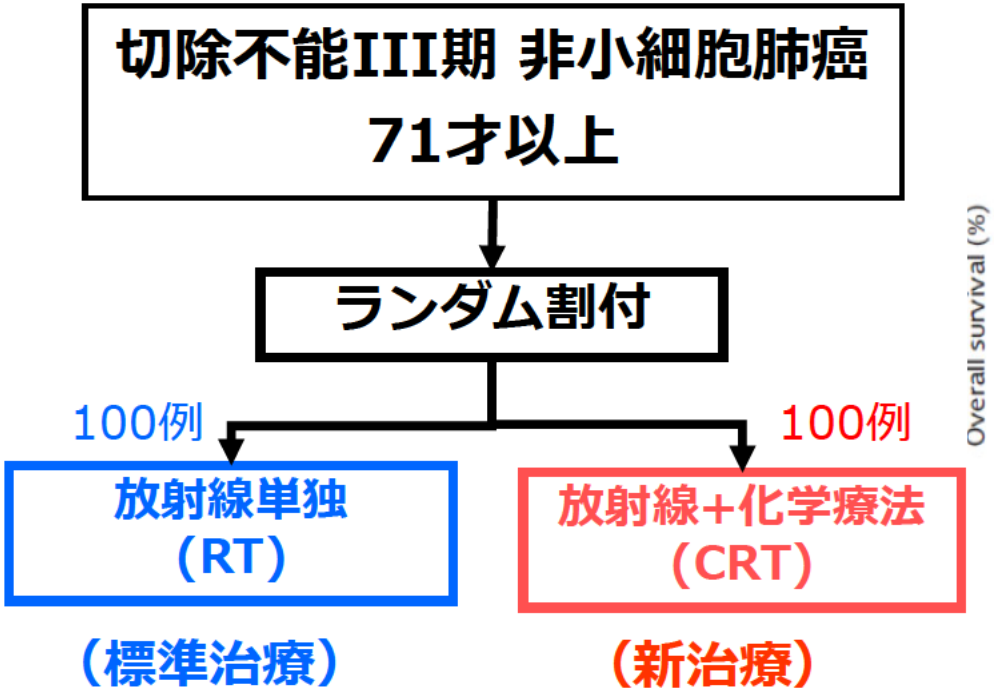
検定結果

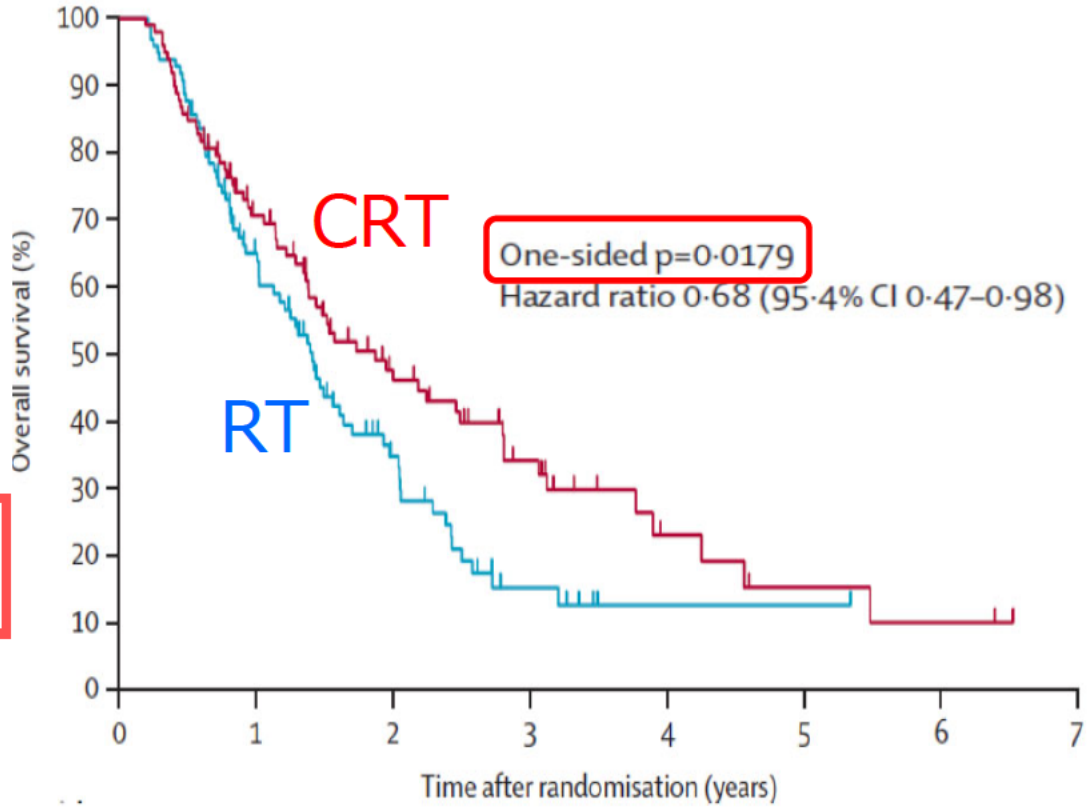
- P値=3.5%だった
  - 実際に観察された結果は【差がない】が真実だとしたら、1000回中35回くらいしか起こらないような結果
- 有意水準を**5%**に設定していたとしたら、有意差あり
- 有意水準を**2.5%**に設定していたとしたら、有意差なし



JCOG0301の場合

- **p=0.0179**：両群に差がないとしたら100回中1～2回くらいしか起こらない
  - 事前に決めた規準 p値≦5%を満たす（有意水準5%）
  - CRTはRTと比較して優れていると判断





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

αエラー、βエラー、検出力

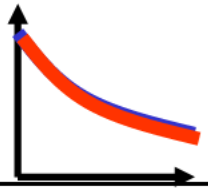
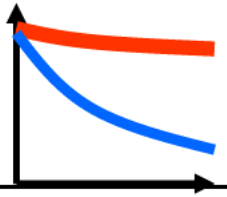
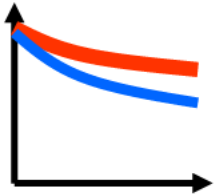
仮説検定の結果は絶対正しい？？

必ずしも検定の結果は正しいとは限らない

- 実際に得られた結果はP値=3.5%
  - これは稀にしか起こらないので、【差がない】という仮説は誤っていると判断した
  - 逆に言えば、【差がない】が真実の場合に稀には起こる
- 真実が【差がない】時に、誤って【差がある】と判断してしまうのは誤った判断をしていることになる
  - この誤りのことを**αエラー**という
  - 【差がない】時に【差がある】と判断する確率は有意水準以下なので、αエラーを起こす確率は有意水準以下

【差がある】のに有意差なしとしてしまう

- この誤りのことを、**"β エラー"** と呼ぶ
  - 本当は効果がある治療を真実に反して捨ててしまう誤り
- **検出力**（確率は**1-β**）
  - 「差がある」ものを正しく「差がある」と判断する確率

		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**）

治療効果の推定
p値ではわからないこと

CRTはどのくらい良い治療？

肺がん内科グループ
JCOG0301

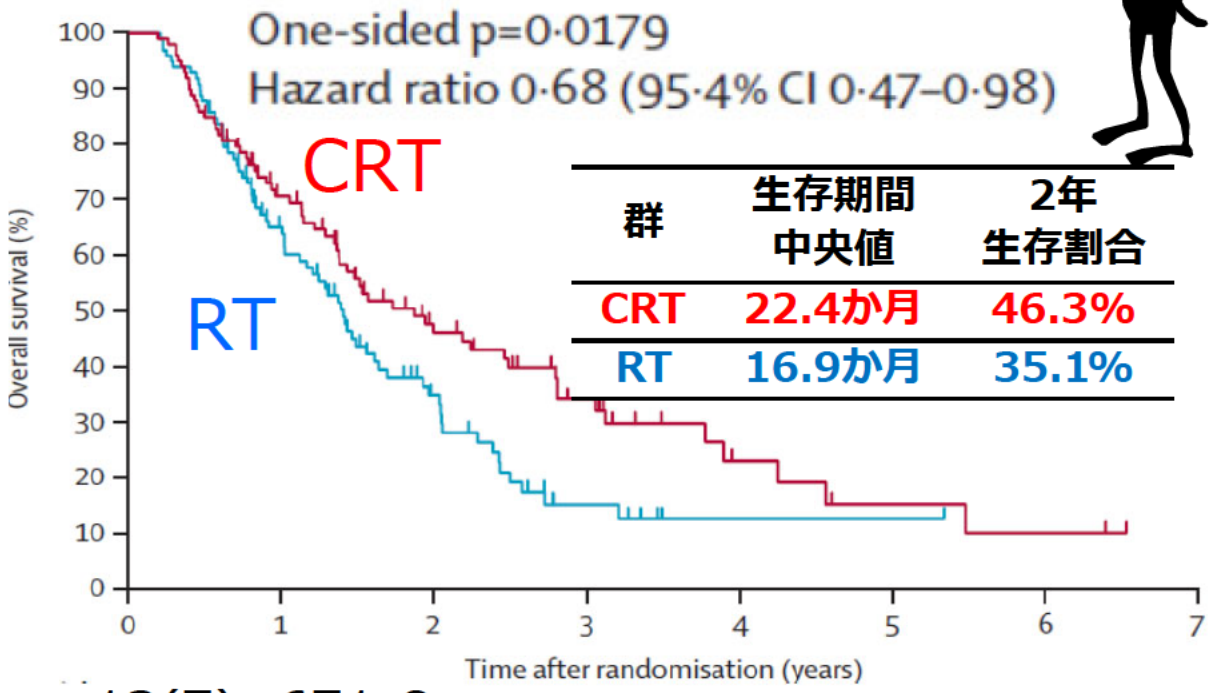
切除不能III期 非小細胞肺癌
71才以上

ランダム割付

100例 **放射線単独（RT）** **（標準治療）**

100例 **放射線+化学療法（CRT）** **（新治療）**

CRT群がRT群よりも良いことはわかった。
でも、どのくらい良い治療法なの？
P値が小さければ良い治療？



群	生存期間 中央値	2年 生存割合
CRT	22.4か月	46.3%
RT	16.9か月	35.1%

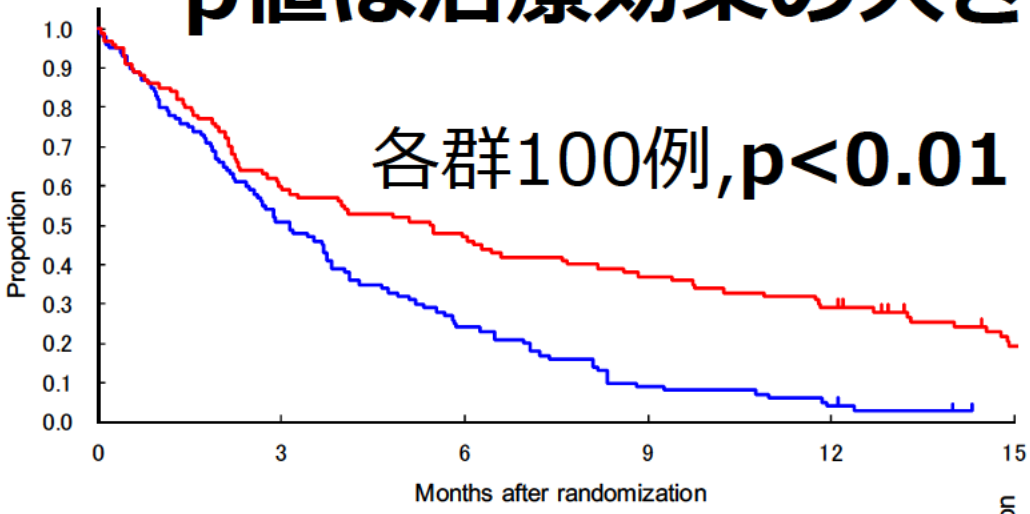
Atagi *et al.* (2012) Lancet Oncology 13(7): 671-8.

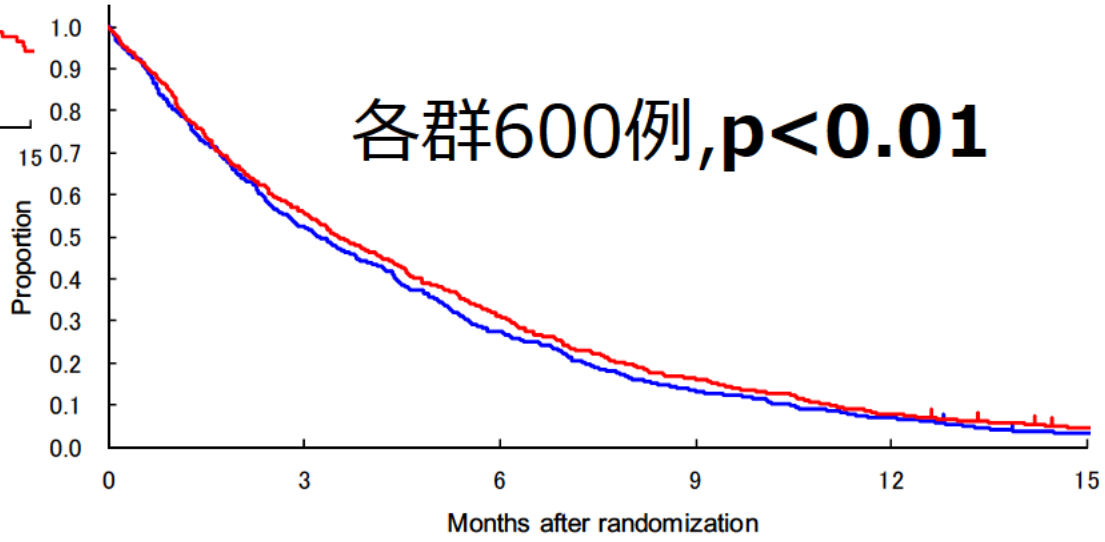
注：仮想例

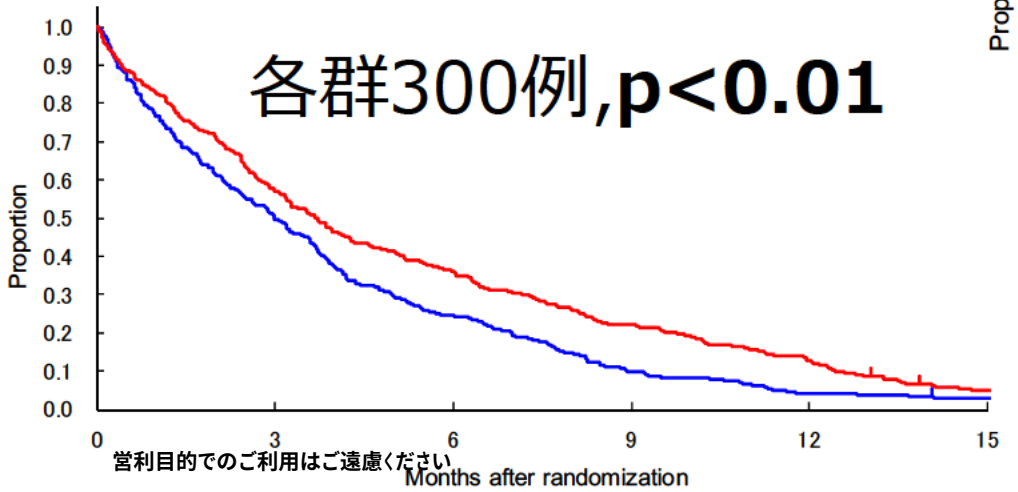
統計的有意差 ≠ 臨床的有意差

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





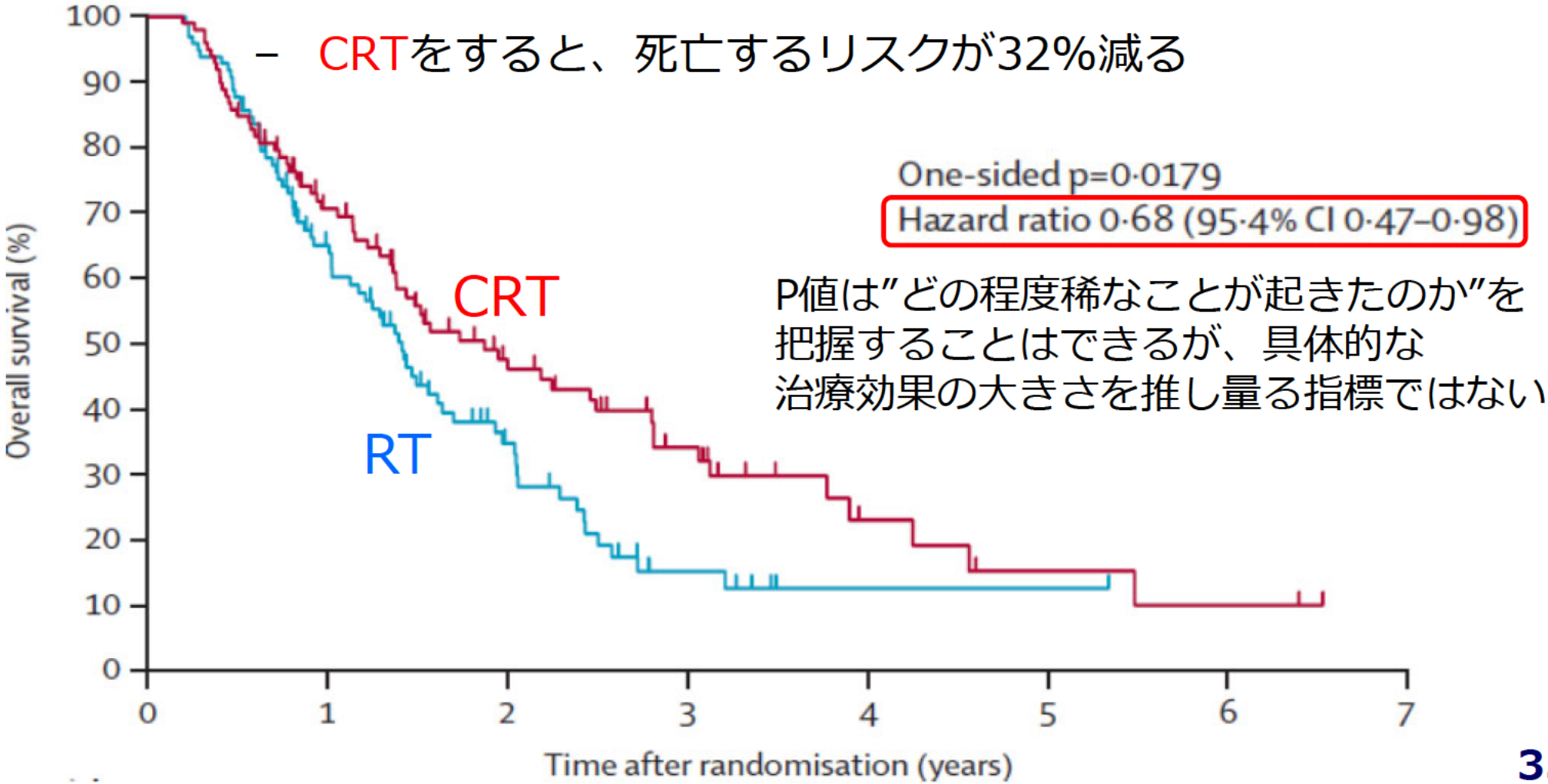


治療効果の大きさを表す指標

- 曲線のある1時点に着目した指標
  - 年次生存割合の差
    - 2年生存割合 CRT:46.3% vs RT:35.1%
  - 生存期間中央値（MST）の差
    - CRT: 22.4か月 vs RT:16.9か月
- 曲線全体を一つの効果にまとめた指標
  - **ハザード比**(**HR** : **H**azard **R**atio)
    - 群間の**ハザード（瞬間死亡率）**の比をとったもの

JCOG0301における解釈

- RT群に対するCRT群の**ハザード比(HR)が0.68**
  - CRTをすると、死亡するリスクが0.68倍になる
  - CRTをすると、死亡するリスクが32%減る



P値は”どの程度稀なことが起きたのか”を把握することはできるが、具体的な治療効果の大きさを推し量る指標ではない

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

Intent(ion)-To-Treat

ちゃんと治療していない患者の扱いは？

肺がん内科グループ
JCOG0301

切除不能III期 非小細胞肺癌
71才以上

ランダム割付

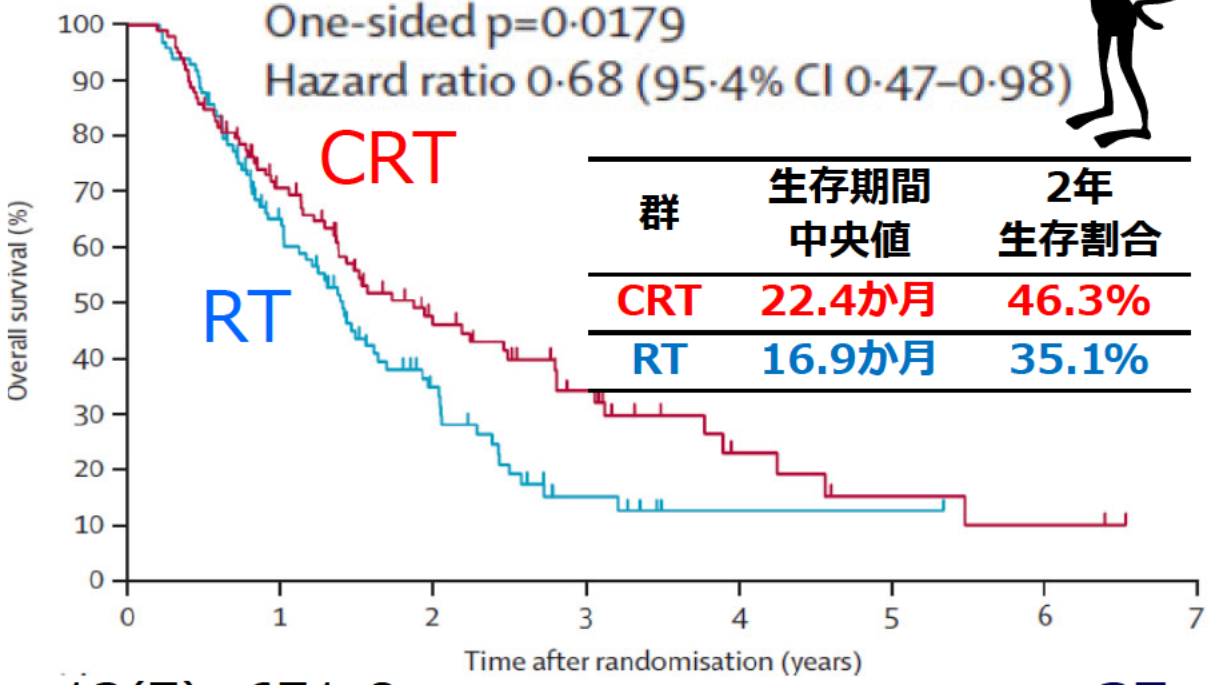
100例 **放射線単独 (RT)**
- **RT施行例 98例**
- **RT完遂 93例**
- **RT非完遂 5例**
- **未治療例 2例**

（標準治療）

100例 **放射線+化学療法 (CRT)**
- **CRT施行例 96例**
- **CRT完遂 88例**
- **CRT非完遂 8例**
- **未治療例 4例**

（新治療）

RT群にもCRT群にも、ちゃんと治療を出来てない人がいるぞ！？
この患者は生存曲線の群間比較の解析に含まれているのかな？

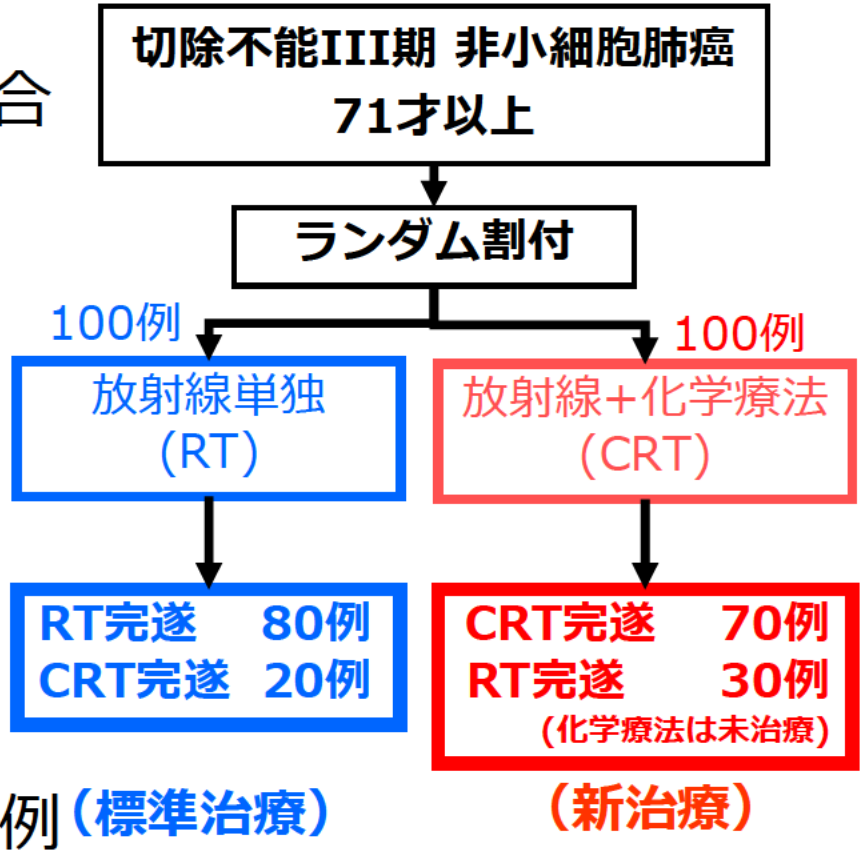


群	生存期間 中央値	2年 生存割合
CRT	22.4か月	46.3%
RT	16.9か月	35.1%

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

問題② あなたならどの解析法を選びますか？

- 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？
- どれか一つを選択してください



① 治療完遂例どうしで比較

RT 80例 vs. CRT 70例

② 実際に行った治療どうしで比較

RT (80+30)例 vs. CRT (70+20)例

③ ランダム化で割り付けられたどうしで比較

RT 100例 vs. CRT 100例

どんな結果が予測されるか考えると

αエラーup

差がないのに誤って差があると言ってしまう確率αup

① **治療完遂例どうしで比較**　RT 80例 vs. CRT 70例

- RT例は、CRTができた元気な人を除いた集団
- CRT例は、RTしかできなかった元気でない人を除いた集団

② **実際に行った治療どうしで比較** RT (80+30)例 vs. CRT (70+20)例

- RT例は、RTしかできなかった元気でない人を含む集団
- CRT例は、CRTができた元気な人を含む集団

⇒　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

検出力down

差がない時は正しく差がないと言える

③ **ランダム化で割り付けられたどうしで比較** RT 100例 vs. CRT 100例

- 本当にCRTに効果がある場合、CRT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う

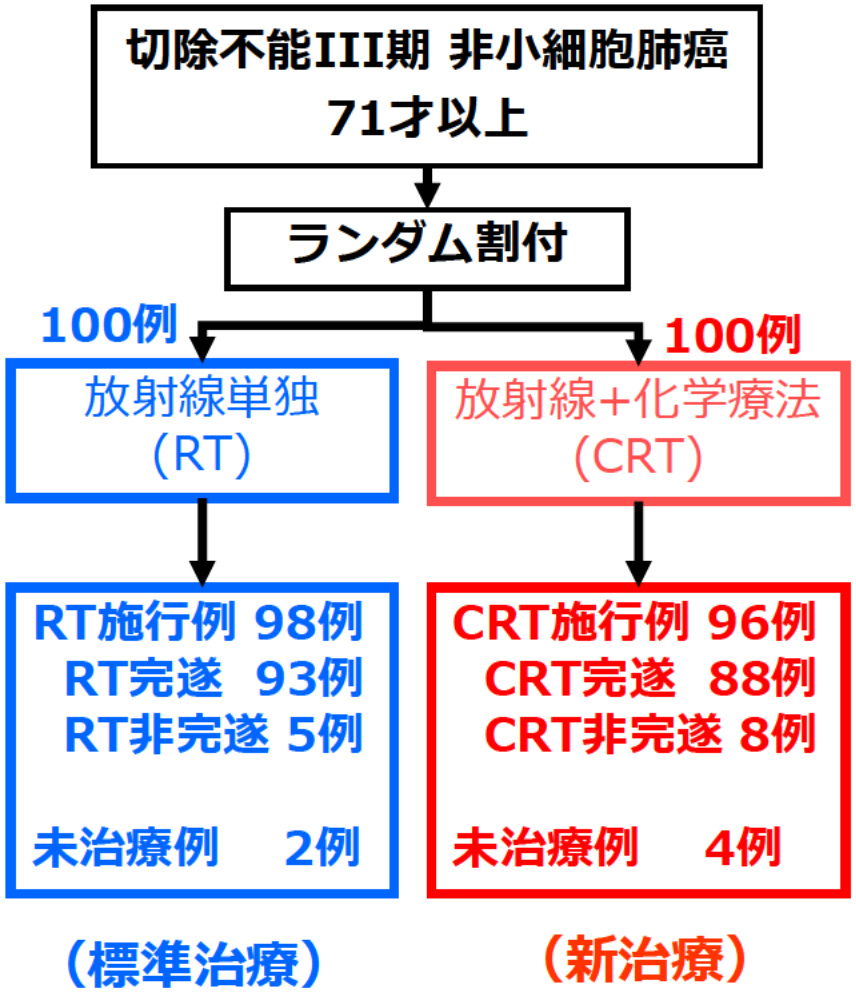
Intention-to treat解析(ITT解析)

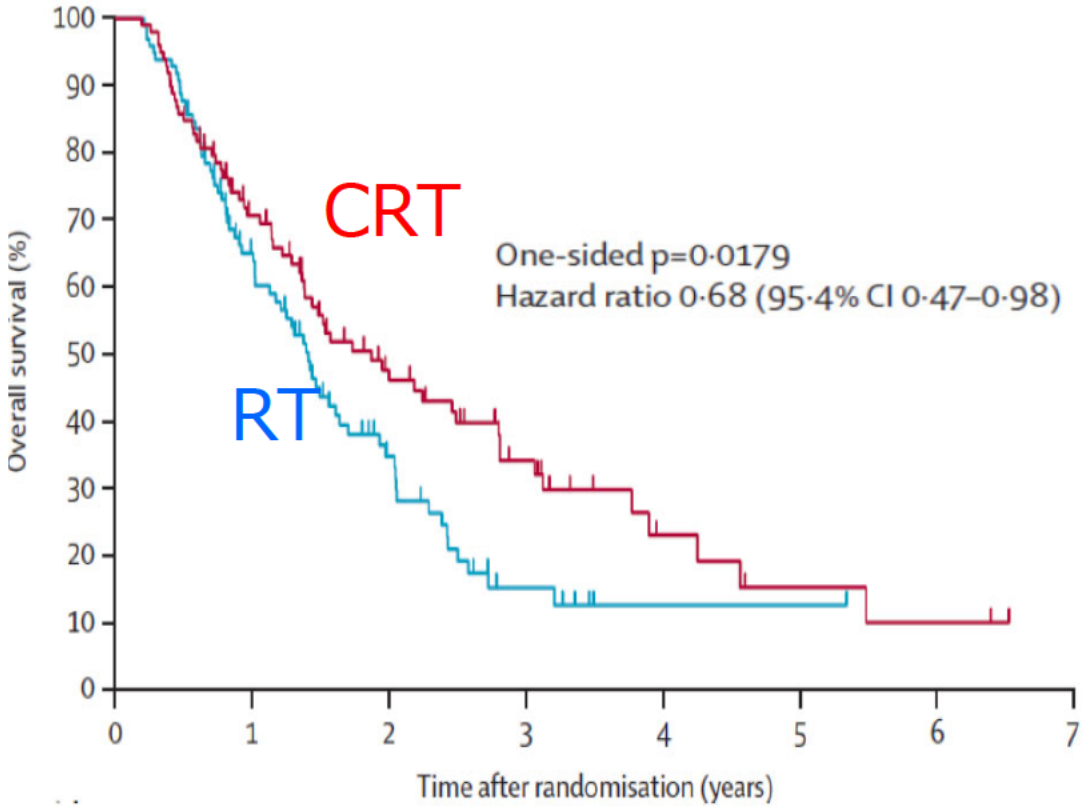
- ランダム化によって**割り付けられた通り**の治療群で行う解析（③の対象で解析する方法）のこと
  - **ITT解析**をすれば、αエラーを起こす確率は大きくならない
  - 有意になりにくいという意味で「保守的(conservative)」な解析法

- 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ります！」と言える
  - **ITT解析がランダム化比較試験の主たる解析の標準的方法**

JCOG0301の場合

- 主たる解析は治療非完遂例などを含めた**ITT解析**
  - CRTはRTを上回っていると判断できる





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

まとめ

- **生存曲線**は生存割合を時間に対してプロット。結果を視覚的に判断できる
- **ランダム化**によって交絡を除去し、治療法の適切な比較が可能
- 結果は仮説**検定**によって求めた**p値**が**有意水準（α）**を下回ったら差があると判断
- 治療効果の大きさはp値ではなく、**ハザード比**や生存割合で判断
- 臨床試験の主たる解析は割り付けられた通りの治療群で行う**（ITT解析）**