

がん臨床試験のデザイン

JCOG運営事務局 SC部門

門田 智裕

第21回JCOG臨床試験セミナー

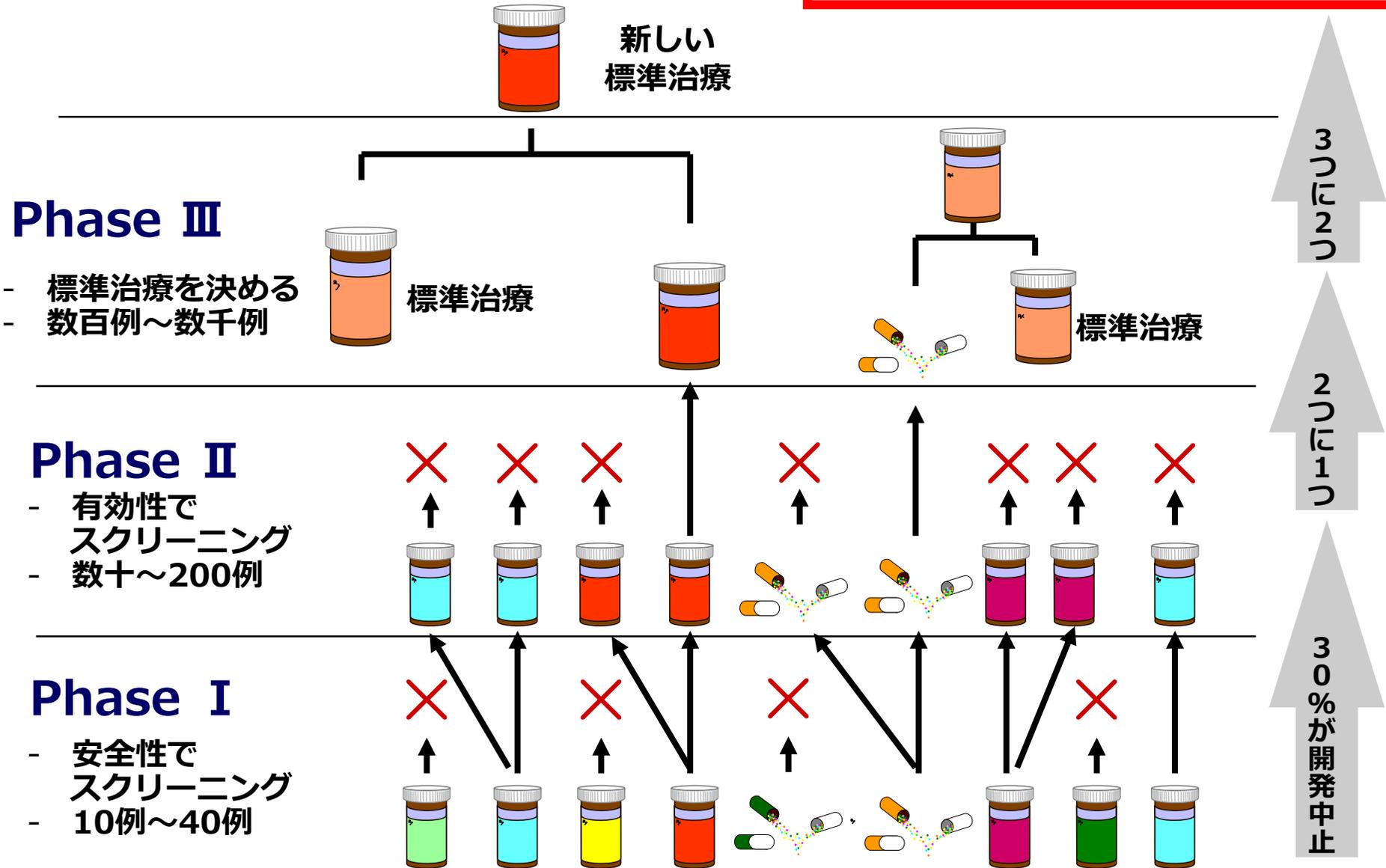
2018/10/13

Outline

- エンドポイントの種類
- Phase I のデザイン
- Phase II のデザイン
- Phase III のデザイン

がん臨床試験の流れ

成功確率は5分の1(1999-2004年)
無秩序にやると非効率的・非倫理的
「正しい」評価が結局は早道!



臨床試験の原則

臨床試験は“比較”である

Phase III

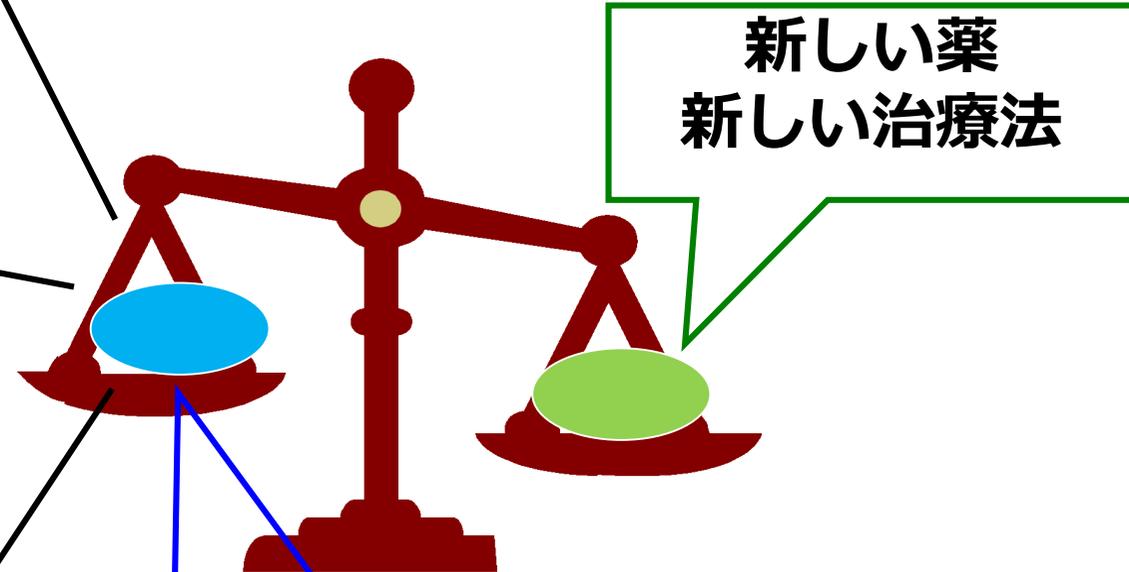
標準薬・標準治療と
総合的な有効性と安全性を
直接同時比較

Phase II

過去のデータから導かれる
有効性の水準や目標と比較

Phase I

過去のデータから導かれる
安全性の許容水準と比較
(許容される毒性の割合)



比較する相手は常に
標準薬・標準治療

何を比較するのか

- **患者さんの利益 (benefit) を比較する**

- 今までの標準薬と比べて

- よく効く
- 副作用が少ない
- 安い
- ● **有効性**の比較
- ● **安全性**の比較
- ● **経済性**の比較

※ただしコストベネフィットは国際的なコンセンサスがなく、社会制度によって大きく異なるので 一般論は成り立たない

Benefitをどのように測るか？

- エンドポイント (endpoint、評価指標、評価項目)

「患者さんの benefit を測る**“ものさし”**」

“ criterion by which patient benefit is measured”

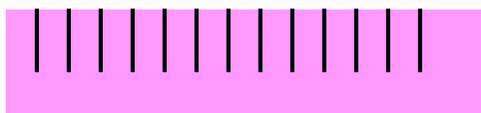
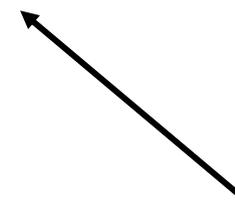
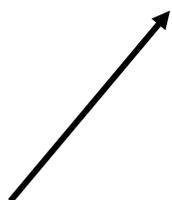
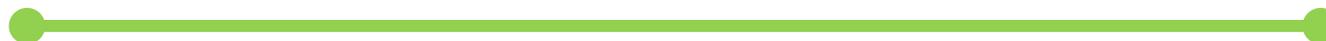
- Richard Simon



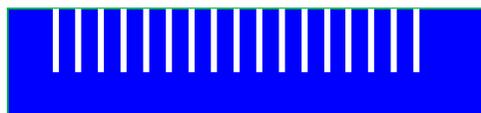
エンドポイント（ものさし）

benefit 小

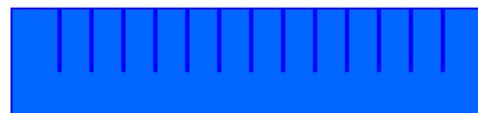
benefit 大



有害事象発生割合



全生存期間
(OS)



奏効割合
(腫瘍縮小効果)



QOLスコア

Primary endpoint : 研究の結論を下すために最も重要なエンドポイント

Secondary endpoint : その他のエンドポイント

エンドポイントの分類

(患者のベネフィットを直接反映する) 真のエンドポイント

代替(サロゲート)エンドポイント

ハードなエンドポイント

(誰が見ても・何度見ても同じで、他の要因の影響を受けにくい)

全生存期間(OS)
(死亡するまでの期間)

誰が見ても同じ
予後延長はベネフィット

無再発生存期間(RFS)
(手術後に再発or死亡するまでの期間)

新病変の判定はぶれにくい
無治療の期間延長は患者のベネフィット

腫瘍縮小の判定はぶれやすい
腫瘍縮小はベネフィットと言い切れない

奏効割合
(腫瘍縮小効果)

腫瘍増悪の判定はぶれやすい
患者の直接のベネフィットになりにくい

無増悪生存期間(PFS)
(増悪or死亡するまでの期間)

QOLスコア

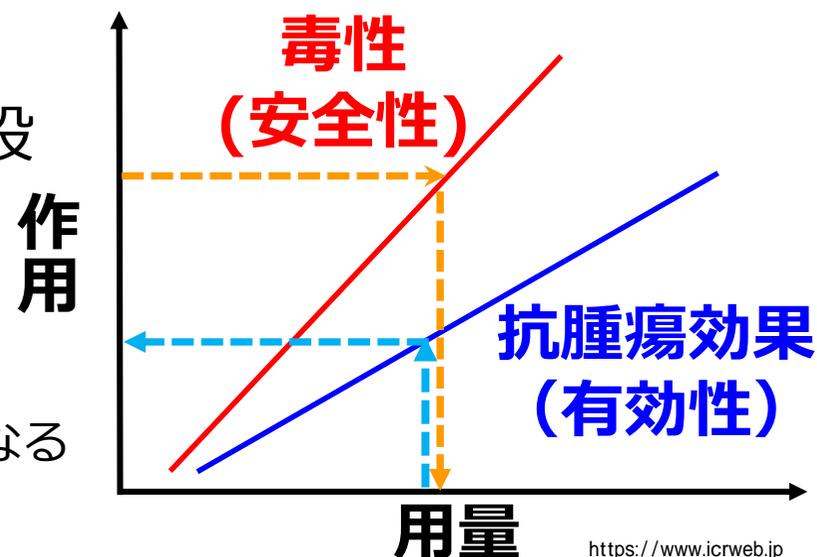
同じものさしで測れない
もし正確に測れば、ベネフィットを反映

ソフトなエンドポイント

Phase I のデザイン

Phase I 試験の目的と概要

- 目的：Phase IIに進むかどうかを決めること
 - Phase II以降に用いる試験治療の推奨用量の決定
- 典型的なデザイン
 - 対象：既に標準治療を施行した正常臓器機能の進行期の患者
 - 登録数：10-40例
 - 参加施設：単～数施設の専門施設
 - **毒性**によるスクリーニング
 - 前提：**毒性**が**有効性**のサロゲート
 - 毒性が高くなると、有効性も高くなる

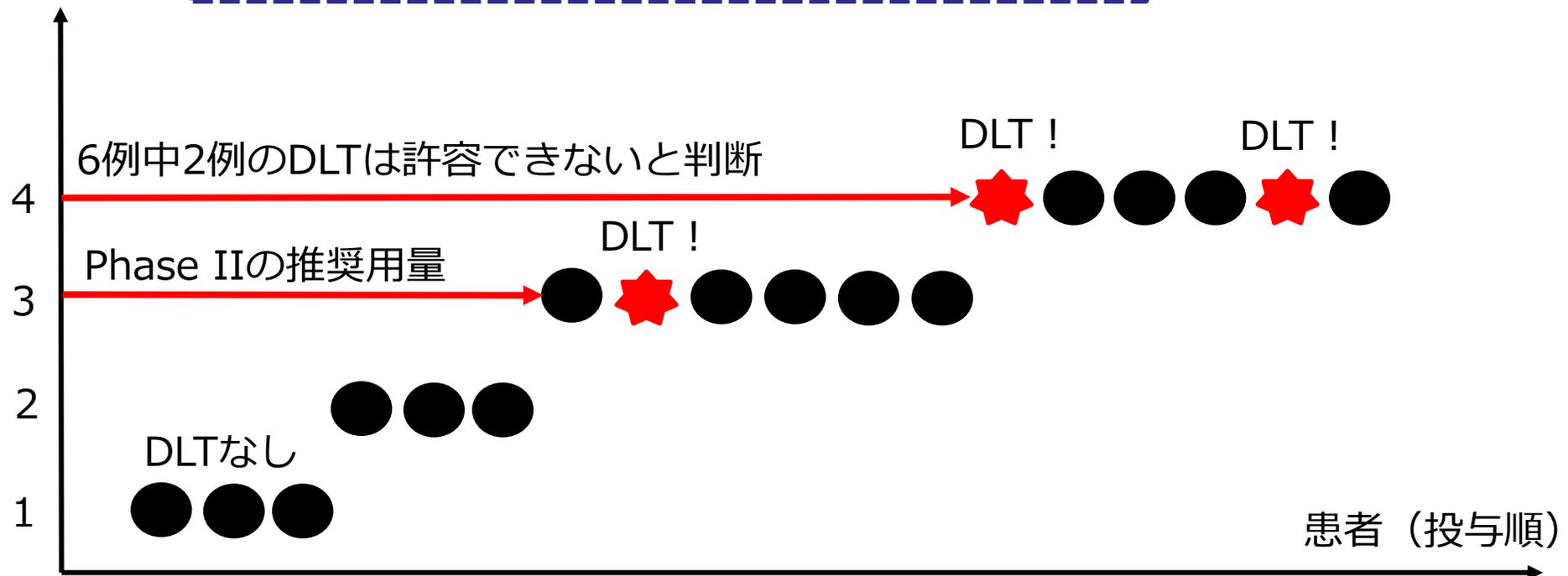


伝統的なphase I デザイン：3例コホート

- ① ある用量レベルで3例投与し、3例中
 - ①-1 DLT=0例なら、次用量レベル。①へ
 - ①-2 DLT=1例なら、3例追加。②へ
 - ①-3 DLT \geq 2例なら、終了
- ② 3例追加して、6例中、
 - ②-1 DLT=1例なら、次用量レベル。①へ
 - ②-2 DLT \geq 2例なら、終了

用量レベル

DLT：用量制限毒性



3例コホートの欠点

- 低用量で投与される患者さんが多い！
 - Phase I 段階とはいえ、登録患者が期待しているのは有効性
 - 低用量(≡ 有効性が低い)の投与例は出来るだけ少なくしたい

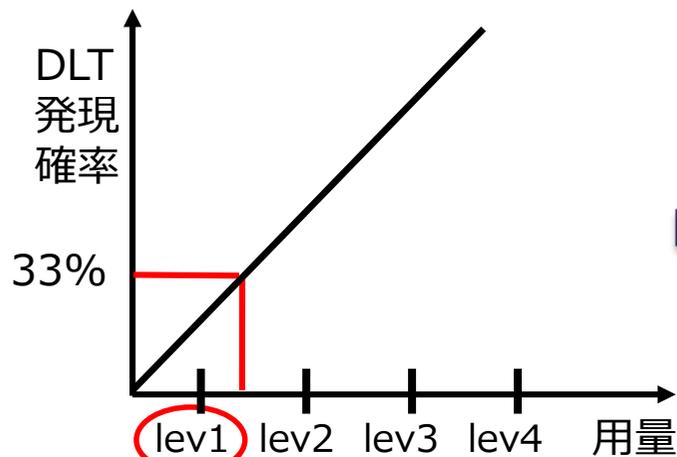
各薬剤の最初に人に投与したphase I 試験の患者数と用量レベル数

薬剤名	FDA承認年	患者数	用量レベル数
パクリタキセル	1992	34	11
ゲムシタビン	1996	47	12
イマチニブ	2001	83	14
ペメトレキセト	2004	38	10
パニツムマブ	2006	96	13

CRM : Continual Reassessment Method

- 用量とDLT発現確率の関数を置く (数学的モデルを設定)
- 登録された患者のDLT有無の情報を使って数学的モデルを更新
- 次に登録される患者の投与レベルは、数学的モデルを元に
DLT=33%を超えない最大のレベルを投与

6例投与された時点の
用量と毒性の関係
(数学的モデルで決まる)



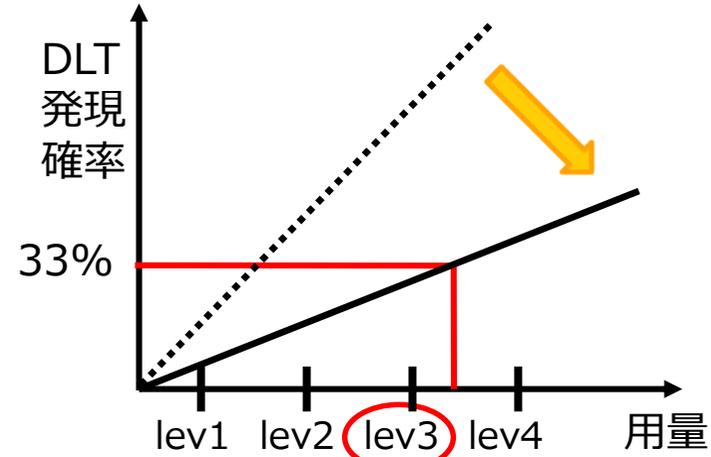
7例目の投与レベル

営利目的での使用はご遠慮ください

7例目はDLTなし



7例投与された時点の
用量と毒性の関係
(数学的モデルで用量と毒性の関係を更新)



8例目の投与レベル

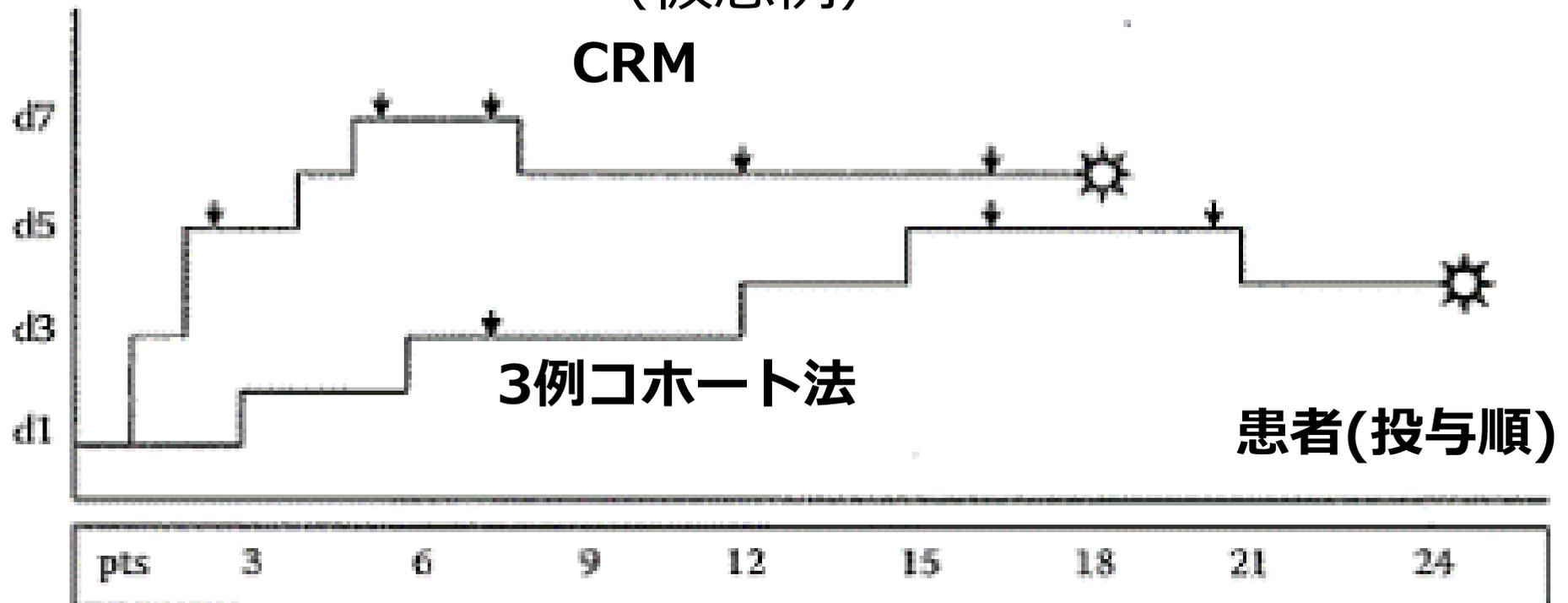
<https://www.icrweb.jp>

CRMのメリット

- 3例コホート法よりも正しい推奨用量を選択できる確率が高い
- 低用量で投与される患者数が少なくて済む可能性が高くなる
- 最適な用量付近で投与される患者数が多くなる
- ただし、煩雑さの問題もありCRMの使用は今のところ限定的

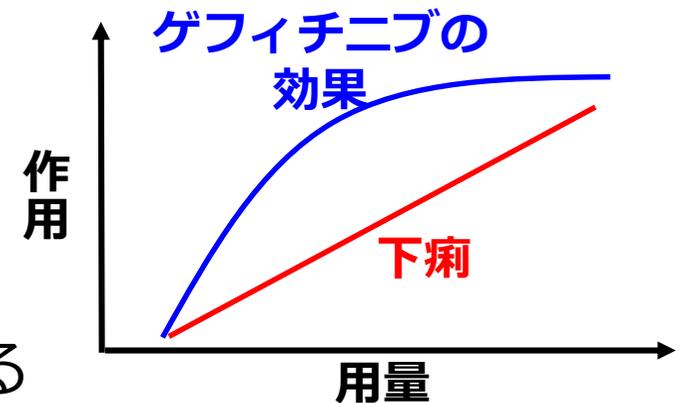
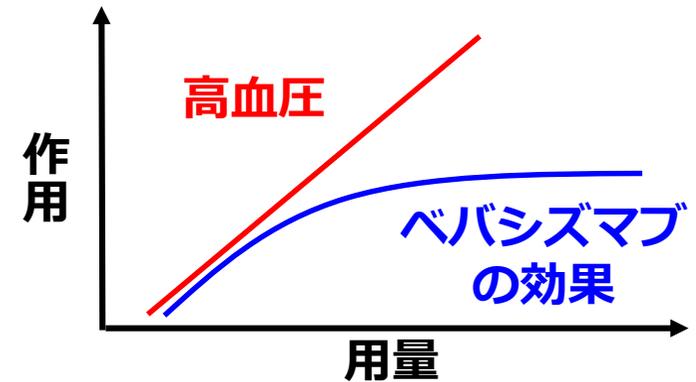
用量レベル

(仮想例)



分子標的薬などの場合のエンドポイント

- 毒性が有効性のサロゲートとして適切ではないかもしれない
 - 例1：ベバシズマブの高血圧
 - 毒性は用量依存だが、有効性は依存しない
 - 例2：ゲフィチニブの下痢、皮疹
 - 毒性が発生する前に、効果がプラトーに
- 毒性以外で推奨用量を決定することもある
 - 血液中の薬物量の指標（AUC、Cmaxなど）
 - 血液中の標的分子に期待される効果の発現
 - 画像の効果判定による腫瘍縮小効果



拡大コホート (Expansion cohort)

- 推奨用量を中心に数例～数十例追加し、安全性・有効性を確認する方法

1. 推奨用量は3-6例の投与のため、推奨用量での安全性を再検討

- パゾパニブ：43例の検討で800 mg 1日1回となったが、400 mg 1日2回投与なども含めて20例追加し安全性を再検討

Hurwitz HI, et al. Clinical cancer research 2009;15(12):4220-7.

- 拡大コホートを用いた13%の試験で元々の推奨用量を減量

Manji A, et al. Journal of clinical oncology. 2013;31(33):4260-7.

2. がん種やバイオマーカーを特定して安全性・有効性を確認したい

- ペンブロリズマブ：固形がん全体で開始したが、腫瘍縮小効果の見られた悪性黒色腫と非小細胞肺癌で拡大コホートを追加

Prowell TM, et al. The New England journal of medicine. 2016;374(21):2001-3.

Phase I 試験デザインのまとめ

- Phase I 試験は主に安全性による候補薬のスクリーニング
 - 殺細胞薬の97%、分子標的薬の58%は毒性による推奨用量の決定

Jardim DL, et al. Clinical cancer research. 2014;20(2):281-8.

- 統計的観点からはCRMの方が3例コホートより好ましい
 - 低用量で投与される患者数の減少
 - 最適な用量付近で投与される患者数の増加

- 煩雑さの問題もありCRMの使用は今のところ限定的？

- 07-08年の181試験のうち6試験(3%) Le Tourneau C, et al. J Natl Cancer Inst. 2009;101(10):708-20.
- 11-12年の162試験のうち7試験(4%) Riviere MK, et al. Annals of oncology. 2015;26(4):669-74.

- 効率的な治療開発を目的として拡大コホートの利用が増えている

- 06-11年の611試験のうち149試験(24%) Manji A, et al. JCO. 2013;31(33):4260-7.

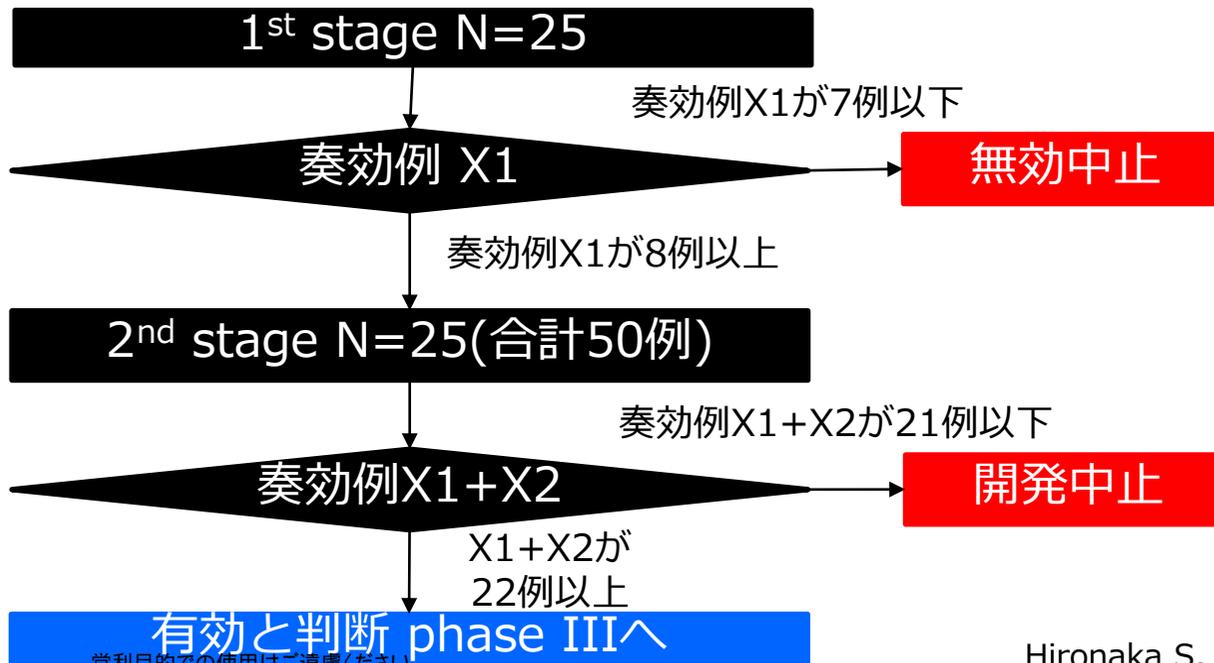
Phase II のデザイン

Phase II 試験の目的と概要

- 目的：Phase III に進むかどうかを決めること
 - 標準治療に勝てそうな有効性があるかどうか？
 - Phase I で観察されなかった毒性プロファイルの充実
 - 用量・用法・治療変更規準などの最適化
- 典型的なデザイン
 - 対象：がん種を限定
 - 登録数：20-60例
 - 参加施設：phase I よりも多くの施設
 - **単群**試験でRECISTによる**奏効割合**などの短期的なエンドポイント
 - できるだけ早くphase III に進むかどうかを決めるため
 - **有効性**によるスクリーニング
 - 前提：腫瘍が縮小することが予後（OS）を延長することのサロゲート

閾値と期待値を用いた単群デザイン

- 閾値：これくらいしか効かなければ、開発をあきらめる値
 - 期待値：これくらい効いたら、開発を続けたい値
 - 奏効割合がprimary endpoint
 - 途中で試験の無効中止を考慮する（2段階のデザイン）
- 例：JCOG0807（切除不能・再発食道がんに対するDCF療法）
 - 閾値35%(CF療法の奏効割合)、期待値50%、 $\alpha=10\%$ 、検出力=80%



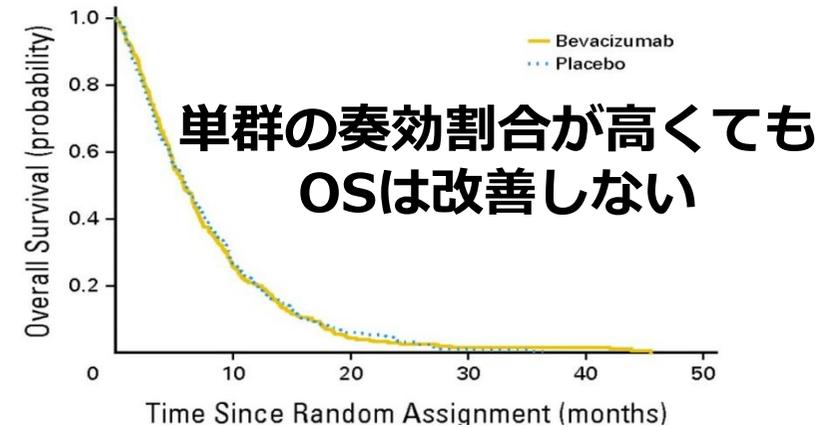
実際の結果：33/53=62%
↓
第III相試験
JCOG1314 (CF vs DCF)
実施中
(UMIN000015107)

単群でよい？奏効割合でよい？

● 膵臓癌

- 単群phase II試験の結果
- 奏効割合
 - ベバシズマブ群：21%
 - 当時の標準治療である
ゲムシタビン単剤：10%未満

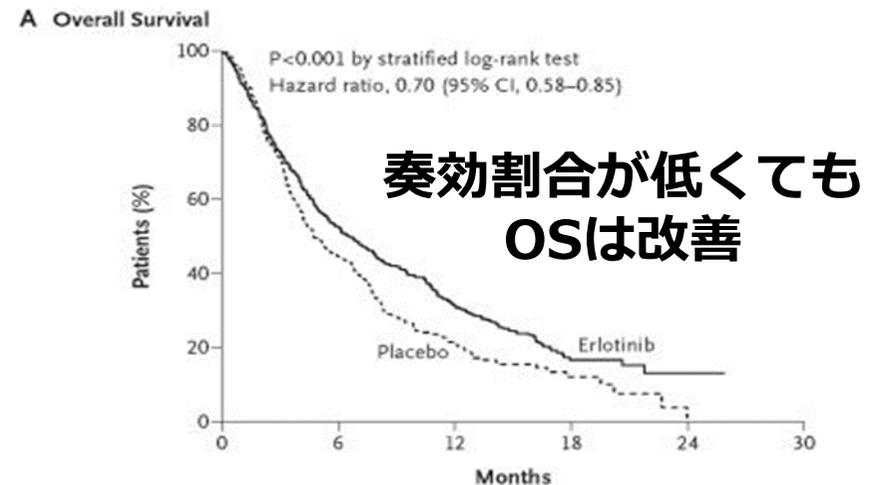
Phase IIIへ



Kindler, H. L., et al. *JCO* 2010; 28: 3617-3622

● 既治療IIIB/IV期 非小細胞肺癌

- ランダム化比較phase III試験の結果
- 奏効割合
 - プラセボ群：0.7%
 - エルロチニブ群：8.2%



Shepherd FA et al. *N Engl J Med* 2005;353:123-132.

<https://www.tcrweb.jp>

単群の年次生存割合、年次無増悪生存割合は？

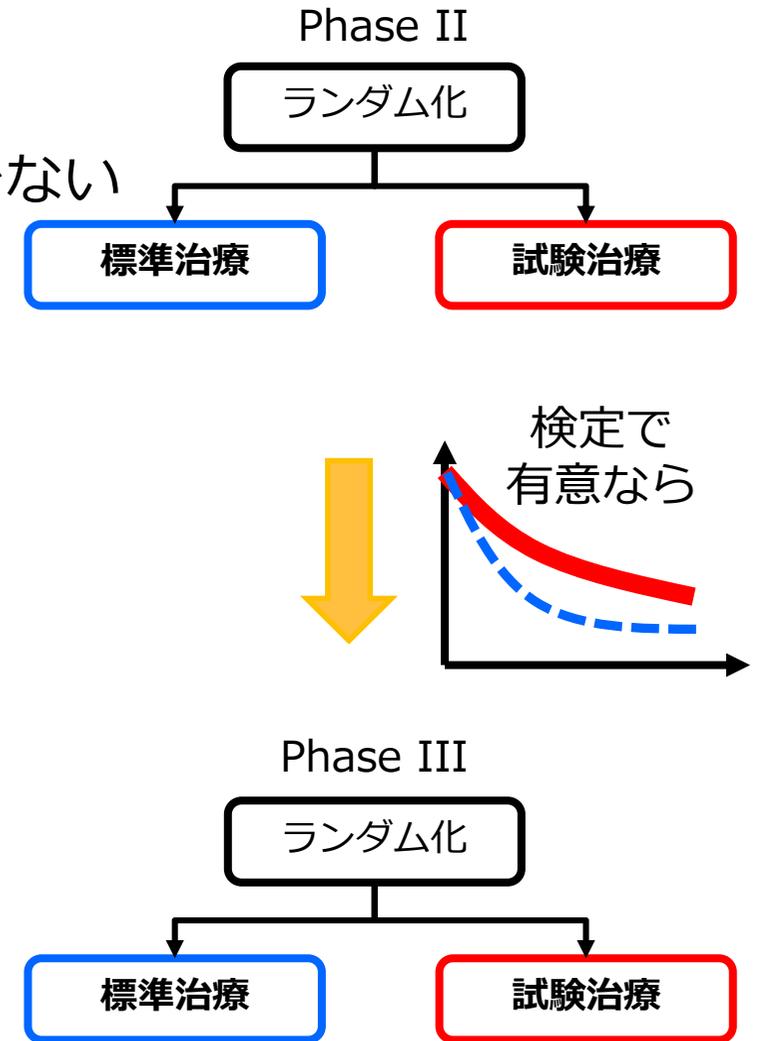
- 年次生存割合（例：1年生存割合）
 - 登録される患者の全身状態によりバラツキが生じやすい
 - 後治療の影響を受けやすい

- 年次無増悪生存割合（例：1年無増悪生存割合）
 - 腫瘍縮小効果は検査間隔の影響を受けやすい
 - 疾患によっては、増悪の定義が試験やグループでバラバラ
 - 前立腺癌では、PSA値や疼痛の増悪の定義が不明確

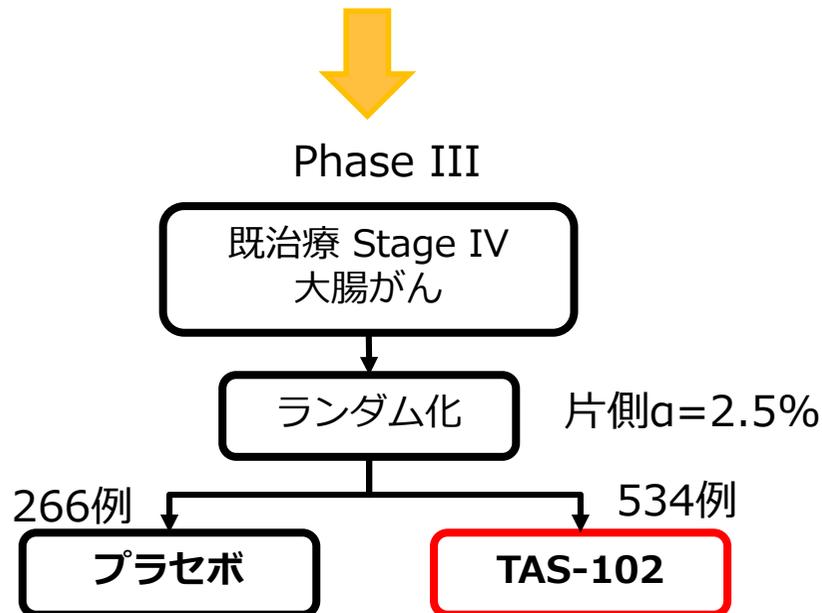
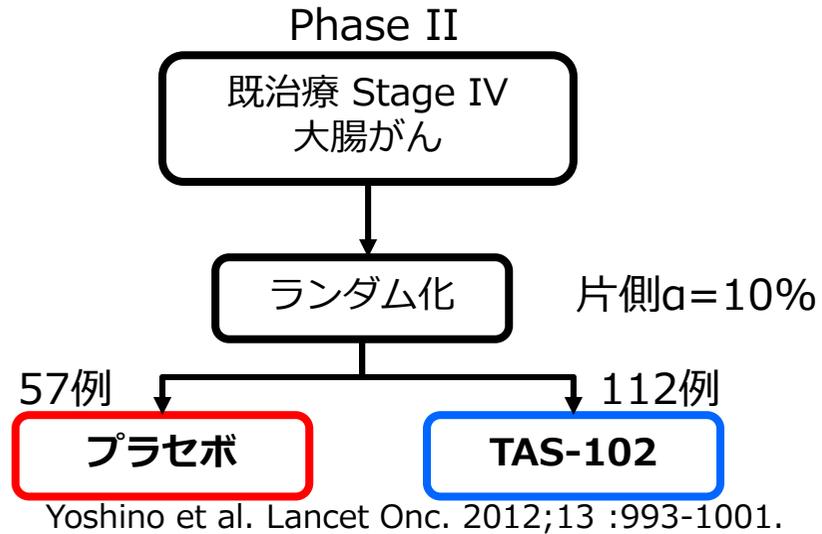
→ **ヒストリカルコントロールとの比較が困難**

ランダム化スクリーニングデザイン

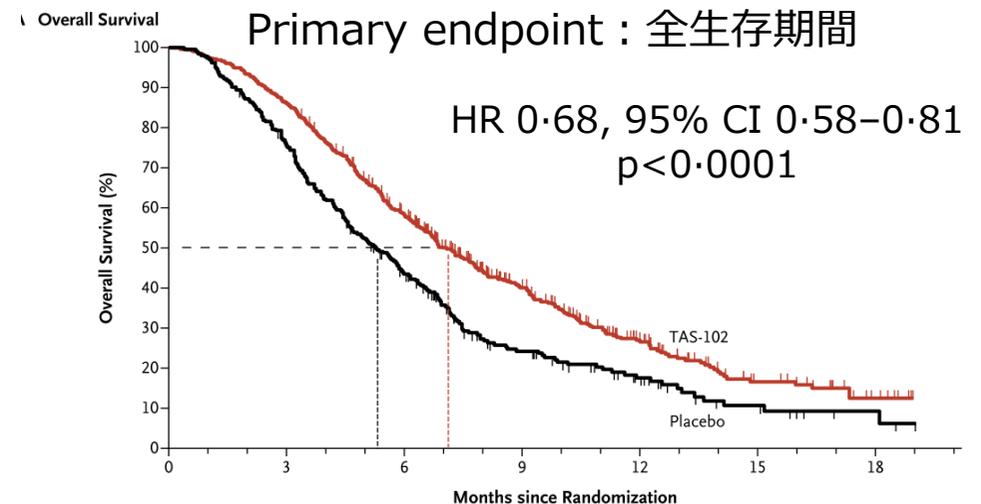
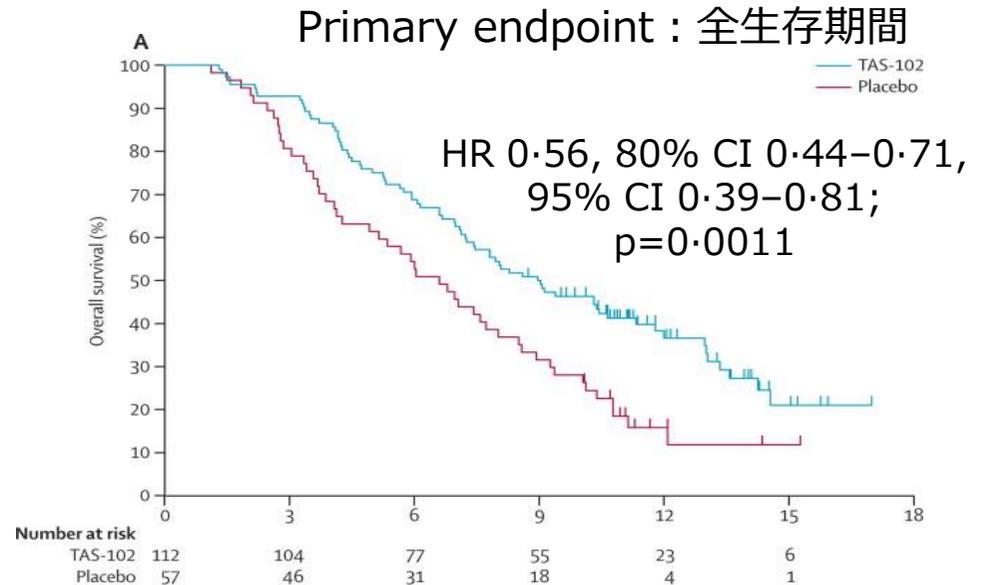
- **標準治療**と**試験治療**をランダム化
- 登録数は単群より多く、phase III より少ない
 - 100-200例前後
 - 有意水準 α は少し大きめ(10-20%)
- エンドポイント
 - 多くの場合はPFS、時折OSや奏効割合
- 規模の小さいphase III ではない
 - あくまでphase III の前の予備試験
 - 有意になってもphase III が必須



ランダム化スクリーニングデザインの例



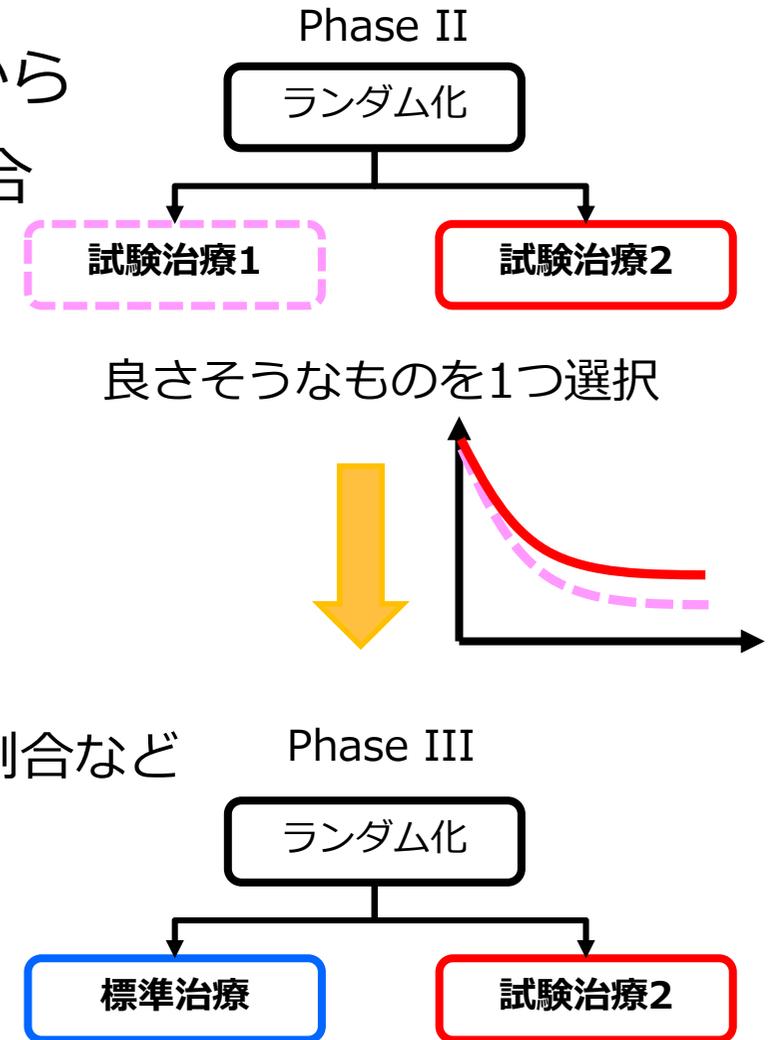
菅井 敬一, et al. NEJM. 2015;372(20):1909-19.



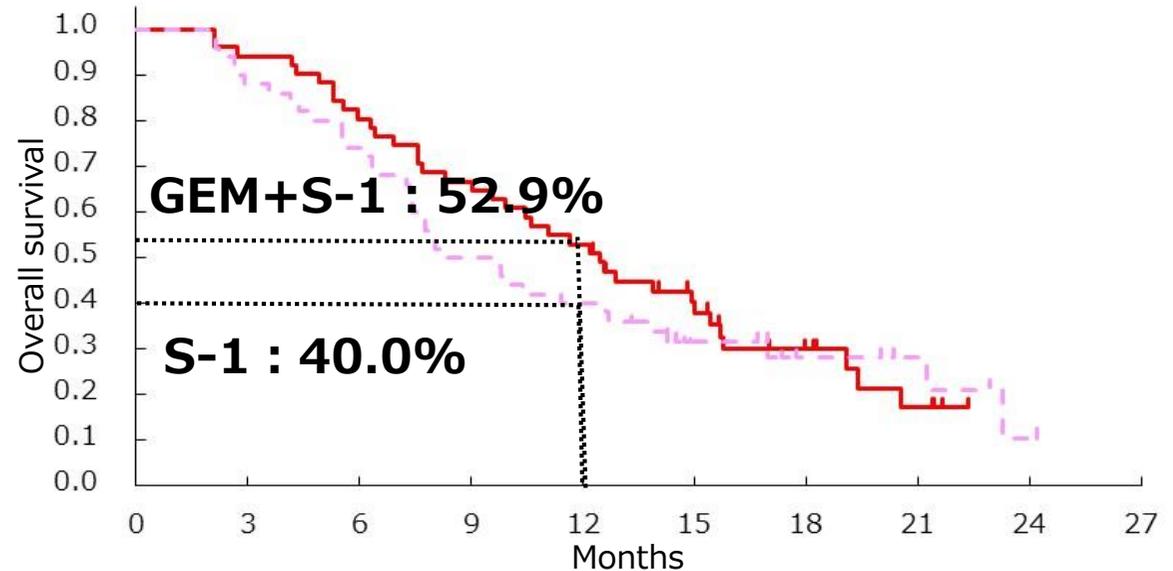
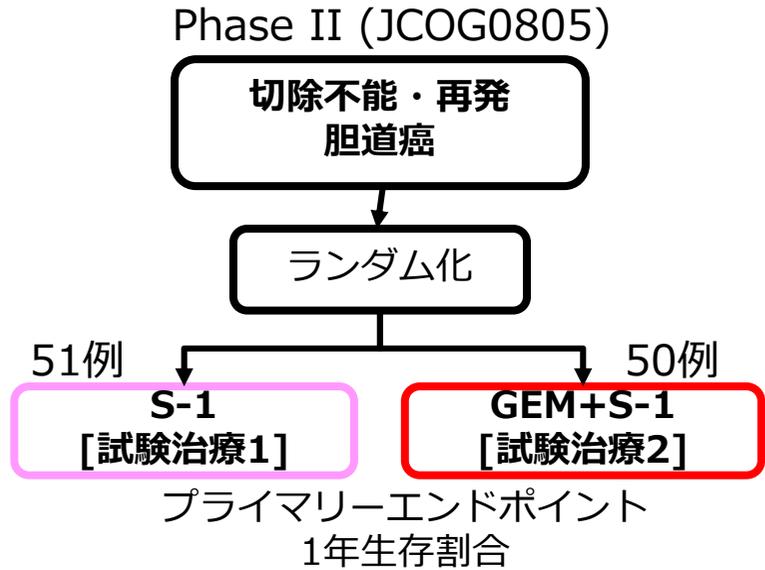
<https://www.icrweb.jp>

ランダム化セレクション(選択)デザイン

- 複数の試験治療の候補があり、その中から phase III にあげる1つを選択したい場合
- 複数の試験治療どうしをランダム化
- 登録数：単群のphase II を約2試験分
 - 100例未満
- エンドポイント
 - 奏効割合、6か月無増悪生存割合、1年生存割合など
- 効果が少しでも良い方を選ぶ
 - 他の試験治療と比べて悪くはないと確信



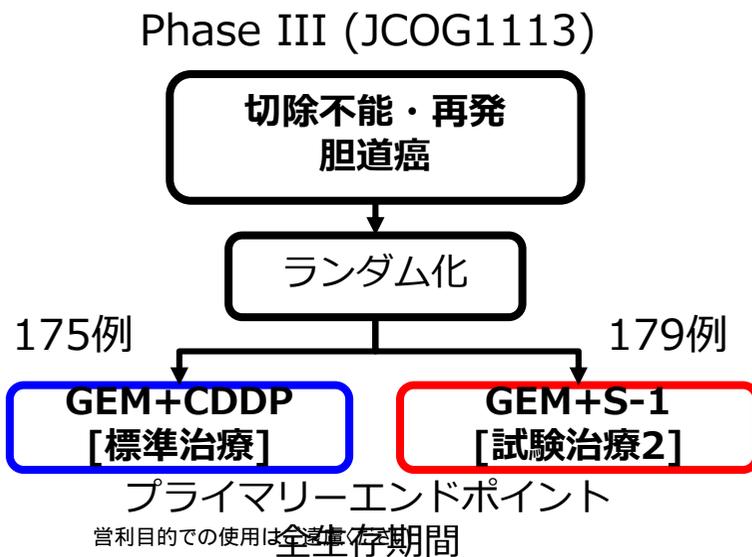
セレクションデザインの例



Morizane C et al. Cancer science. 2013;104(9):1211-6.

1年生存割合が点推定値で優った
(統計的に有意差があるわけではない)

GEM+S-1療法を標準治療の
GEM+CDDP療法とのphase III 試験の
試験治療に選択



Mizusawa J et al. JJCO. 2016;46(4):385-8.

Phase II 試験デザインのまとめ

- Phase II 試験は主に有効性による候補薬のスクリーニング
- 適切なヒストリカルコントロールが存在すれば
 - 奏効割合をprimary endpointとした単群試験
 - 無効中止を考慮した2段階のデザイン
- 適切なヒストリカルコントロールがない場合はランダム化
 - 標準治療と試験治療をランダム化したスクリーニングデザイン
 - 大きな効果が見られて、有意になってもphase IIIが必要
- 試験治療の候補が複数の場合はランダム化
 - 試験治療どうしをランダム化したセレクションデザイン
 - 次は、標準治療と比較するphase IIIが必要

Phase III のデザイン

Phase III 試験の目的と概要

- 目的

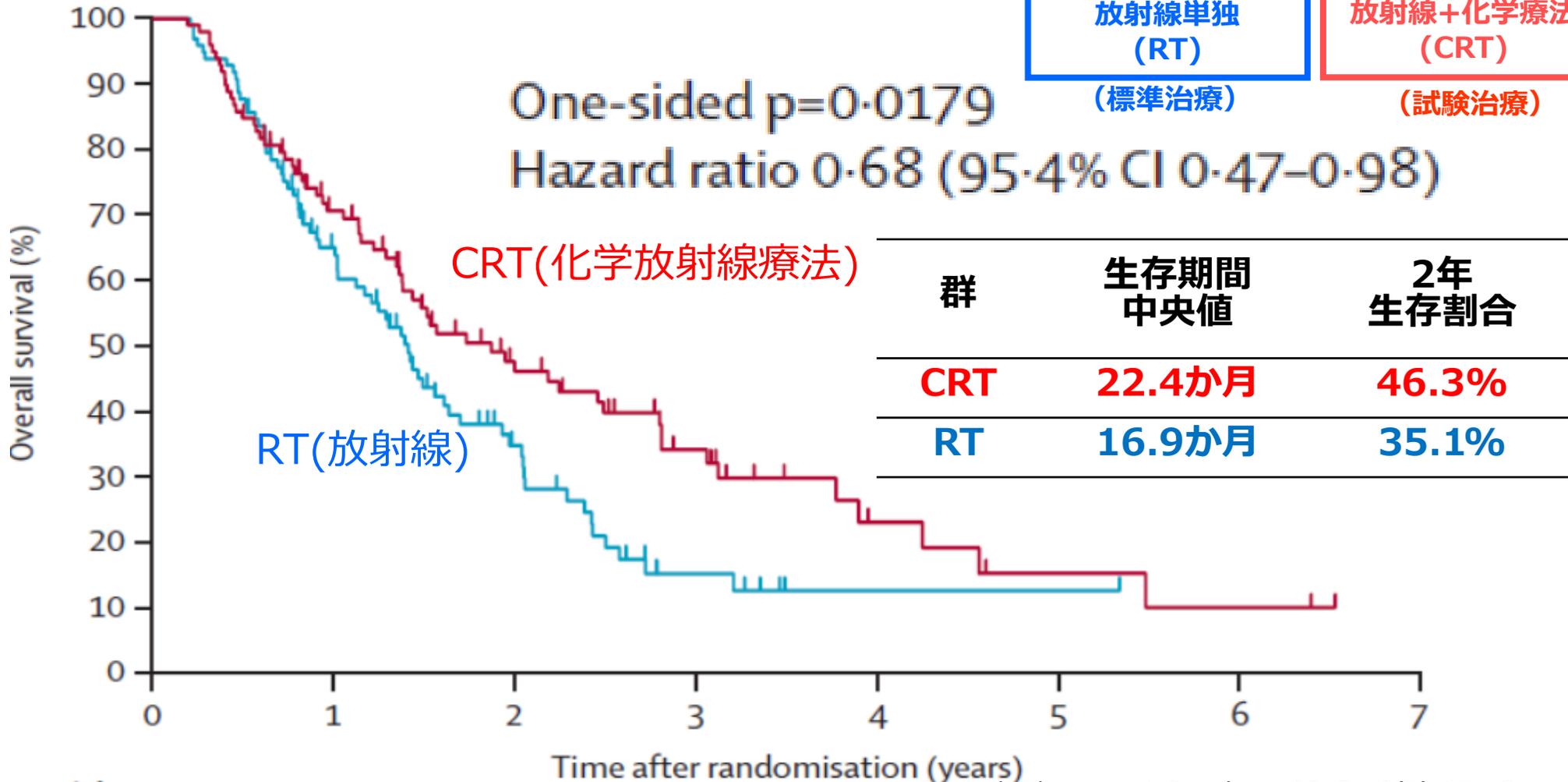
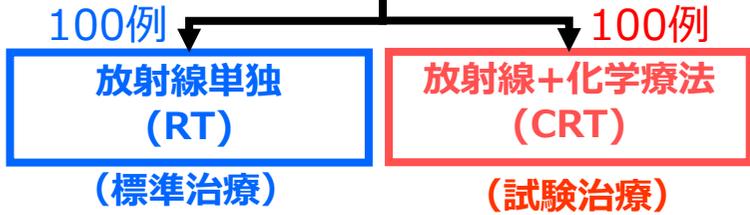
- 試験治療が標準治療（or 承認）になりうるか？を判断

- 典型的なデザイン

- 対象：phase I とphase II よりも広い適格規準を満たした患者
 - 得られた結論を適用できる範囲（一般化可能性 [外的妥当性]）を考慮
- 登録数：数百例-数千例
- 参加施設：一般病院も含む
- エンドポイント：OS、無再発生存期間など
 - 患者のベネフィットを直接反映する真のエンドポイント
- ランダム化による比較可能性の担保

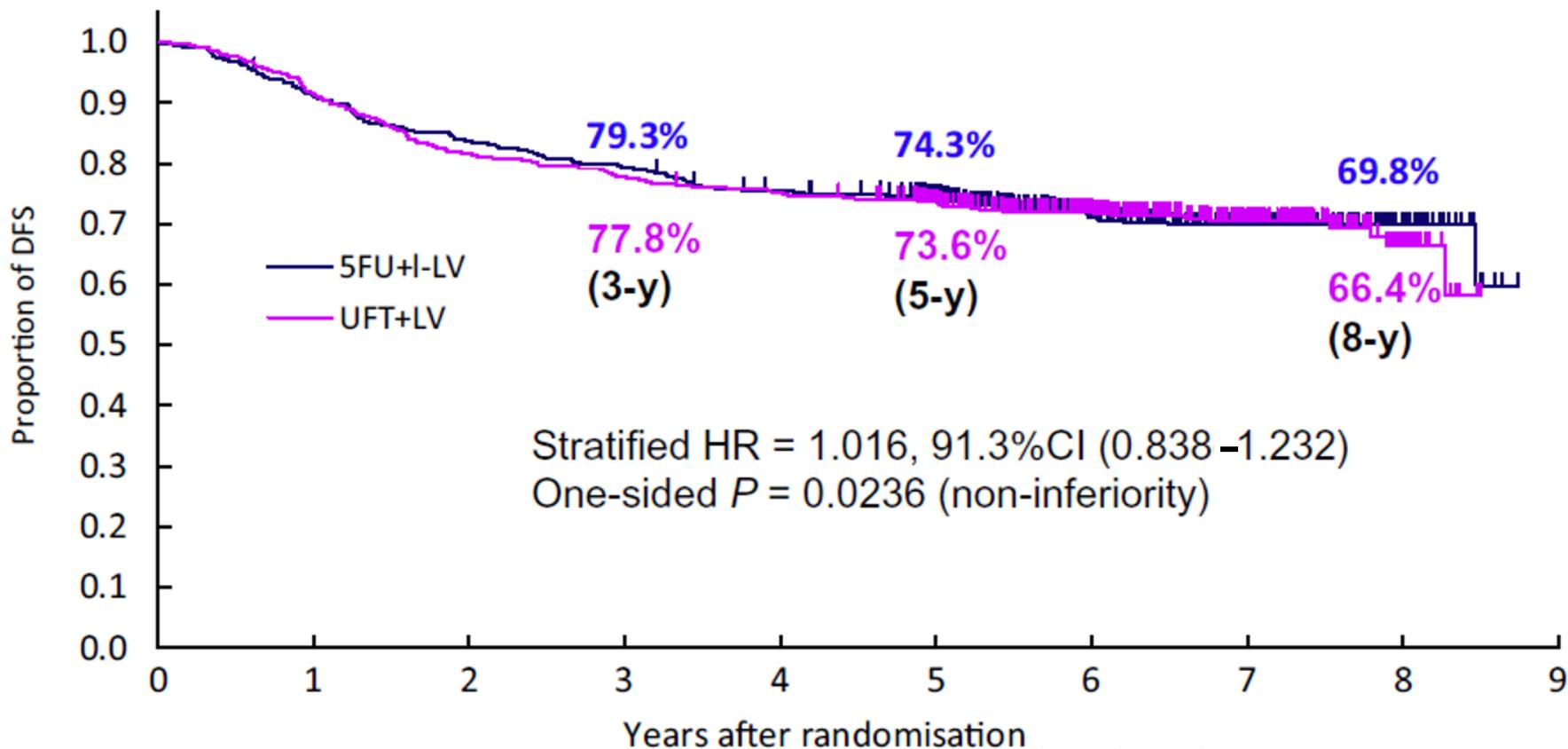
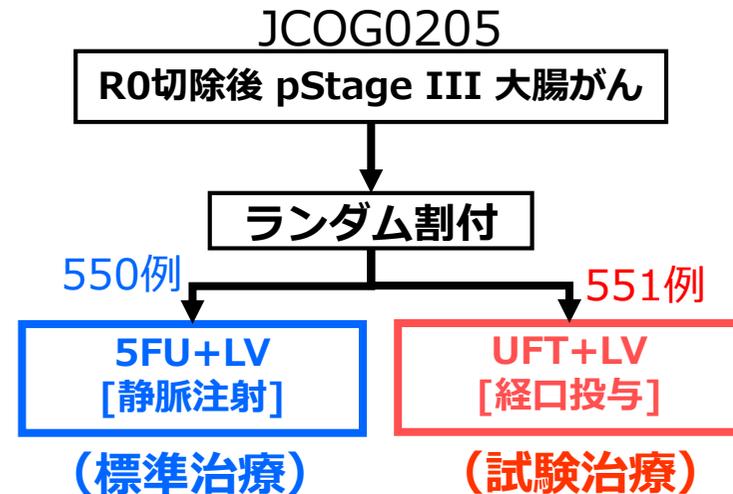
Phase III 試験で勝つということ

試験治療(CRT)が標準治療(RT)を全生存期間で上回っているので、CRTが新たな標準治療

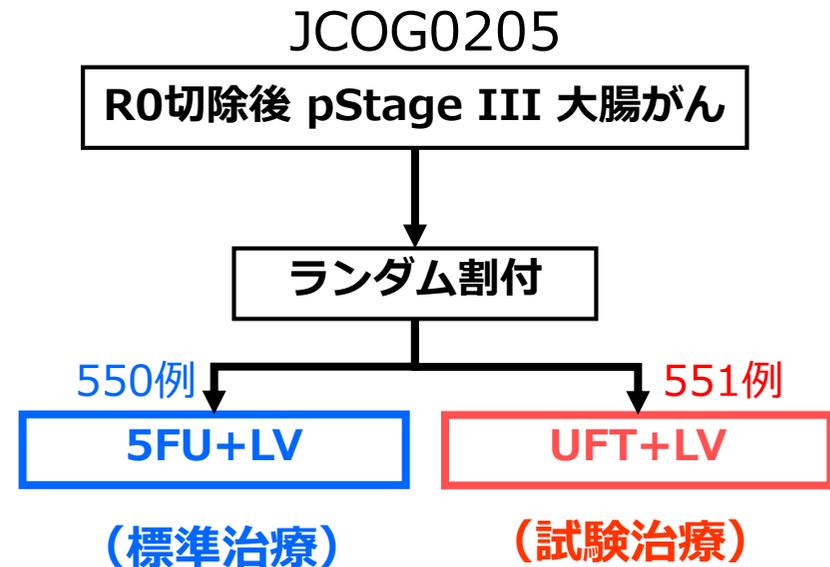
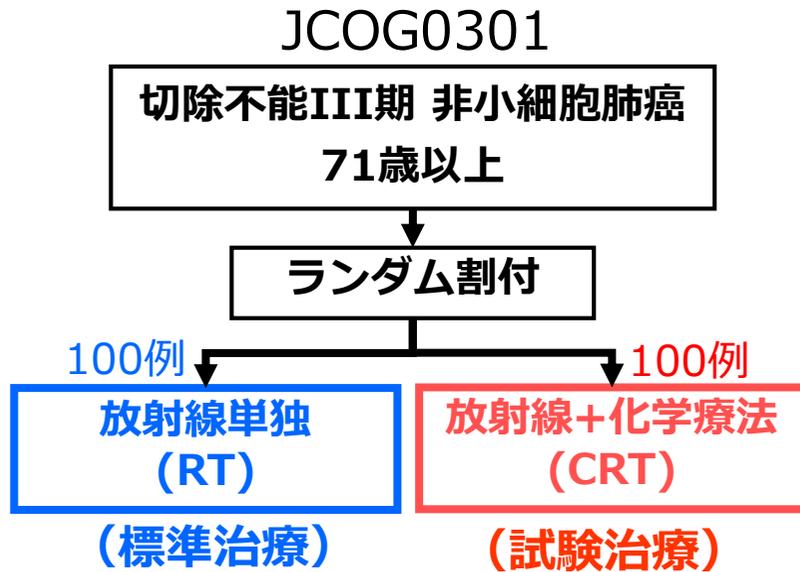


試験治療は上回らないとダメ？

- 試験治療(UFT：経口剤)が標準治療(5FU：静脈注射)を無病生存期間で上回っていないので、標準治療は5FUのまま？



比較の種類：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
- 試験治療は標準治療より毒性等が強い(Toxic new)

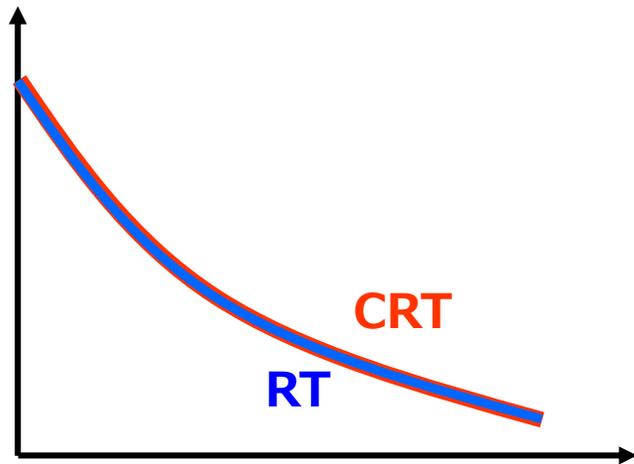
非劣性試験

- 試験治療は有効性で一定以上劣らなければ良い
- 試験治療は標準治療より毒性等が弱い(Less toxic new)

試験計画時の優越性/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

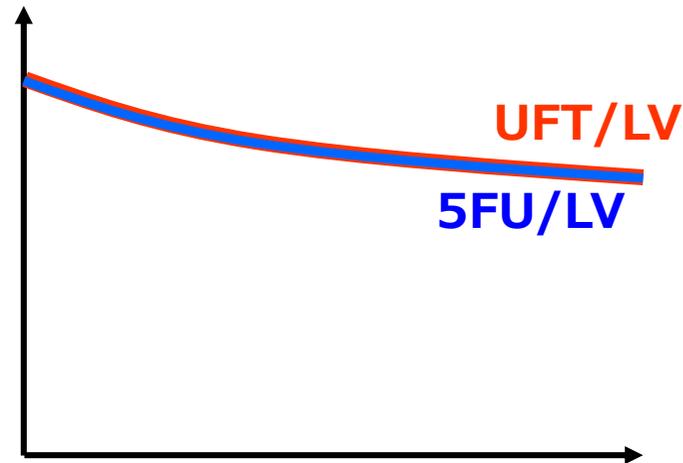
JCOG0301の場合



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

営利目的での使用はご遠慮ください

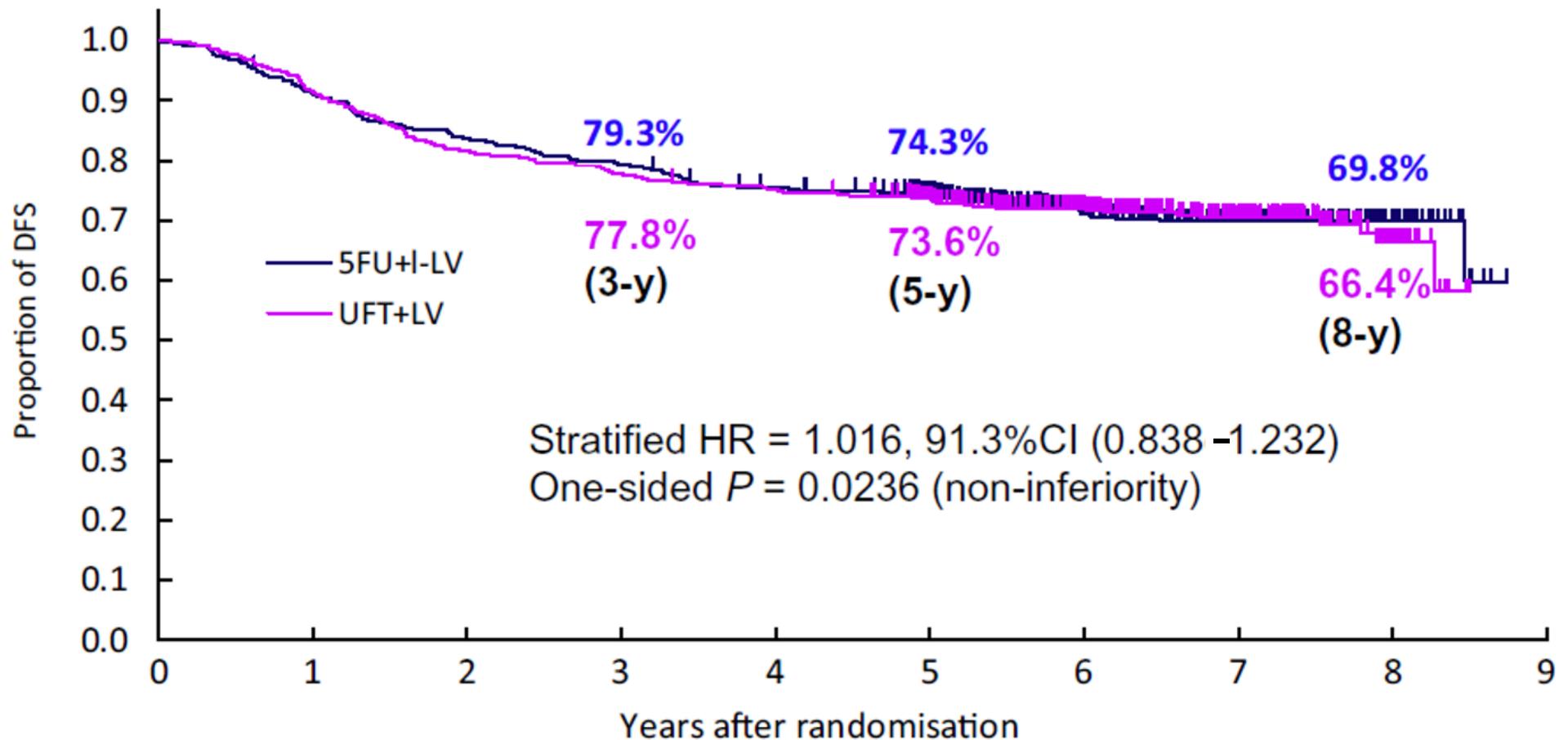
JCOG0205の場合



経口剤の**UFT/LV群 (Less toxic new)**
と静注の**5FU/LV**の**無病生存期間**が同じ
なら、標準治療は**UFT/LV**
→ 非劣性試験

非劣性試験の例：JCOG0205

試験治療(UFT：経口)が標準治療(5FU：静脈注射)に劣っていないので、
UFTが新たな標準治療



Phase III 試験デザインまとめ

- Phase III試験は標準治療と試験治療の直接比較
- エンドポイントは患者のベネフィットを直接反映するもの
 - 理想的にはOS
- 優越性試験と非劣性試験がある
 - 試験治療が有効性で勝たなければならないなら**優越性試験**
 - 試験治療が有効性で劣らなければならないなら**非劣性試験**

Take home message

- Phase I
 - **安全性**によるスクリーニング：エンドポイントは毒性（DLT）
 - 推奨用量の決定
- Phase II
 - **有効性**によるスクリーニング：エンドポイントは奏効割合など
 - 信頼できるヒストリカルコントロールが
 - ある：単群
 - ない：標準治療とのランダム化スクリーニングデザイン
 - 試験治療の優先順位付け：セレクションデザイン
- Phase III
 - 従来の**標準治療との決勝戦**：真のエンドポイントを用いて評価
 - **優越性試験**：標準治療 vs. **toxic new** な試験治療
 - **非劣性試験**：標準治療 vs. **less toxic new** な試験治療