

がん臨床試験のデザイン

JCOGデータセンター

佐立 峻

第27回JCOG臨床試験セミナー

2024/10/5

Outline

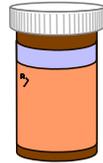
- がん臨床試験の流れとエンドポイント
- Phase I のデザイン
- Phase II のデザイン
- Phase III のデザイン

がん臨床試験の流れ

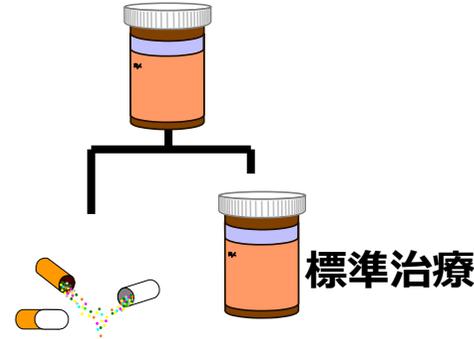
新しい
標準治療

Phase III 試験

- 標準治療を決める
- 数百例～数千例



標準治療

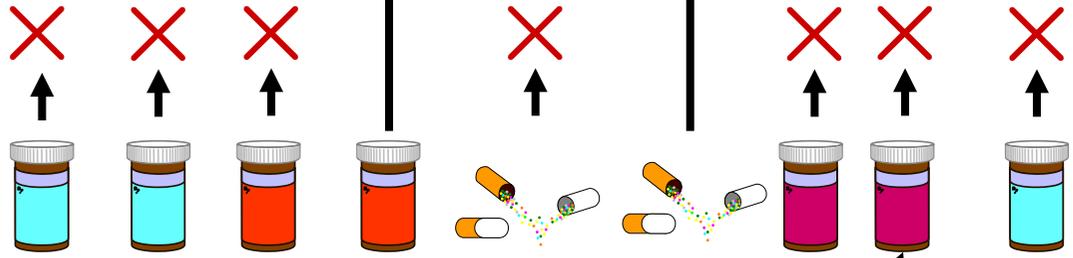


標準治療

50%

Phase II 試験

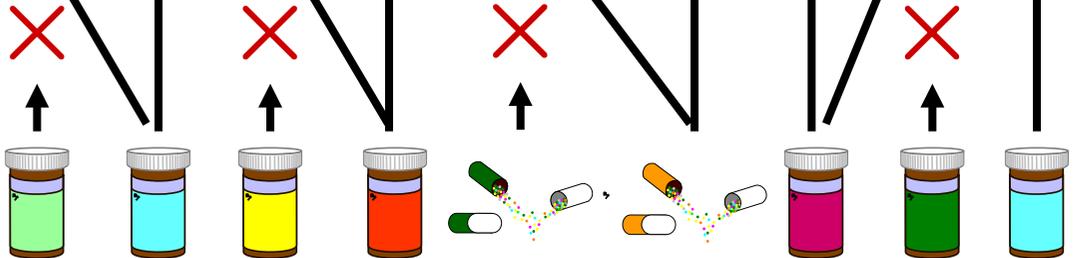
- 有効性で
スクリーニング
- 数十～100例



30%

Phase I 試験

- 安全性で
スクリーニング
- 10例～40例



70%

臨床試験の原則

臨床試験は“比較”である

Phase III 試験

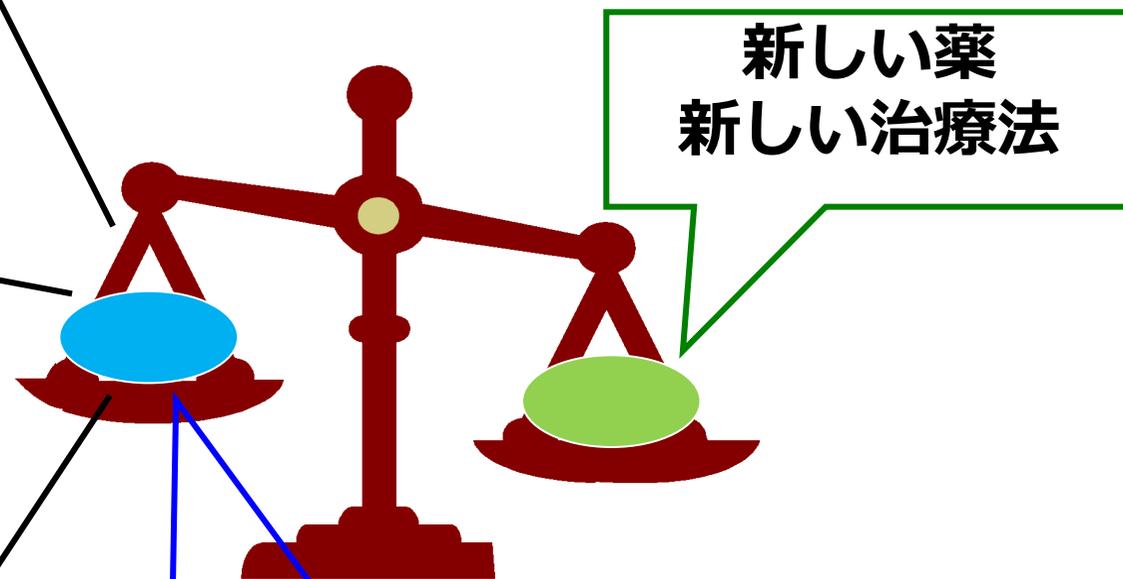
標準薬・標準治療と
総合的な有効性と安全性で
直接同時比較

Phase II 試験

過去のデータから導かれる
有効性の水準で比較

Phase I 試験

過去のデータから導かれる
安全性の許容水準で比較
(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（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”

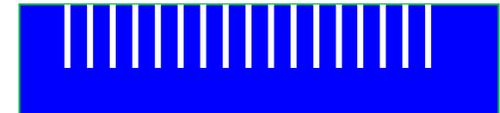
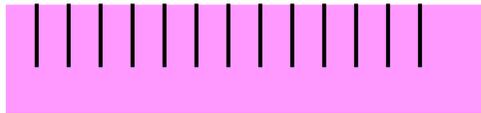
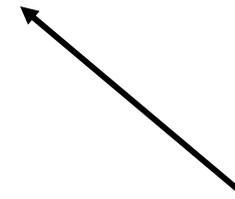
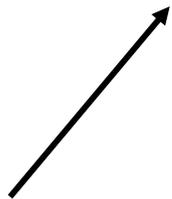
- Richard Simon



エンドポイント (ものさし)

benefit 小

benefit 大



有害事象発生割合

奏効割合
(腫瘍縮小割合)

QOLスコア

全生存期間
(OS)

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

営利目的でのご利用はご遠慮ください

<https://www.icrweb.jp>

エンドポイントの分類

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

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)

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

(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

真のエンドポイントとはいえない

Time to event 型 エンドポイントの定義

- Time to event 型：ある特定の時点（起算日）から目的のイベントのうち、いずれか最も早いものが起きるまでの時間

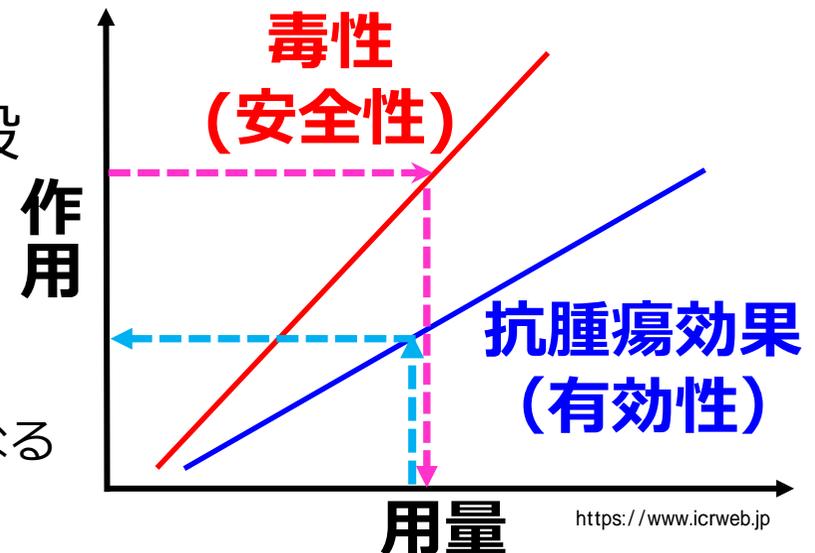
エンドポイント	イベント（いずれか早いもの）		
全生存期間 Overall survival (OS)	あらゆる死亡 (ハード)	-	-
無増悪生存期間 Progression-free survival (PFS)	あらゆる死亡	増悪 (ソフト)	
無再発生存期間 Relapse-free survival (RFS)	あらゆる死亡	再発 (ハード)	
無病生存期間 Disease-free survival (DFS)	あらゆる死亡	再発	二次がん



Phase I 試験のデザイン

Phase I 試験の目的と概要

- 目的：Phase II 試験に進むかどうかを決めること
 - Phase II 試験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



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

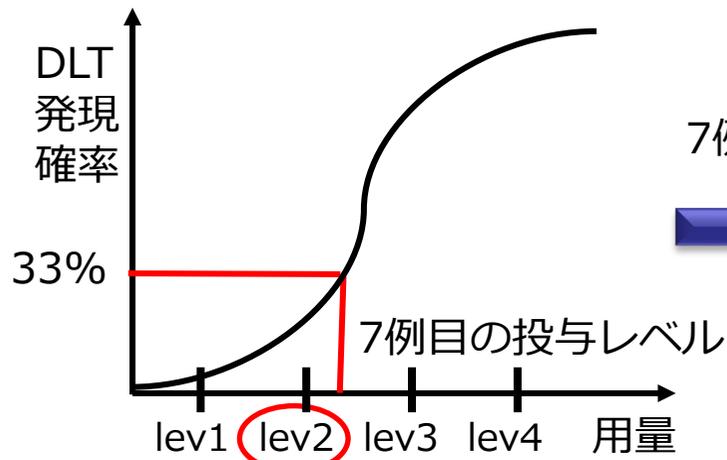
各薬剤の最初に人に投与したphase I 試験の患者数と用量レベル数

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	RDは用量レベルが高いものと考ええると、用量レベル数が多いということは、低レベルの患者はRDよりもかなり低用量で投与されたと考えられる
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	

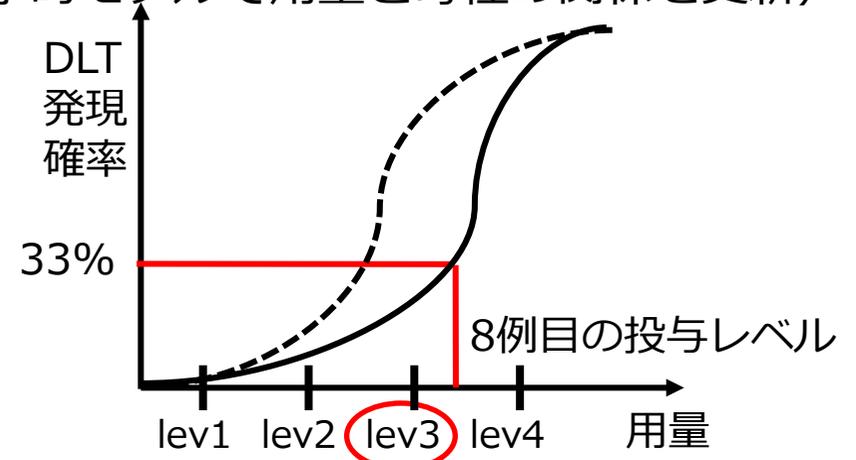
CRM : Continual Reassessment Method

- 用量と毒性の係に数学的モデルを設定
- 登録された患者のDLT有無の係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
用量と毒性の係
(数学的モデルで決まる)



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

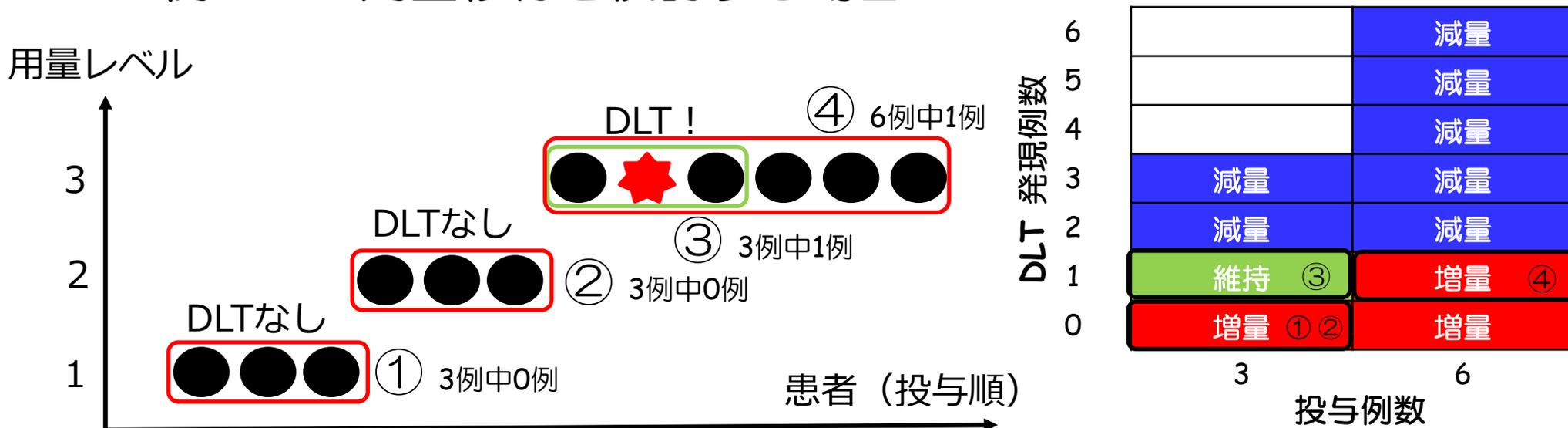


素早く用量レベルをあげられるので、低用量で投与される患者数が少なく、最適な用量付近で投与される患者数が増える可能性が高くなる

BOIN : Bayesian Optimal Interval design

- 用量ごとに毒性の発現確率に数学的モデルを設定
- 次用量の決定ルールを事前に増減表としてプロトコールに記載することができる
- 現在用量に登録された患者数とDLT発現数の情報を使って次用量の決定

3例ごとに用量移行を検討する場合



事前に増減ルールを明示できるため、CRMよりも運用コストを下げることもできる
 最適な用量付近で投与される患者数が多くなる可能性が高くなる

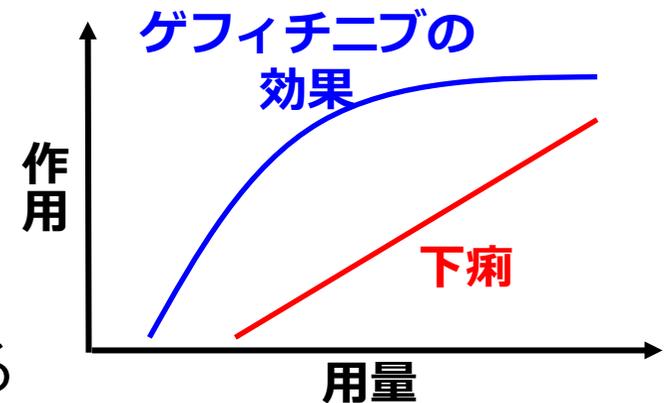
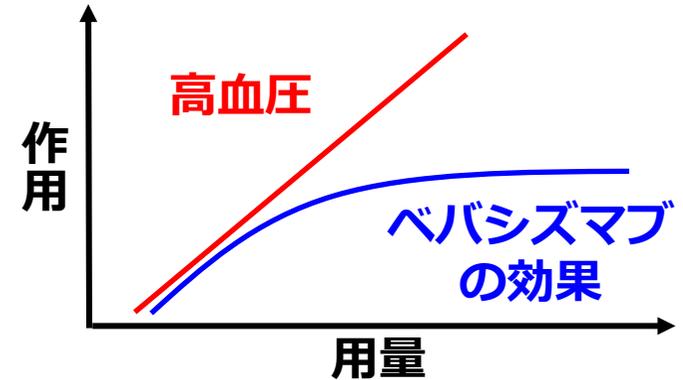
拡大コホート (Expansion cohort)

- Dose escalation phaseの後、目的に応じたコホートに数十例の患者を追加し、安全性・有効性などを確認する方法
 - 推奨用量での安全性を再検討 (ex : パゾパニブ)
 - 腫瘍縮小効果の見られたがん種の有効性の検討 (ex : ペムブロリズマブ)
- 分子標的薬や免疫チェックポイント阻害薬の登場と開発の迅速化により利用が増えつつある
 - 効果のありそうなシーズの増加 & 効果のある患者の特定が開発成功の鍵に
 - 2006年は12% → 2011年は38%で使用

Manji A., J Clin Oncol 2013;31:4260-7.

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
 - 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



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やBOINの方が3例コホートより好ましい
 - CRM: 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
 - 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

Phase II 試験のデザイン

Phase II 試験の目的と概要

- 目的：Phase III 試験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 試験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
 - (バスケット試験のようにがん種を限定しないデザインも登場)
 - 登録数：20-60例
 - 参加施設：phase I 試験よりも多くの施設
 - **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
 - **有効性**によるスクリーニング
 - できるだけ早くphase III試験に進むかどうかを決める
 - 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：この程度しか効果がないとすると、開発をあきらめる値
 - 期待値：ある一定程度効果があれば、開発を続けたい値
 - 奏効割合：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%



単群でよい？奏効割合でよい？

- 切除不能膵臓癌

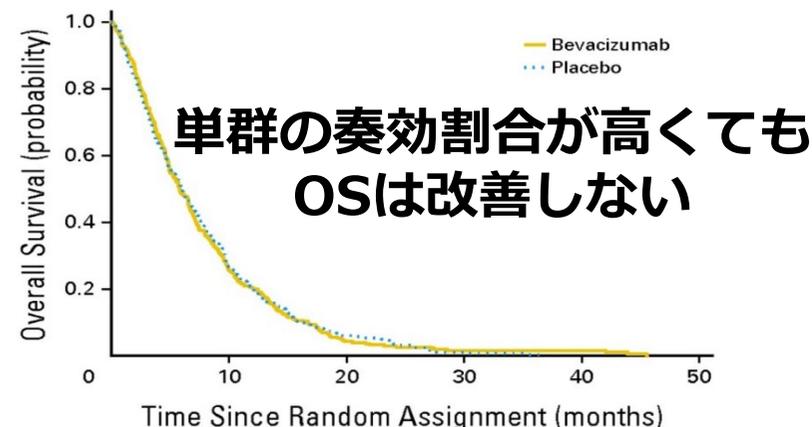
- 単群phase II試験の結果
- 奏効割合
 - ベバシズマブ群：21%
 - 当時の標準治療である
ゲムシタビン単剤：10%未満

Phase IIIへ

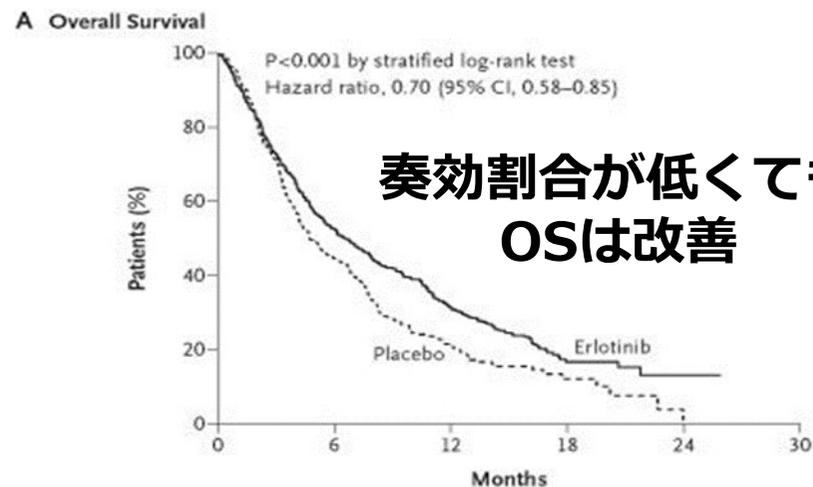
Kindler, Hedy L., et al. *JCO* 2005;23

- 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%



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



単群の年次生存割合、年次無増悪生存割合は？

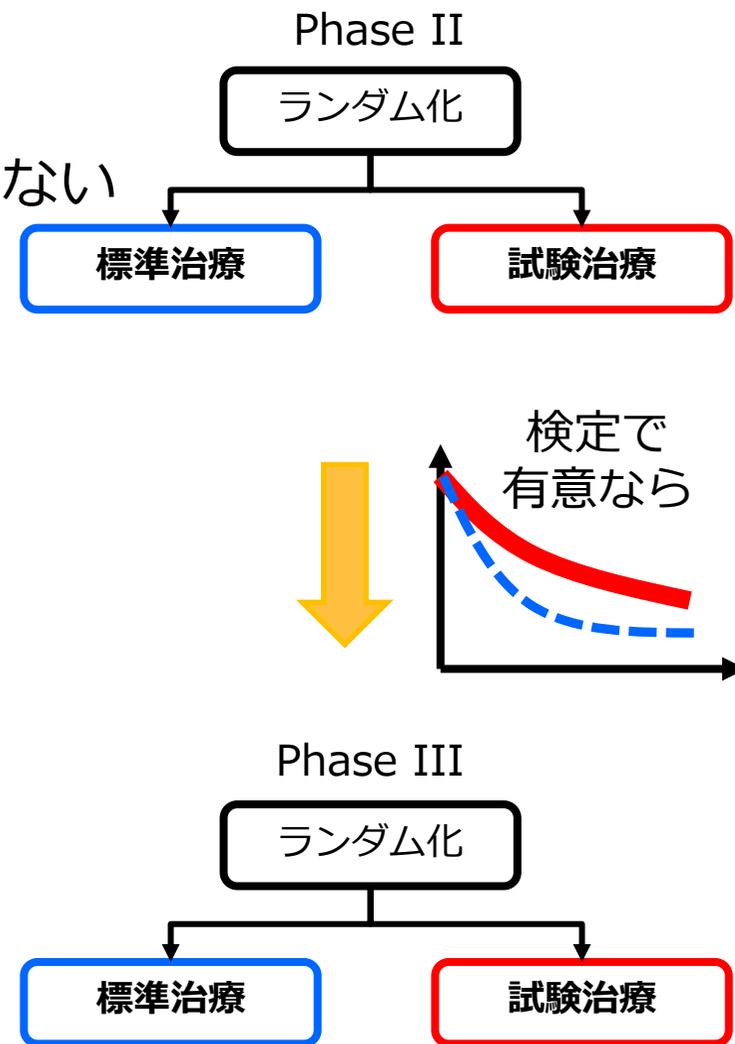
- 年次生存割合（例：1年生存割合）
 - 登録される患者の全身状態によりバラツキが生じやすい
 - 後治療の影響を受けやすい

- 年次無増悪生存割合（例：1年無増悪生存割合）
 - 増悪時期は検査間隔の影響を受けやすい
 - 増悪の定義ががん種や試験で統一されていない
 - 前立腺癌では、PSA値や疼痛の増悪の定義が不明確

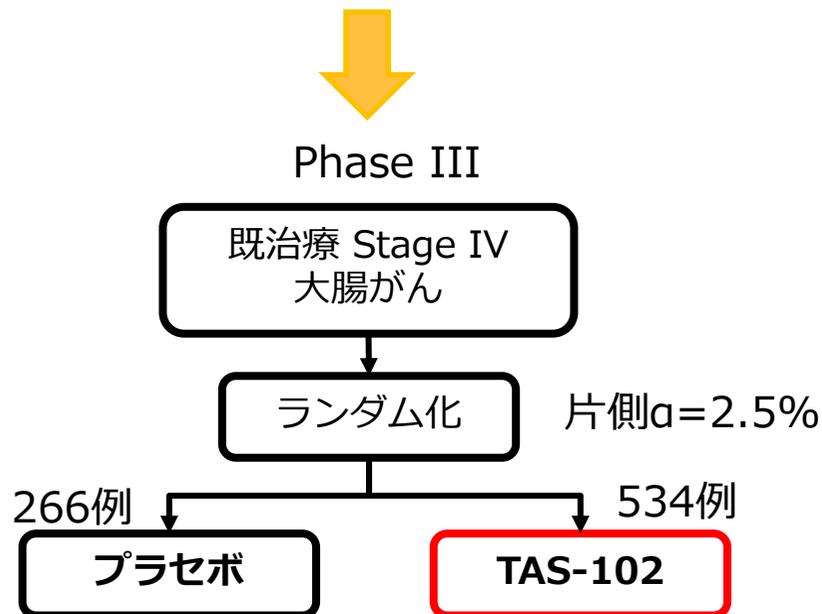
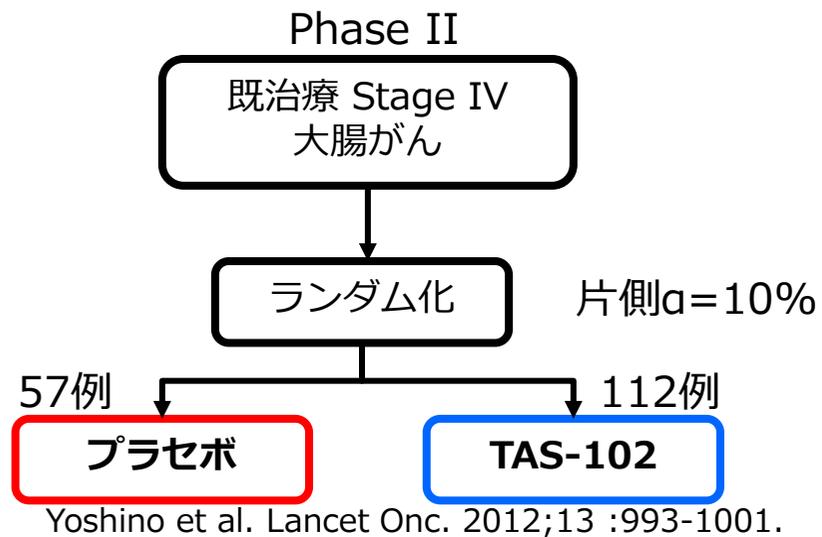
→ **ヒストリカルコントロールとの比較が困難**

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

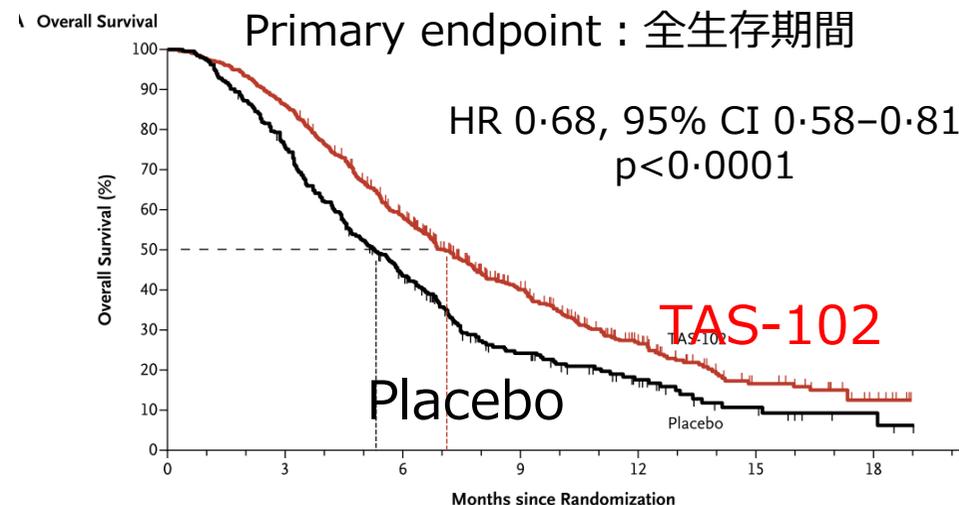
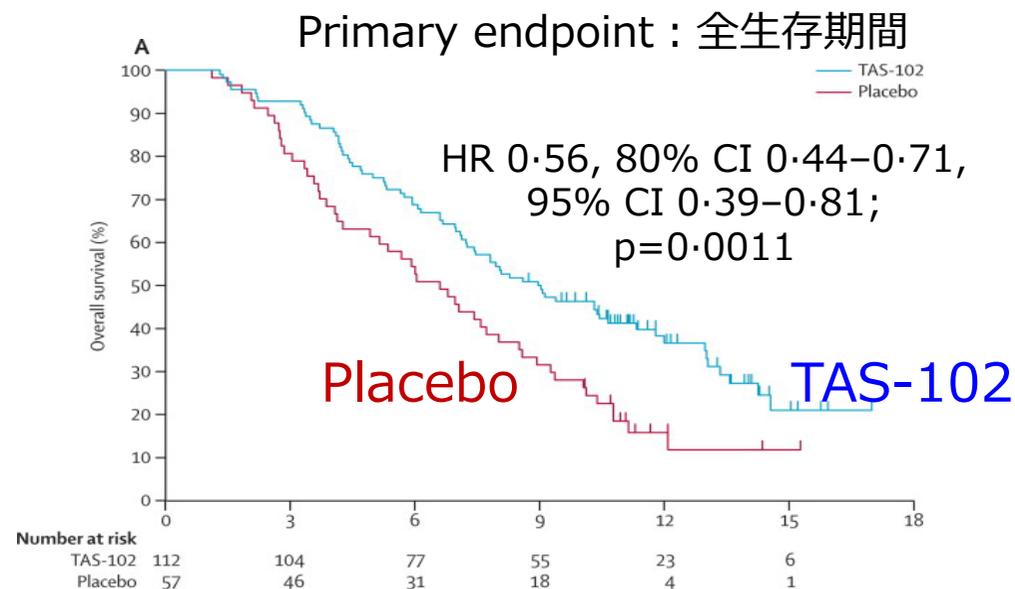


ランダム化スクリーニングデザインを用いた治療開発の例



営利目的でのご利用はご遠慮ください

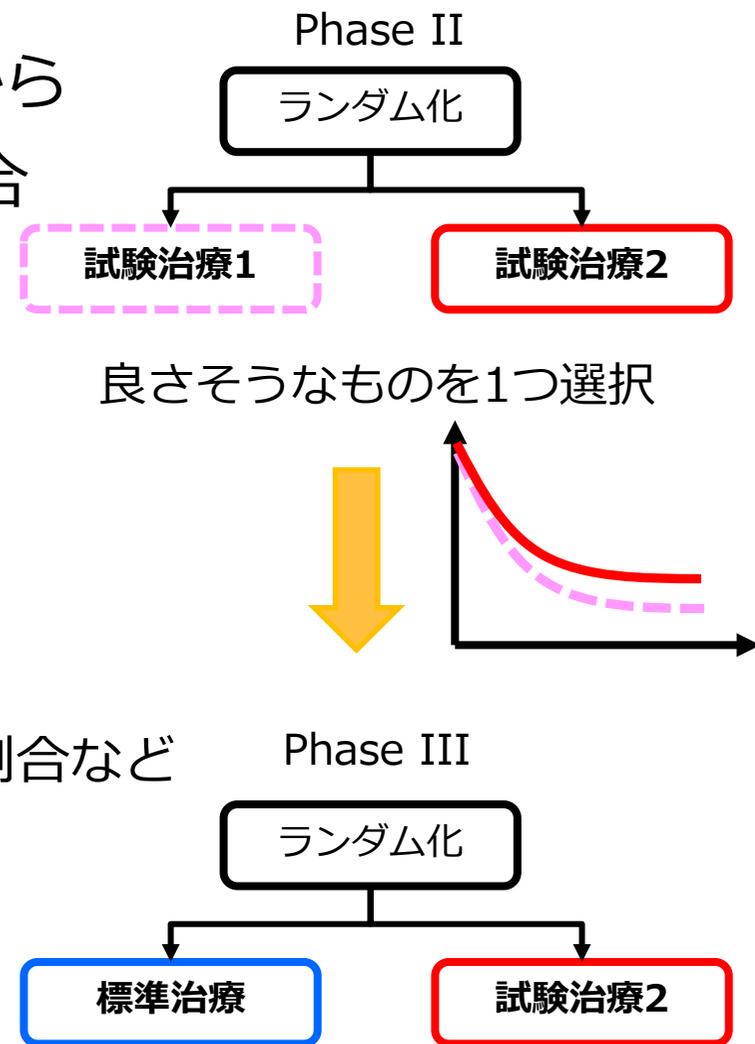
Mayer RJ, et al. NEJM. 2015;372(20):1909-19.



<https://www.icrweb.jp>

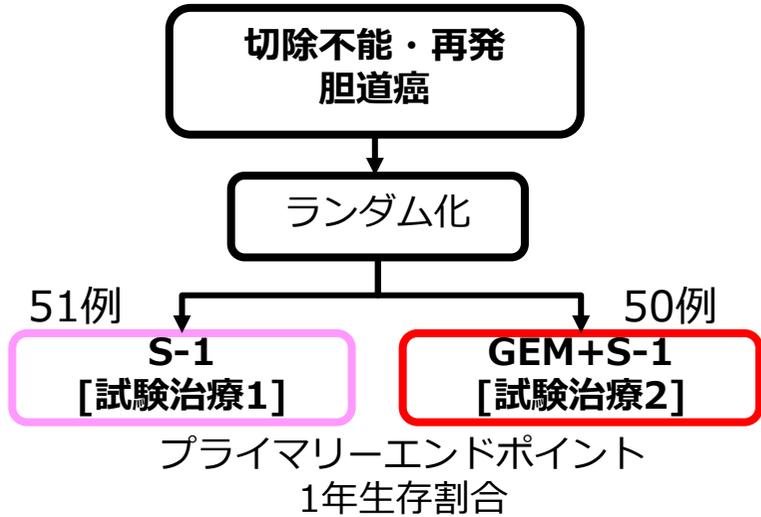
ランダム化セレクション(選択)デザイン

- 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例程度が目安
- エンドポイント
 - 奏効割合、1年生存割合、6か月無増悪生存割合など
- 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

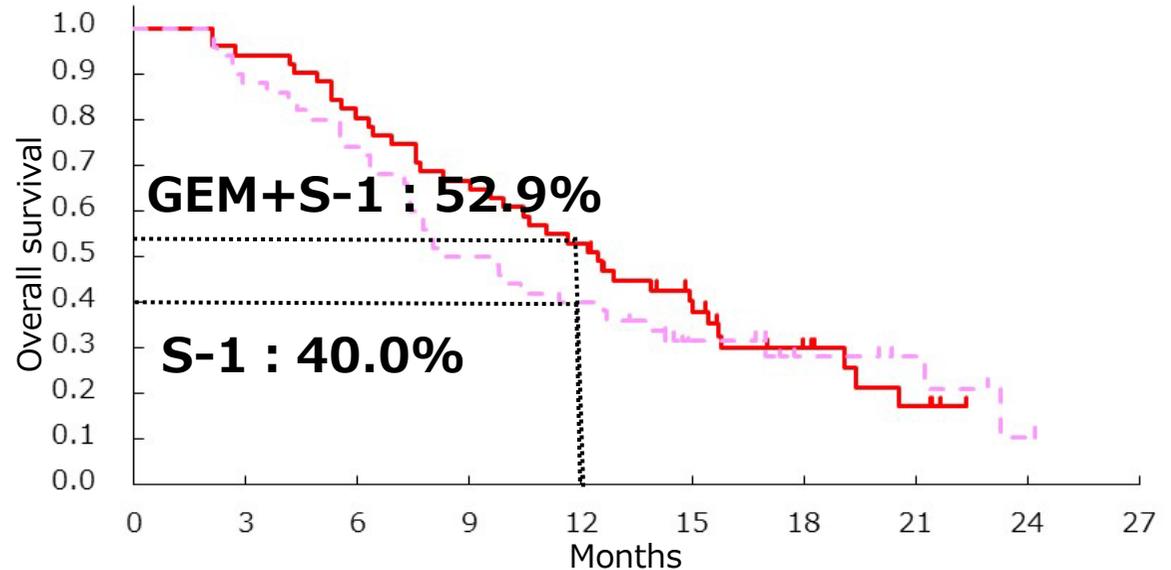
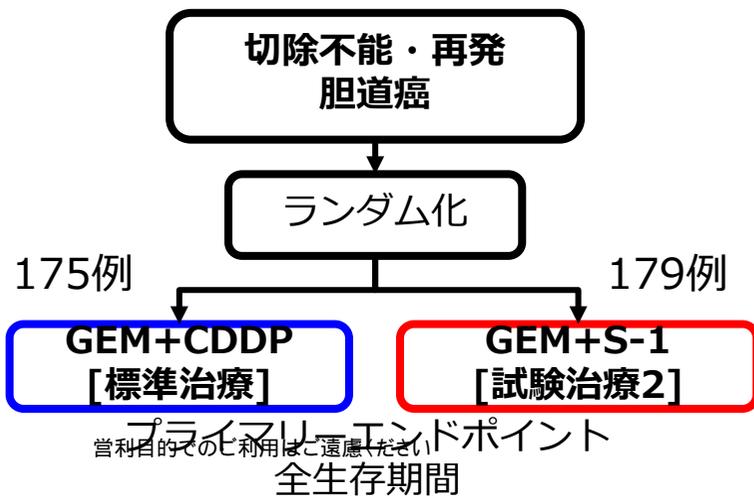


セレクションデザインを用いた治療開発の例

Phase II (JCOG0805)



Phase III (JCOG1113)



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

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試験のデザイン

Phase III 試験の目的と概要

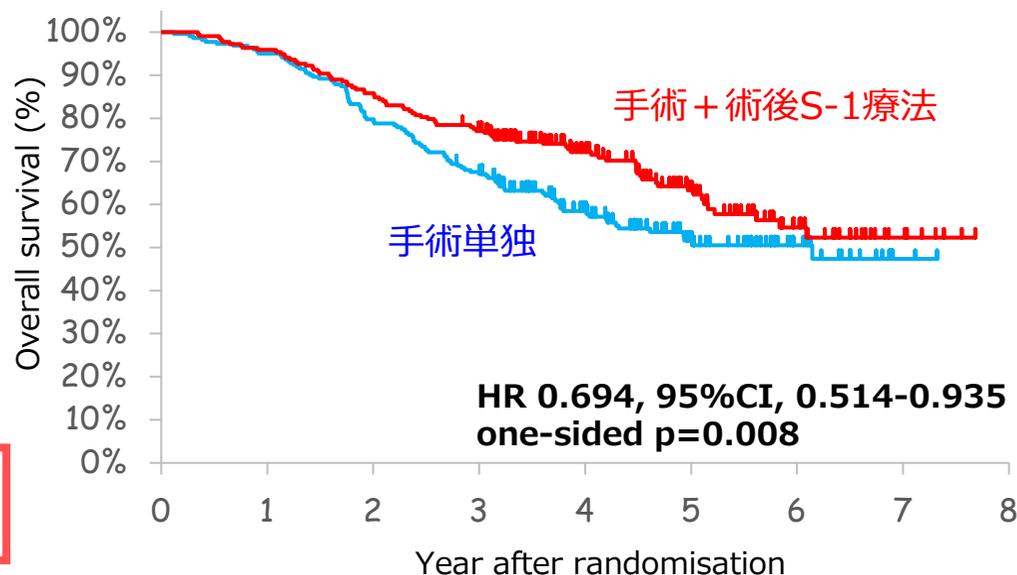
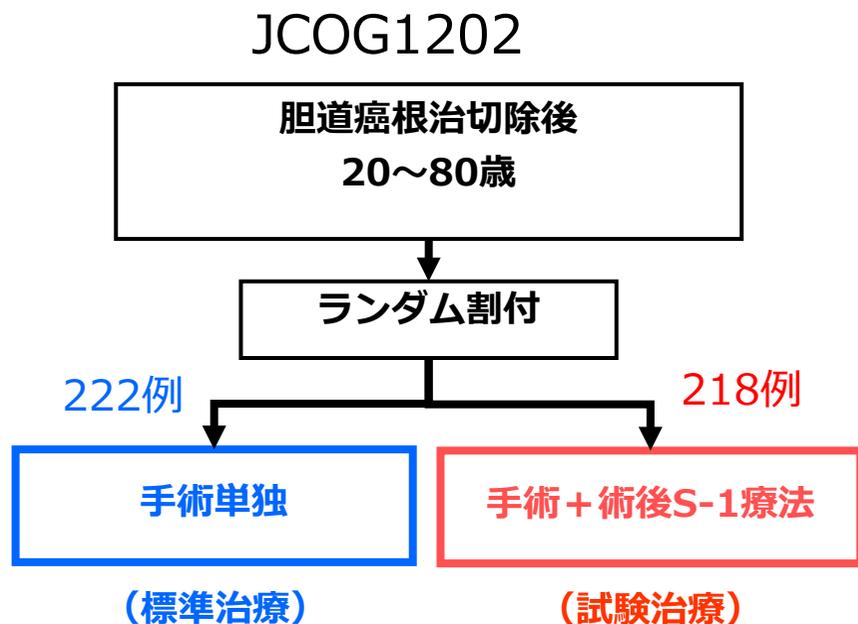
- 目的

- 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：全生存期間、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

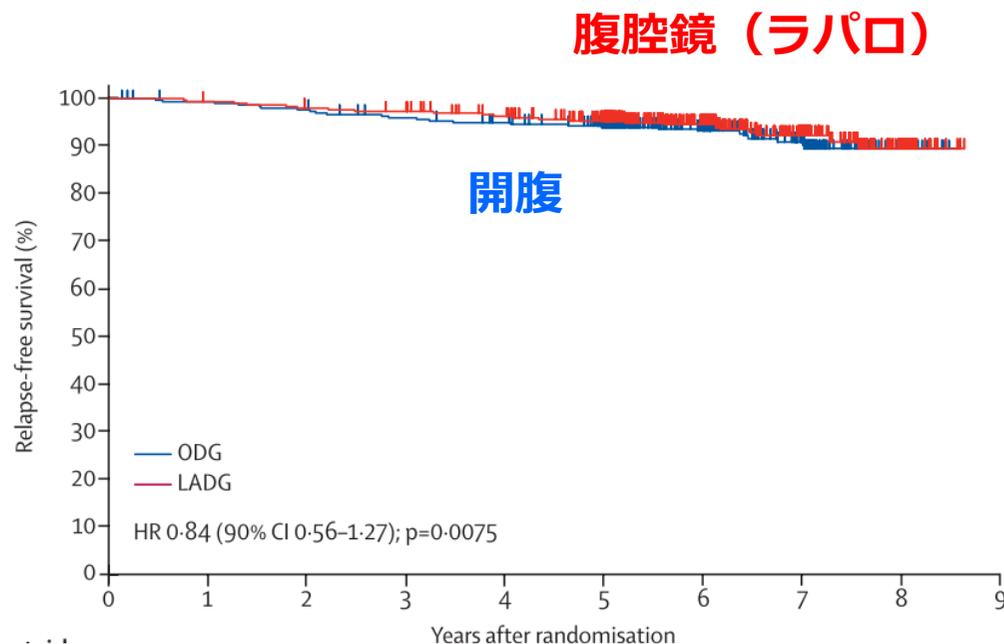
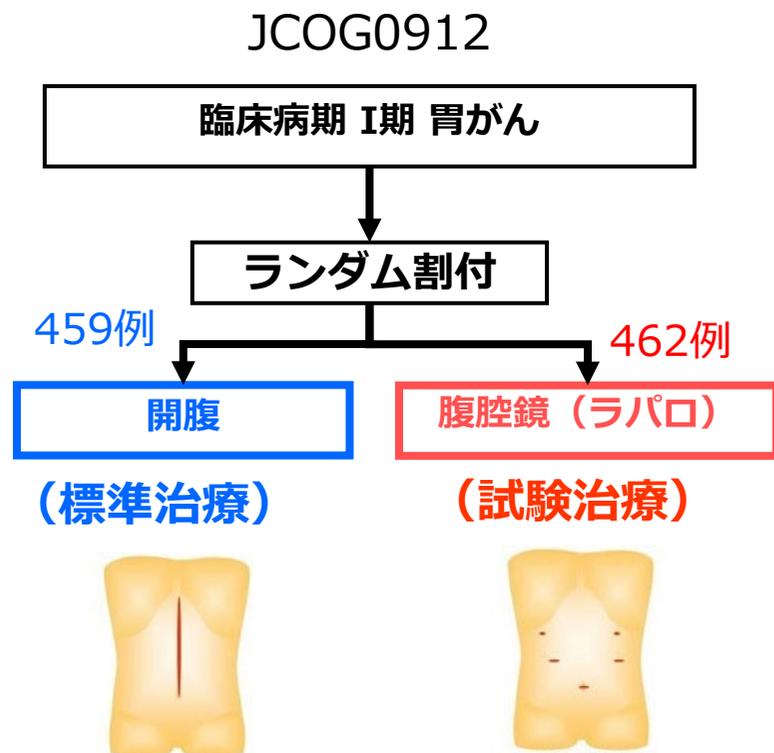


群	3年生存割合
手術+術後S-1療法	46.3%
手術単独	35.1%

試験治療(手術+術後S-1療法)が標準治療(手術単独)を

3年生存割合で上回っているので、手術+術後S-1療法が新たな標準治療

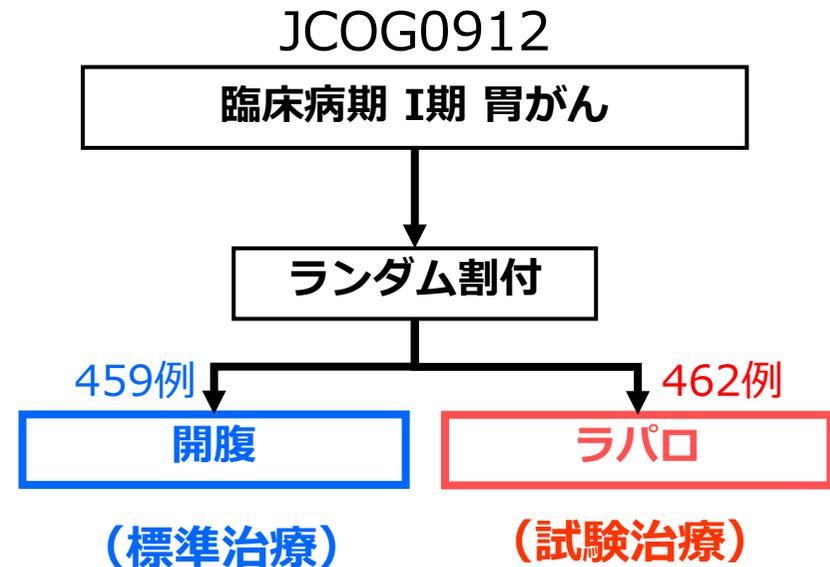
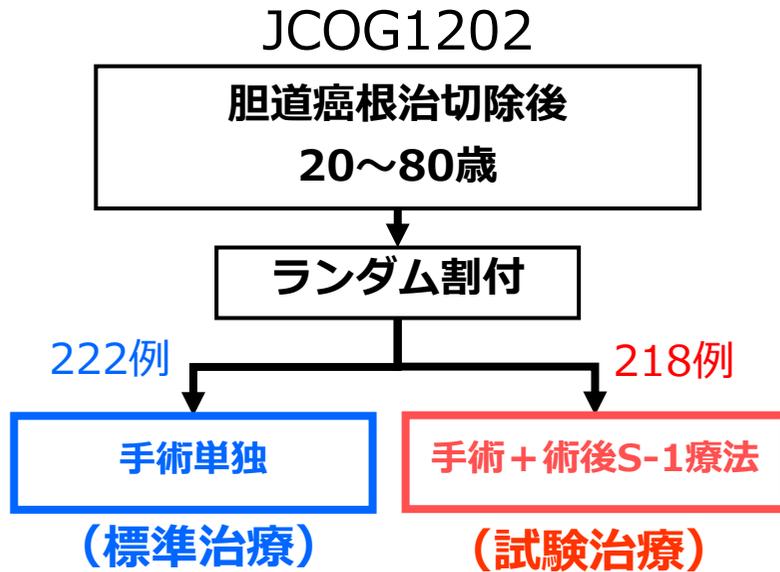
試験治療は上回らないとダメ？



試験治療(ラパロ)が標準治療(開腹)を
無再発生存期間で上回っていないので、標準治療は開腹のまま？

※ 試験治療(ラパロ)が標準治療のひとつとして位置づけられた

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**)

営利目的でのご利用はご遠慮ください

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**)

<https://www.icrweb.jp>

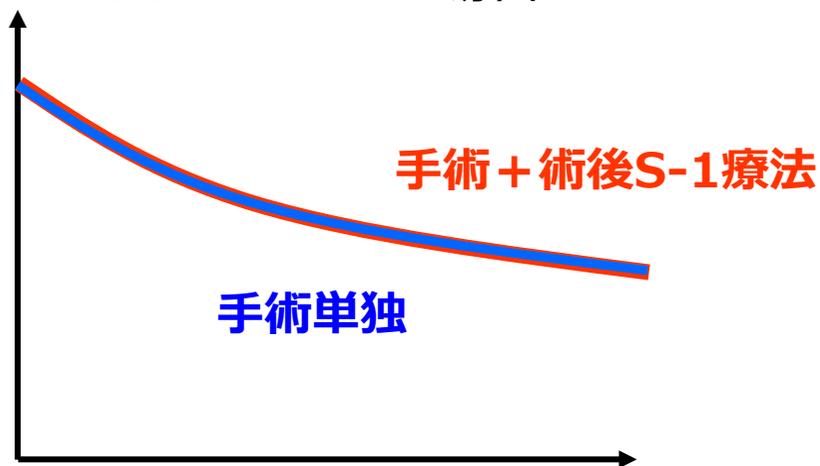
試験計画時の優越性/非劣性の決め方

- 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**

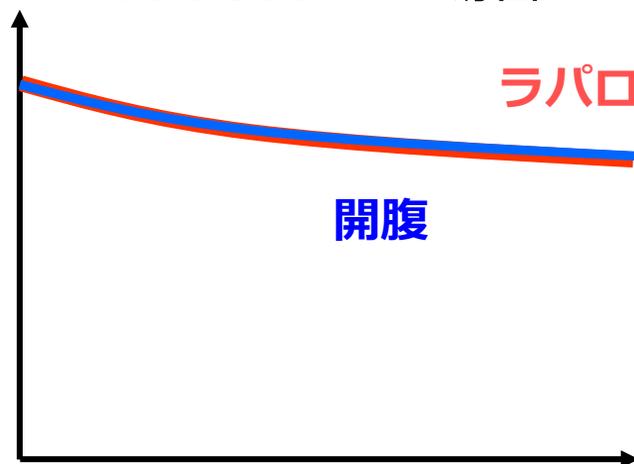
- **試験治療**を選択する → **非劣性試験**

JCOG1202の場合



毒性が強く、治療に手間がかかる
手術+術後S-1療法群 (Toxic new)と
手術単独群の全生存期間が同じなら、
標準治療は**手術単独**
→ **優越性試験** ご確認ください

JCOG0912の場合



傷が小さい**ラパロ (Less toxic new)**と
開腹の無再発生存期間が同じなら、
標準治療は**ラパロ**
→ **非劣性試験**

Phase III 試験デザインのまとめ

- Phase III試験は総合的な有効性と安全性による総合的評価
- エンドポイントは真のエンドポイントかつハードなエンドポイント
- Toxic newな治療なのかLess toxic newな治療のか
 - Toxic newな試験治療であれば優越性試験
 - Less toxic newな試験治療であれば非劣性試験

Take home message

- Phase I 試験：癌種問わない、患者さんのリスク大、単～少数の専門施設
 - **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 試験：癌種特定、患者さんのリスク中、施設数限定し専門病院主体
 - **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 試験：癌種特定、患者さんのリスク小、施設数多く一般病院含む
 - 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