

交絡調整のテクニック

～ランダム化・傾向スコアを中心に～

国立がん研究センター
研究支援センター 生物統計部
小川 岳人

平成30年度 臨床研究・治験従事者研修

2018. 12. 8 (土)

希少フラクションにおける治療開発アプローチ

- 外部データを対照群にする
- マスタープロトコール型臨床試験
- ベイズ流のデザイン・解析
- 登録促進のためのネットワーク構築
- 試験期間を長く設定
- サンプルサイズが小さくて済むような集団を対象にする
- α 、 β を緩める
- データの質を担保
- 観察研究の実施

Billingham L, Lancet Oncol 2016.

Outline

- 外部データを対照群とする場合の問題点
 - 交絡とは
 - 交絡調整のテクニック
- バスケット試験
 - 主なデザイン・解析方法
 - 階層ベイズモデルの利点・欠点

外部データを対照群とする場合の 問題点 ～交絡と交絡調整のテクニック～

希少フラクションに対する治療開発

- 新規薬剤の作用機序から、特に効果が見込まれる部分集団を遺伝子情報により同定して開発するアプローチに期待
 - － 例：ALK融合遺伝子転座に対するクリゾチニブやアレクチニブ
- 一方で、遺伝子情報により対象集団を絞り込むと、開発対象が稀な集団（希少フラクション）になることがある
- 希少フラクションに対しては、単群試験で承認が期待できるケースが少なからずある
 - － ただし、単群試験では**“治療効果”が正しく推定できるか不安**

単群試験による治療効果の推定

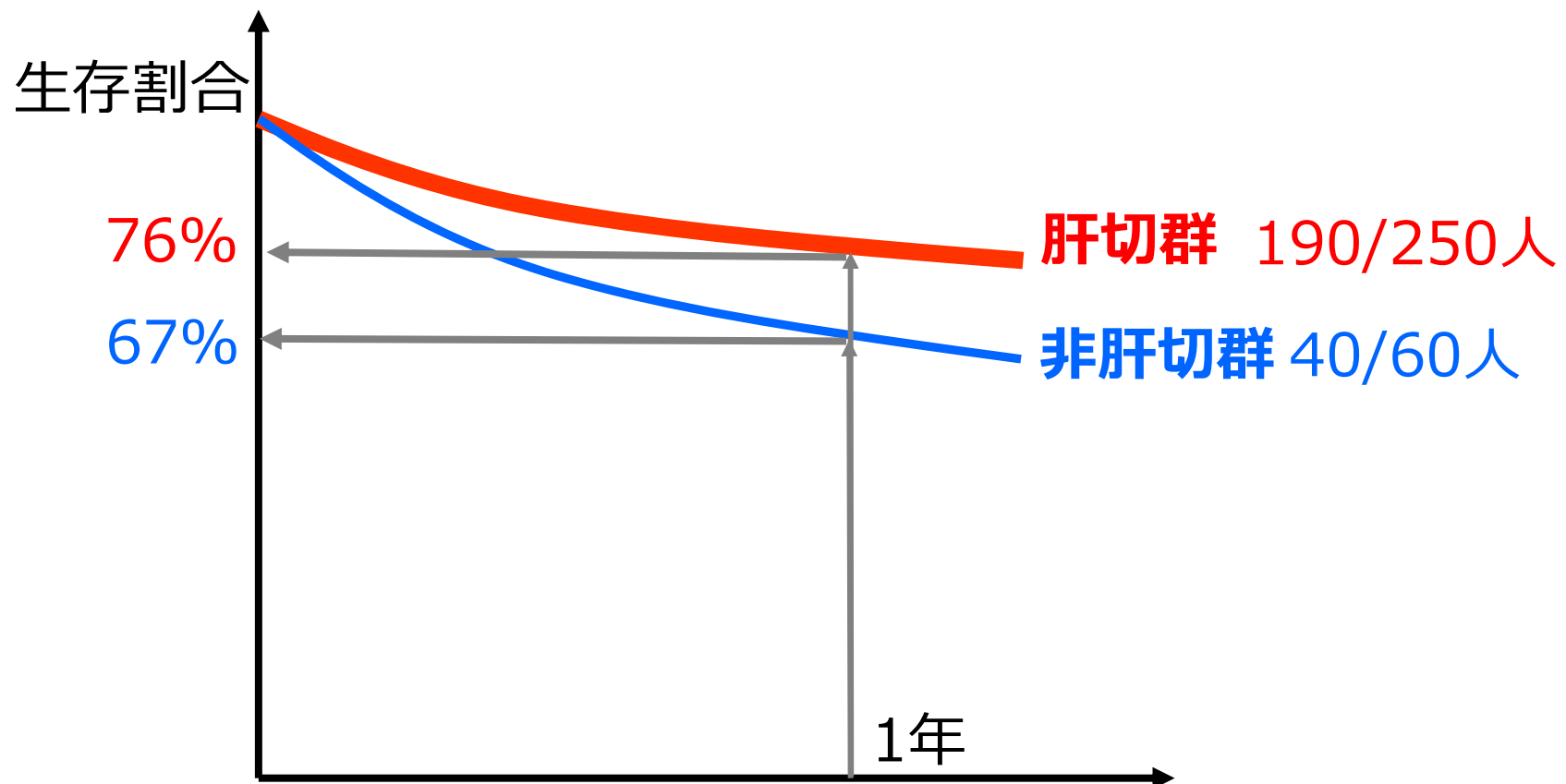
- 興味のある“治療効果”は、
「ある集団が薬Aで治療されたときの効果」と
「ある集団が標準治療で治療されたときの効果」の差
- 単群試験では、標準治療の効果として、既報を参考にした
閾値（ヒストリカルコントロール）を設定
 - 薬Aの効果が閾値を超えていれば、治療効果ありと判断

問題点

- 閾値を設定するための事前情報は十分か？
- 薬Aの効果が閾値を超えていたのは、治療効果があるから？
 - 交絡の可能性

学会で見かける発表

- 当院における肝転移を有する切除不能大腸がん患者を**肝切群**(250例)と**非肝切群**(60例)に分けてレトロスペクティブに検討した
- **肝切群**は**非肝切群**と比較し予後良好であった。

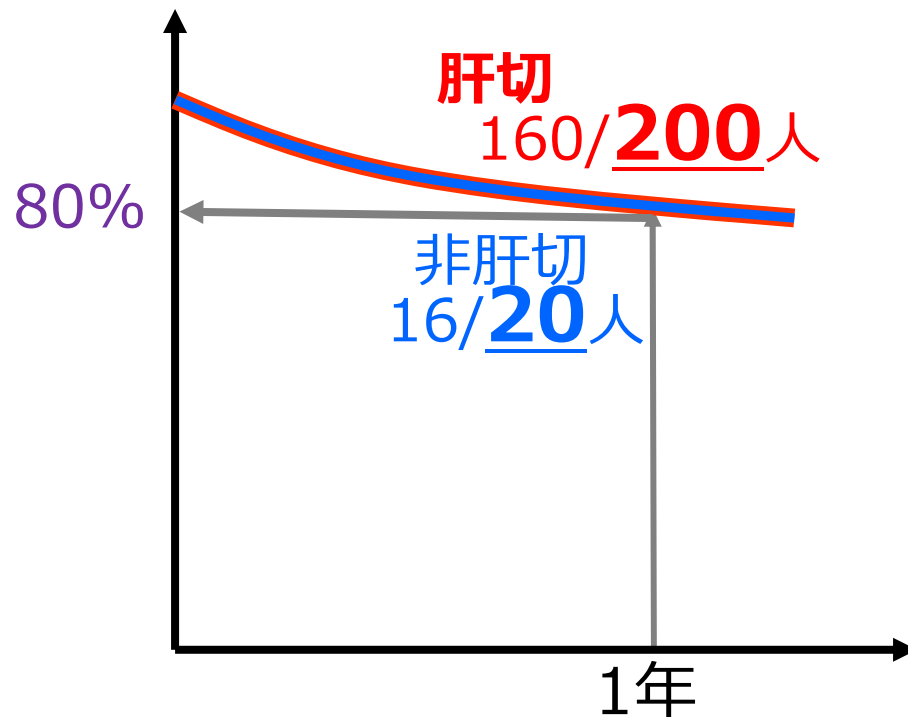


注：仮想例

最良効果で分けた場合の予後

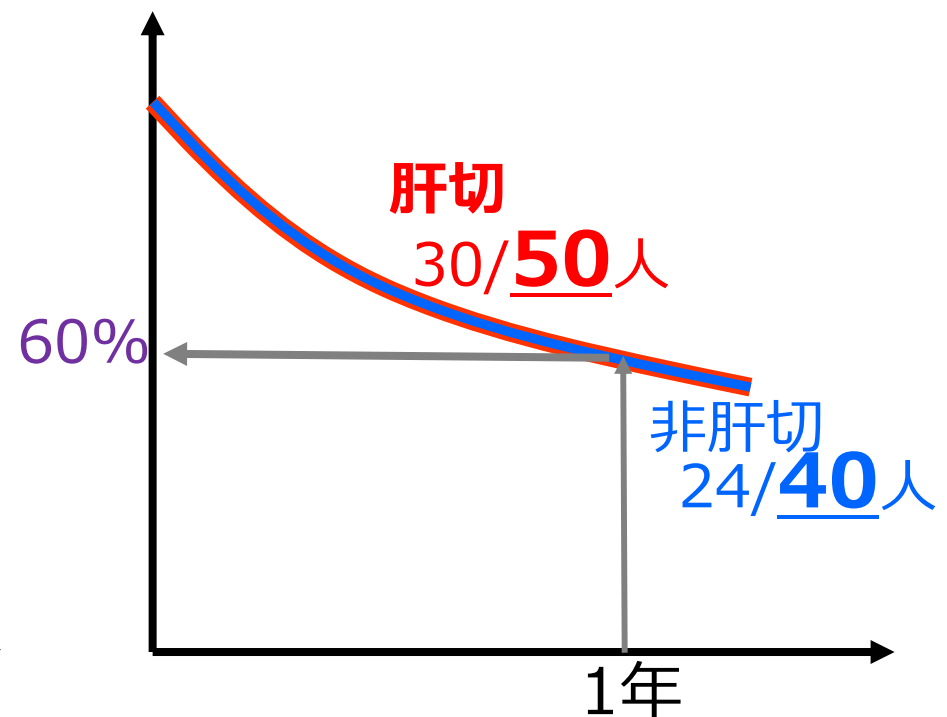
PR or CR
(腫瘍が小さくなった)

生存割合



SD or PD
(腫瘍が小さくならなかった)

生存割合



最良効果によらず**肝切**と**非肝切**の予後は変わらない

比較したいのは**肝切**と**非肝切**の違いだから

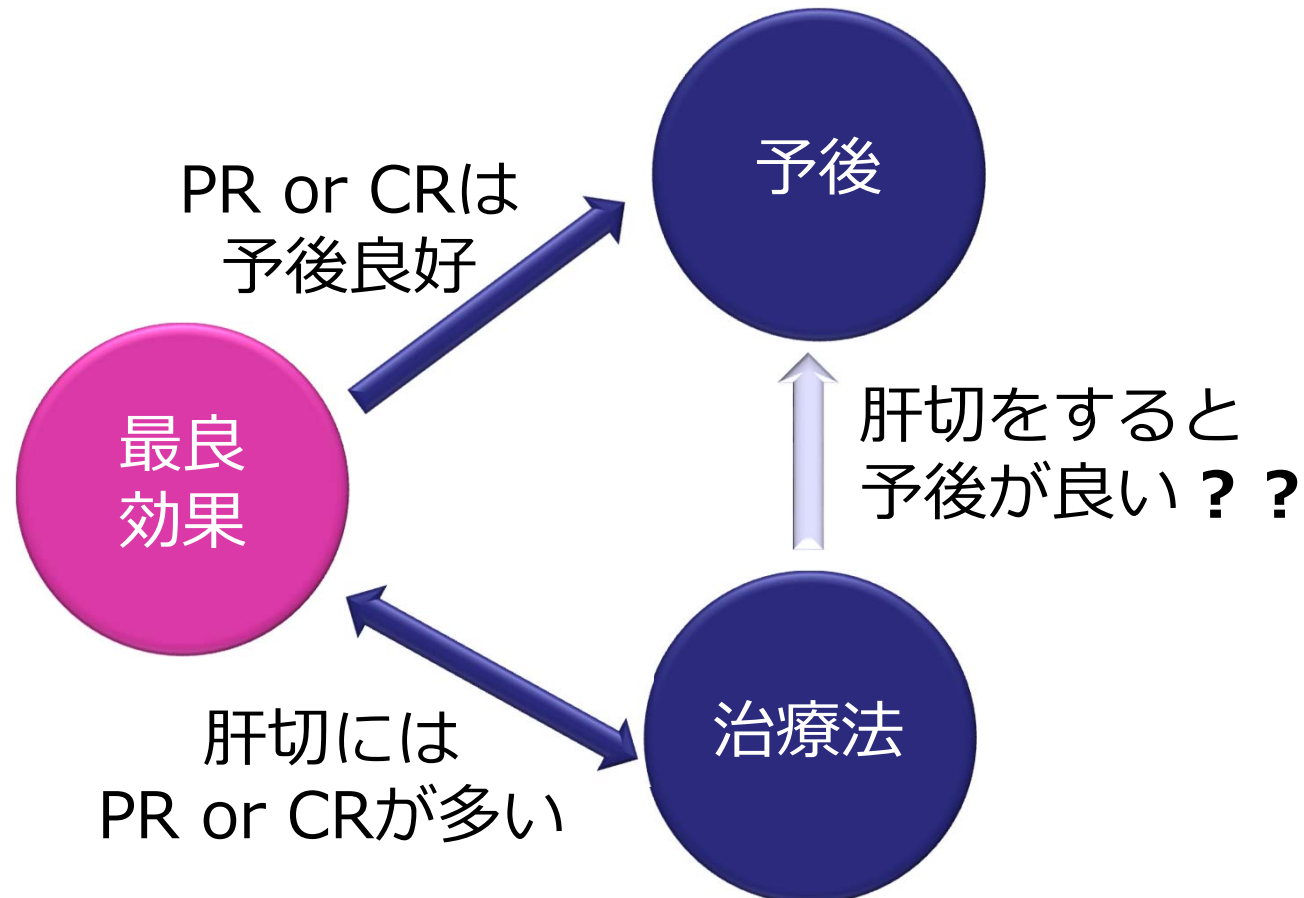
- 治療法以外の予後に影響する要因の条件が同じでなければ、“比較”にならない！！

治療法	PR or CR		SD or PD	合計
肝切	200人 (80%)	>>	50人	250人
非肝切	20人 (33.3%)	<<	40人	60人

- 肝切**は**非肝切**と比べ「**PR or CR**」の割合が高い
- 最良効果によって予後が異なる(PRorCRは予後良)

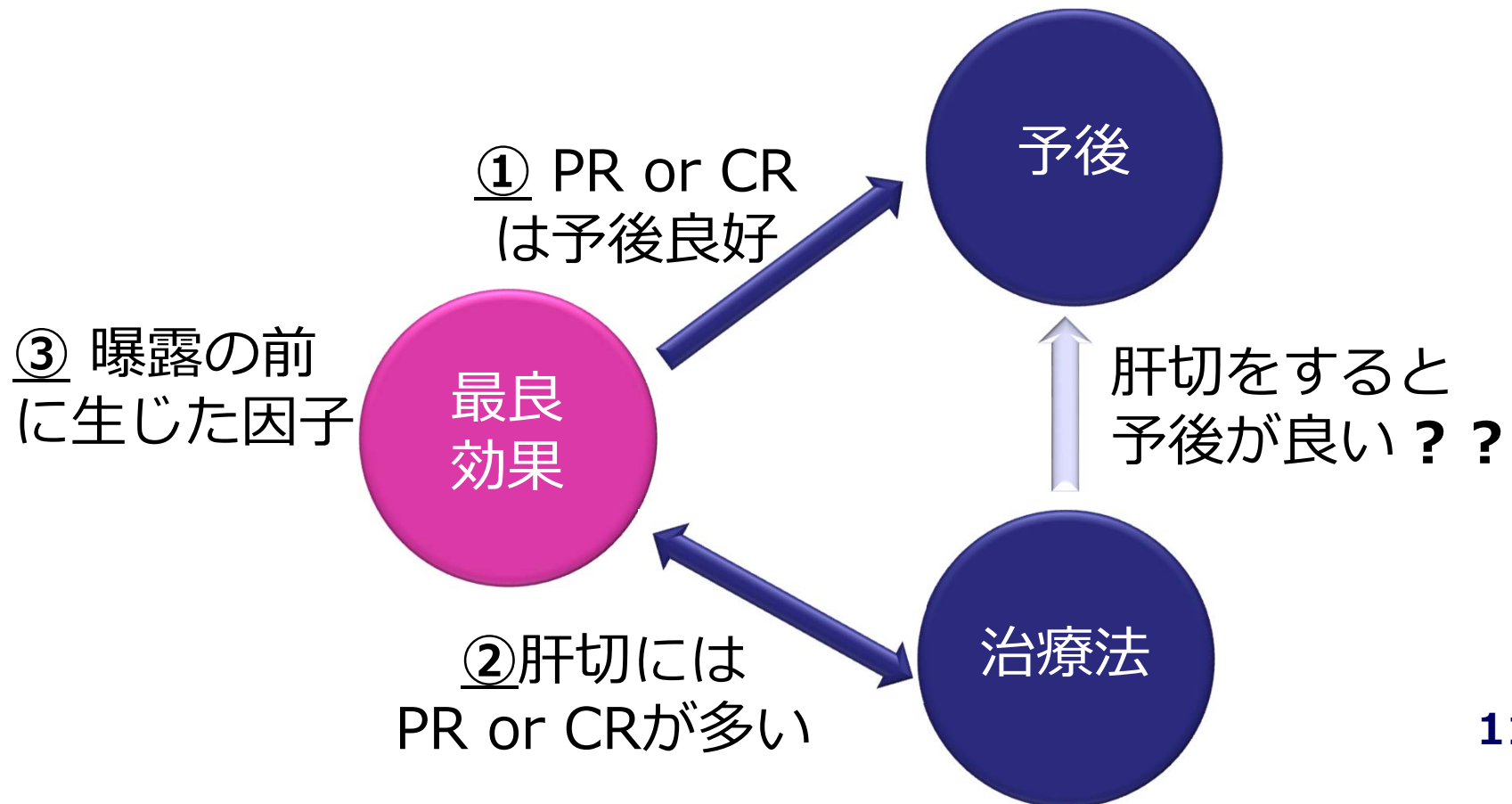
交絡についてのまとめ

- 治療法と予後に関連する第3の因子（最良効果）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(=最良効果)のことを**交絡因子**という

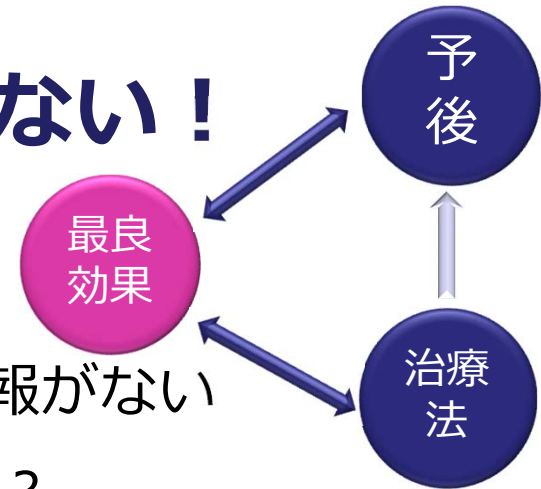


交絡の必要条件

- ① アウトカム(予後)と関係がある
- ② 曝露(治療法)と関係がある
- ③ 中間変数ではない (治療法の前に判明している事象)

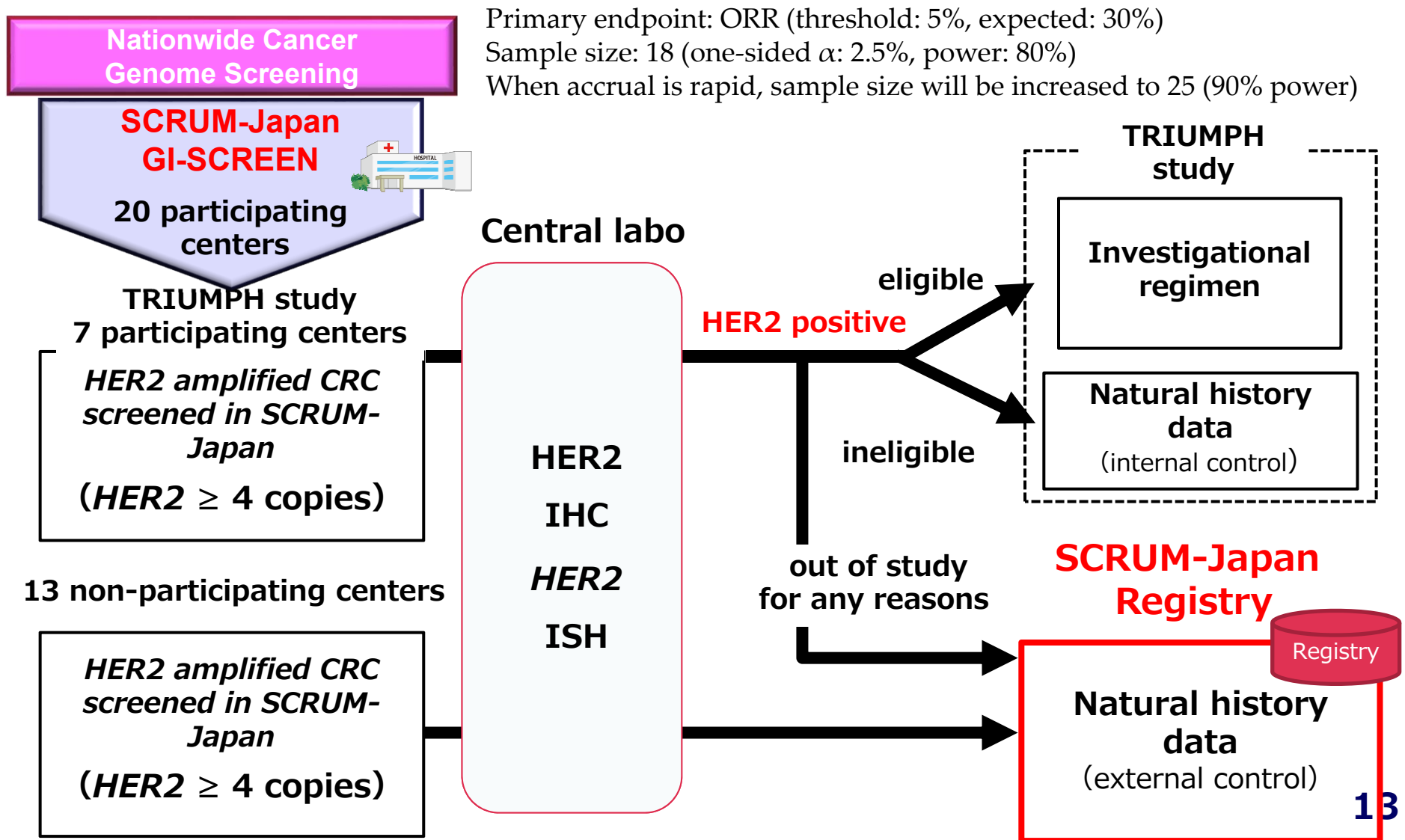


単群試験では、交絡が調整できない！



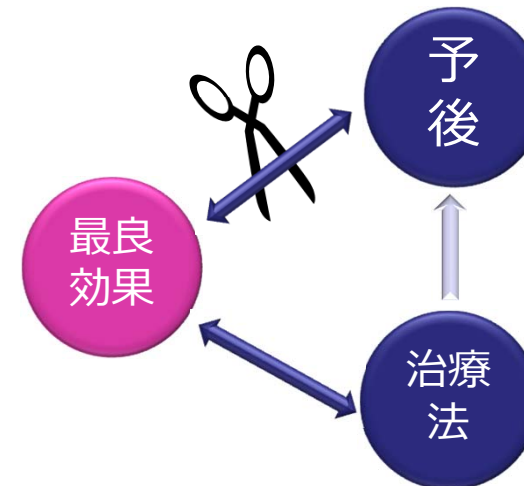
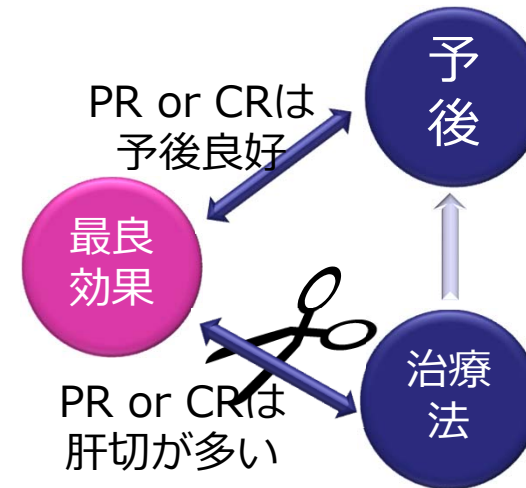
- ヒストリカルコントロールの交絡因子の情報がない
 - 薬Aと閾値の差は、治療効果？交絡による差？
 - phase 3に進んでも良いか、という判断には使えるが、薬Aが良い治療か、という判断は難しい
- 対照群を置けば、交絡を調整して治療効果を推定できる
 - 交絡が調整できれば、薬Aが良い治療か、という判断に使える

例 : TRIUMPH trial (登録中)



交絡調整のテクニック

- デザイン段階でのテクニック
 - 交絡因子と治療法の関連を除去
 - ランダム化
 - マッチング
- 解析段階でのテクニック
 - 交絡因子と予後の関連を除去
 - 多変量解析
 - 交絡因子と治療法の関連を除去
 - サブグループ解析
 - 層別解析
 - 傾向スコアを用いた解析



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）
 - PS（全身状態）
 - その他（未知の因子も含めて）

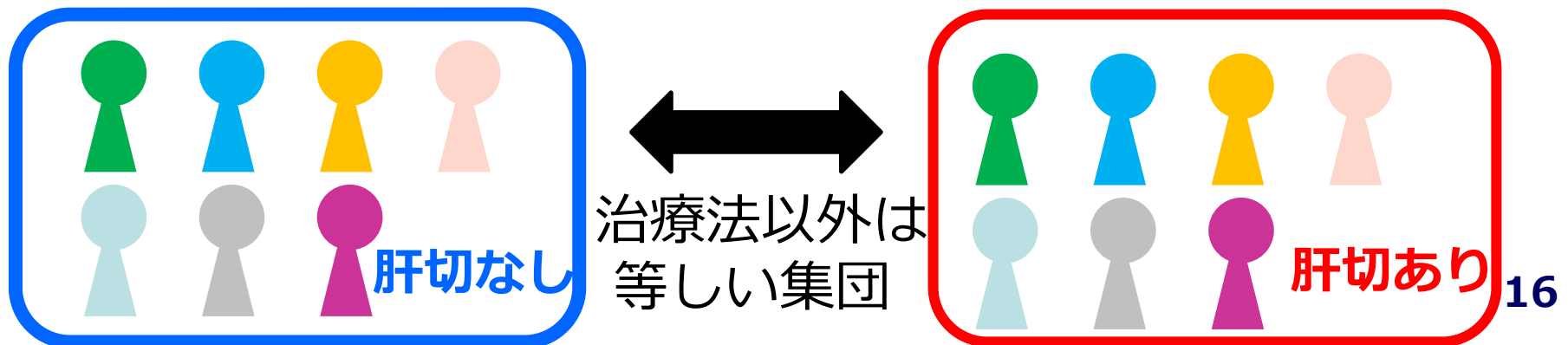
因子がたくさんある・未知の因子があるために
全てを考慮できない



治療法を**ランダム**に決める

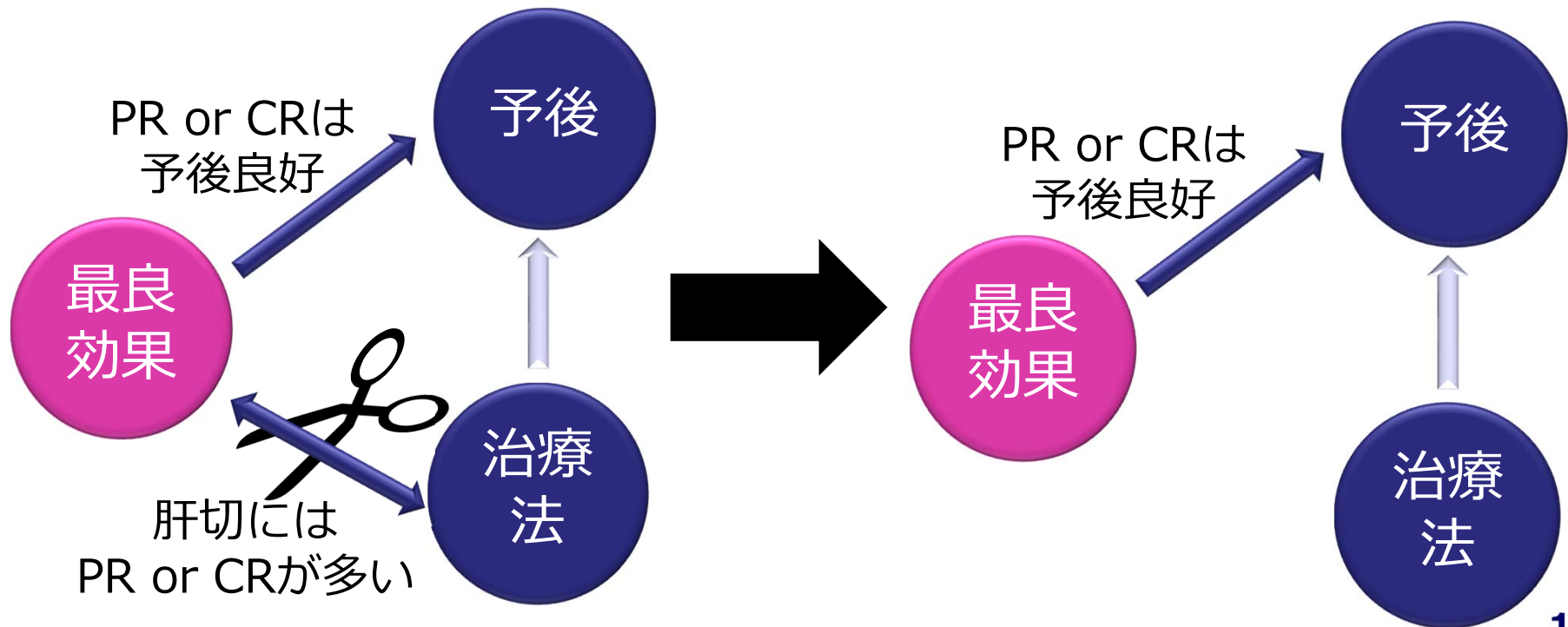
ランダム化 randomization

- 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
 - 状態の良い患者は新薬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



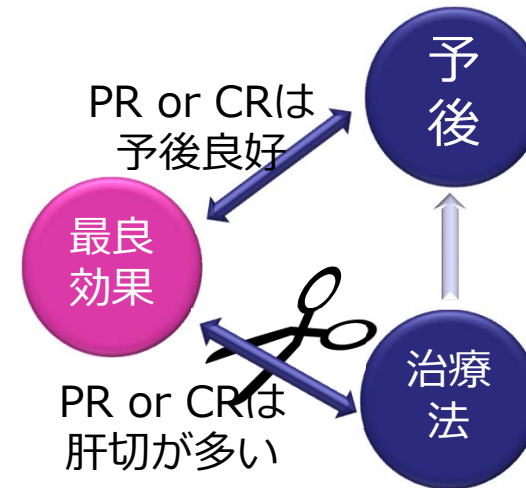
ランダム化の意義

- 治療法と最良効果の関連を除去できる
 - 最良効果による交絡は除去されるので、治療法と予後の関係を評価できる
 - 注：予後と最良効果の関係は残ったまま



交絡調整のテクニック

- デザイン段階でのテクニック
 - 交絡因子と治療法の関連を除去
 - ランダム化
 - マッチング
- 解析段階でのテクニック
 - 交絡因子と予後の関連を除去
 - 多変量解析
 - 交絡因子と治療法の関連を除去
 - サブグループ解析
 - 層別解析
 - 傾向スコアを用いた解析

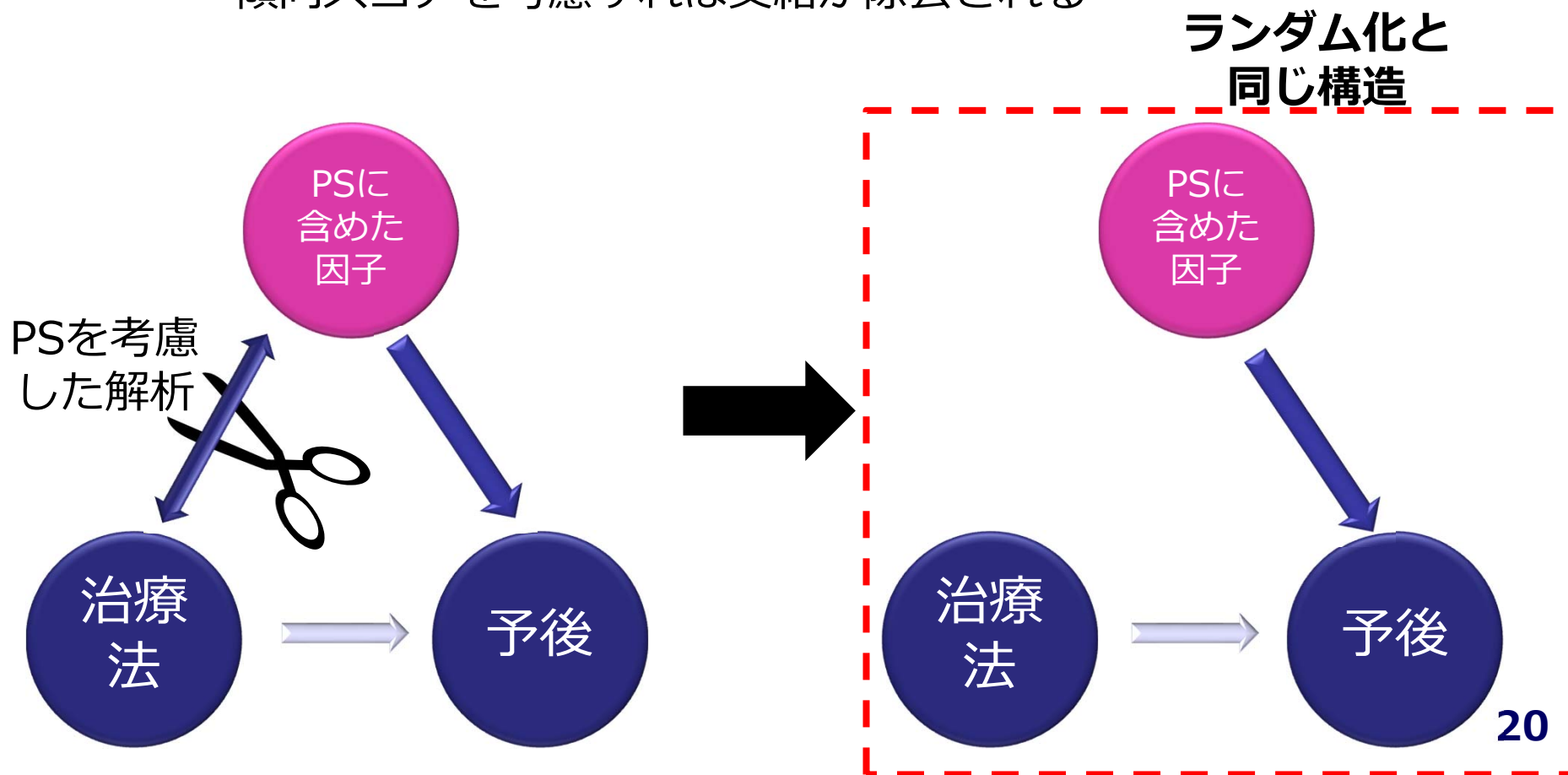


傾向スコア (Propensity Score)

- 傾向スコアとは
 - ある背景因子を持つ対象が、一方の治療法に割り付けられる確率
 - 曝露要約スコア
 - 取り得る値の範囲：0~1
- 治療法の選択に係る複数の因子を要約した1つのスコアにする
 - 1つ1つの背景因子は違っても、傾向スコアが同じであれば、一方の治療法への割り付けられやすさは同じとみなせる
 - 傾向スコアが同じ対象を集団で考えれば、治療群間で背景因子が平均的に同じであることが期待できる
 - ある意味ランダム化と同様の状況が作り出せる

傾向スコアで調整した比較の意義

- 治療法と傾向スコア[PS] (≡背景因子) の関連を除去できる
 - 交絡を除去する構造がランダム化と同じになる
 - 傾向スコアを考慮すれば交絡が除去される



傾向スコアを用いた解析法

- 2段階の手順で行う

① 傾向スコアの推定

- 患者ごとの傾向スコアを推定する

② 傾向スコアを用いて調整した解析

- マッチング
- 層別解析
- 多変量解析
- IPTW

①患者ごとの傾向スコア(PS)の推定

- 傾向スコア (PS:Propensity score)は新薬を投与される確率で、 $0 < PS < 1$ の値を取る
 - ≈ 0 : 新薬を投与される確率はほとんどない
 - ≈ 1 : ほぼ間違いなく新薬を投与される

- 通常はロジスティック回帰モデルを用いて推定する

- 新薬を投与されれば1、投与されなければ0として、

$$\log\left(\frac{PS}{1-PS}\right) = \beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{stage} + \beta_3 \cdot \text{性別} + \dots + \beta_i \cdot \text{併存症}$$



$$PS = \frac{\exp(\beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{stage} + \beta_3 \cdot \text{性別} + \dots + \beta_i \cdot \text{併存症})}{1 + \exp(\beta_1 \cdot \text{年齢} + \beta_2 \cdot \text{stage} + \beta_3 \cdot \text{性別} + \dots + \beta_i \cdot \text{併存症})}$$

- 傾向スコアは「予後(アウトカム)」の情報を用いずに推定可

傾向スコアを用いた解析法

- 2段階の手順で行う

① 傾向スコアの推定

- 患者ごとの傾向スコアを推定する

② 傾向スコアを用いて調整した解析

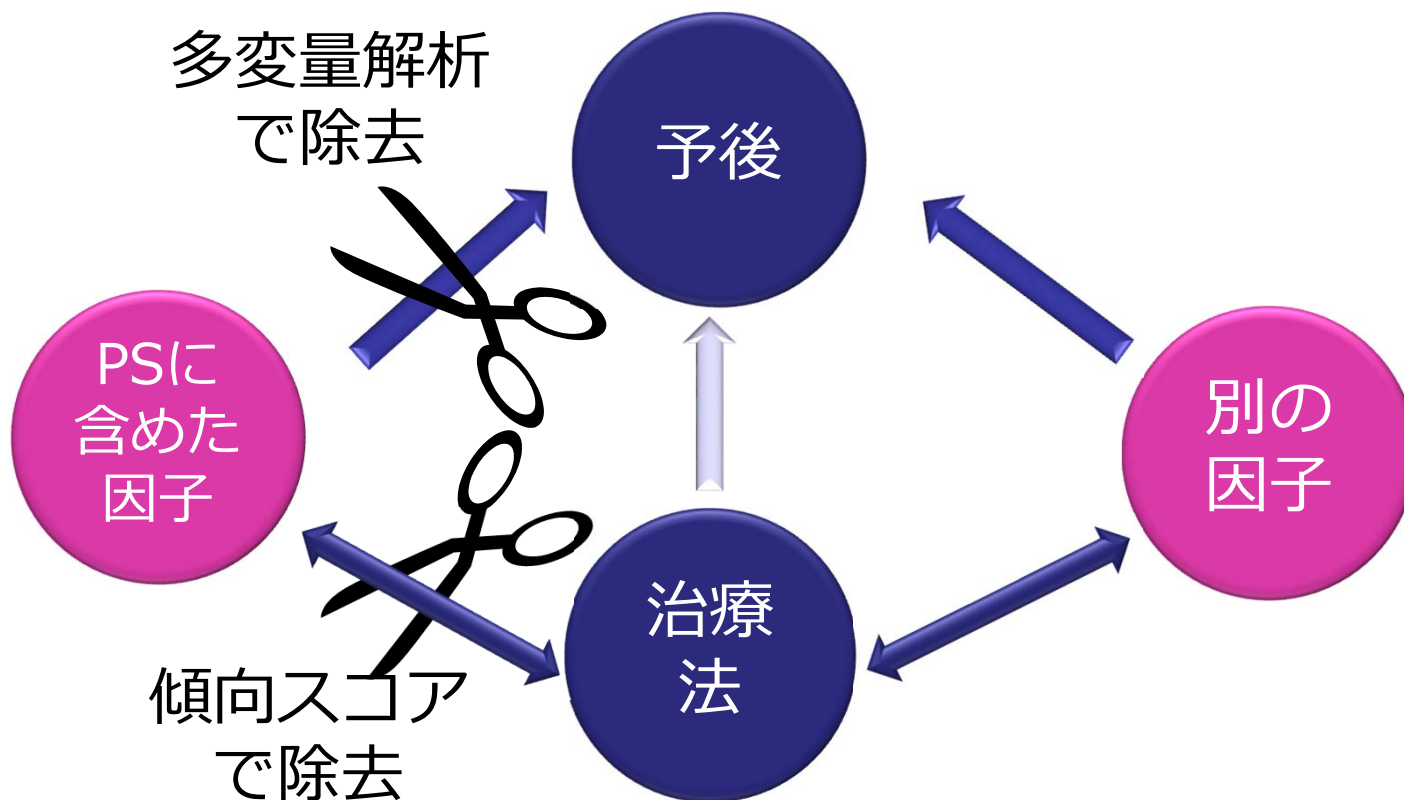
- マッチング
- 層別解析
- 多変量解析
- IPTW

②傾向スコアを用いて調整した解析

- マッチング
 - 各治療群から傾向スコアがほぼ同じ対象をペアにする
- 層別解析
 - 傾向スコアを層に分けて、層ごとに治療効果を算出し、重み付き平均して統合する
- 多変量解析
 - モデル(Cox回帰モデルやロジスティック回帰モデル)に、治療法と傾向スコアを共変量として入れた解析
- IPTW (Inverse Probability of Treatment Weighting)
 - 患者毎の傾向スコア (PS) の逆数で重み付けした解析

傾向スコアの欠点

- 傾向スコア(PS)の解析で調整していない因子の影響が除去できない=ランダム化の代わりにはならない
 - 調整していない因子が影響していたら、バイアスが入ったまま



対照群としてのレジストリの活用

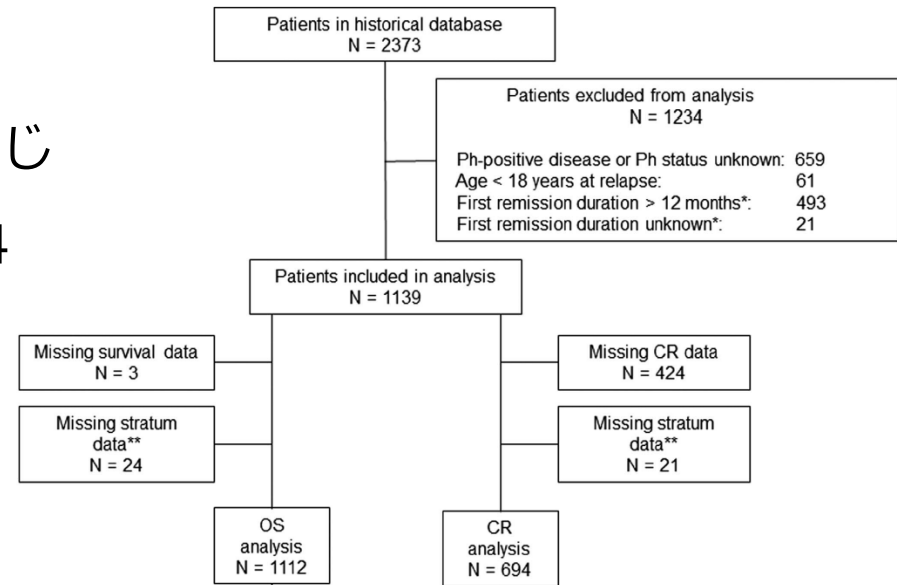
- 交絡調整のテクニックとして最強の方法はランダム化
- 現実的には、ランダム化が困難で、解析段階で交絡を調整するしかない状況も多い
 - 希少フラクションを対象とするため症例数が少ない
 - 倫理的な問題（例：喫煙と非喫煙のランダム化）
- 解析段階で交絡を調整するための対照群に、レジストリが活用できないか？

レジストリを承認申請に活用した事例

- 成人前駆B細胞急性リンパ性白血病に対するblinatumomab
 - 2本のphase2でaccelerated approval @FDA (2014)
 - primary endpointは寛解割合 (%CR/CRh)
 - 206 trial (2010~, 20mo): 寛解割合 = 69% (n=36) > 閾値10%
 - 211 trial (2012~, 21mo): 寛解割合 = 43% (n=189) > **閾値30%**
 - 2013年までにfast track, orphan drug, breakthrough therapyに指定
- 規制当局 (FDA/EMA) からの指摘
 - pivotal試験211 trialの閾値30%は実在のデータを反映しているか？
 - メタアナリシス等の手法で閾値30%の妥当性について検討すべき
 - regular (full) approvalにはOSの比較結果が不可欠

過去対照データを用いた検討 [1/2]

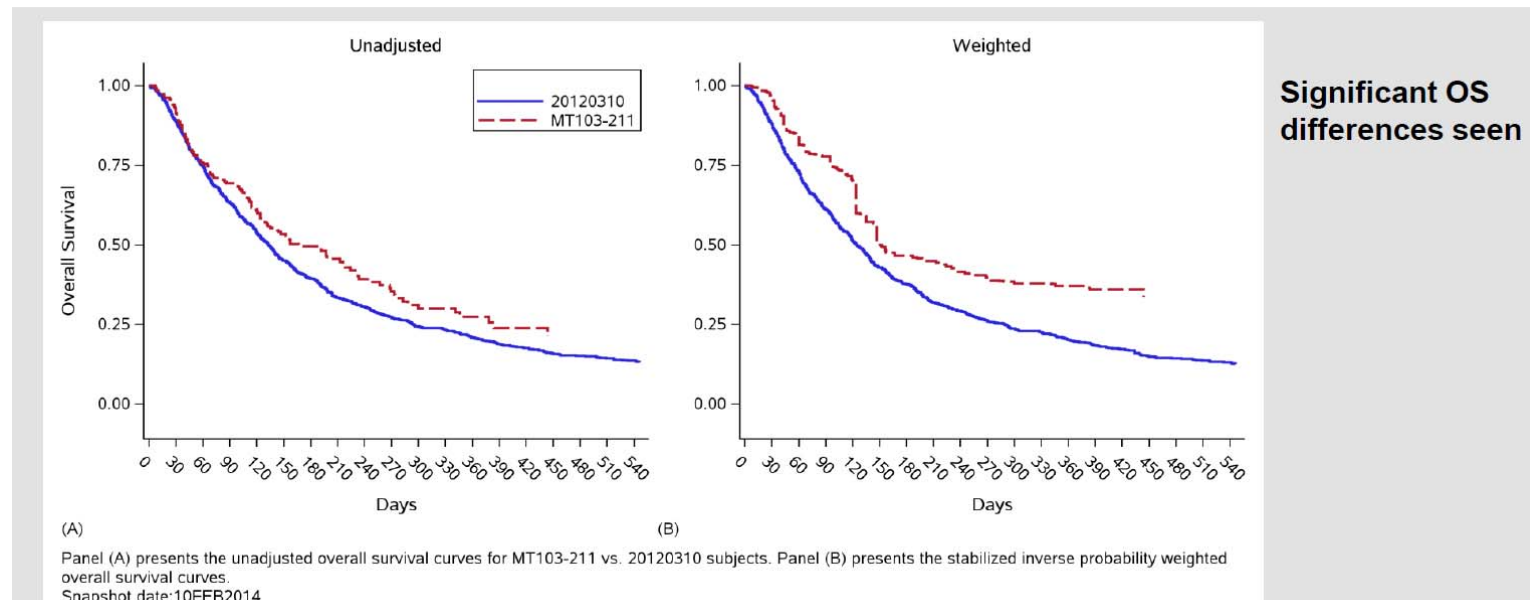
- 収集データ : n=1139
 - 適格規準は2本のphase2と同じ
 - 寛解割合の解析対象 : n=694
(400例以上が欠測)
 - OSの解析対象 : n=1112



- 閾値の妥当性
 - 患者背景を211 trialと合わせた寛解割合=24% [20%-27%]
 - 値が大きく変わらなかったため、「閾値30%は妥当」と結論
- 傾向スコアで調整したOSの解析も実施

過去対照データを用いた検討 [2/2]

- 傾向スコア調整前（左図）、調整後



– HR by IPTW: 0.54 [95% CI, 0.40-0.73]

- この結果を基に、EMAでもconditional approval

レジストリを用いる場合の留意点：解析手法

- アウトカムや背景情報の欠測やデータの質の問題で、実際に解析に使えるデータは少ない可能性がある
 - blinatumomabの例では、 $n=1139$ のうち、寛解割合の解析に使用できたのはわずか $n=694$
- 症例数が少ない場合、傾向スコア等の複雑な解析はできないが、それでも対照群を置くべきか？
 - サブグループ解析等による、交絡についての簡単な検討は可能
 - “治療効果”の推測のためには、交絡が全く調整できない単群試験よりはマシ

レジストリを用いる場合の留意点：信頼性

- 臨床試験（特に治験）の中で収集した対照群のデータ
 - 治験であればGCP準拠で、前向きなデータなので、質が良い
 - データの原資料との一貫性や、完全性等が担保されている
- レジストリから抽出した対照群のデータ
 - 原資料との**一貫性が担保されない**場合がある
例：クリニックから集めたデータの原資料が存在するか？
 - 治験と同水準でのフォローアップはされていないため、**データの欠測や、バイアス**の恐れがある
- 薬事承認を目的としてレジストリデータを対照群として用いる場合、GCP準拠が求められる
 - 現実的には、GCP準拠は困難な場合が多いかもしれない

交絡についてのまとめ

- 予後に関係する背景因子が治療群間で偏ることで、見かけ上の関連が生じてしまう or 見かけ上関連が消えてしまう現象のことを交絡という
 - 単群試験では調整できない
- 交絡調整のテクニック
 - ランダム化
 - 傾向スコア
- ランダム化できない状況でも、レジストリ等を活用して対照群を置けば、ある程度は交絡が調整可能

バスケット試験

～一度に複数のがん種を対象とする新しいデザイン～

希少フラクションにおける治療開発アプローチ

- 外部データを対照群にする
- マスタープロトコール型臨床試験
- ベイズ流のデザイン・解析
- 登録促進のためのネットワーク構築
- 試験期間を長く設定
- サンプルサイズが小さくて済むような集団を対象にする
- α 、 β を緩める
- データの質を担保
- 観察研究の実施

Billingham L, Lancet Oncol 2016.

マスタープロトコール型臨床試験

- 従来、治療法/疾患ごとに独立に行われてきた臨床試験を、1つのマスタープロトコールの下でまとめた臨床試験

– バスケット試験

- 複数のがん種に対して、単一薬剤の有効性を評価
- 例：BRAV V600、SIGNATURE

– アンブレラ試験

- 1つのがん種に対して、複数薬剤の有効性を評価
- 例：BATTLE、NCI-MATCH

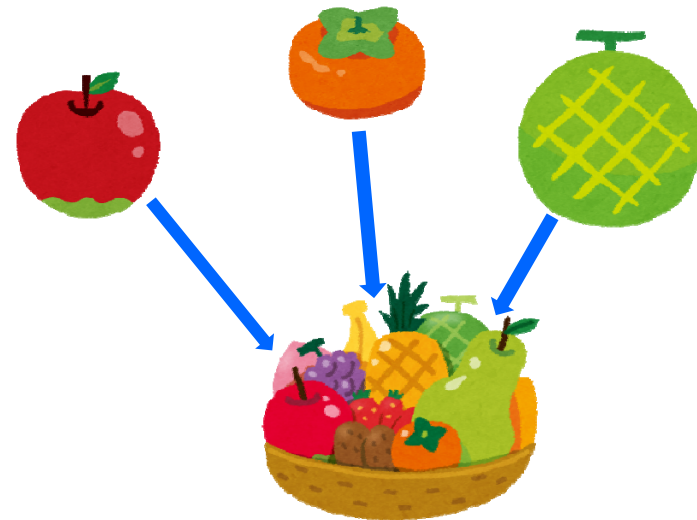
– プラットフォーム試験

- 基本的な構造はアンブレラ試験と同様
- 試験中に、新規薬剤群の追加や、評価済みの薬剤群の卒業
- 例：I-SPY 2

バスケット試験の目的と解析手法

- バスケット試験を行う目的
 - 単一の集団だけでは試験実施が困難な集団の**有効性探索**
 - 有効ながん種の**スクリーニング**
 - 探索的な意味合いが強い（検証的試験の実施例はない）

- 主な解析手法
 - 全がん種をまとめて解析
 - がん種ごとに解析
 - 階層ベイズモデル



どの解析手法がよい？

- 全がん種をまとめて解析すれば、がん種ごとのサンプルサイズは小さくて済む
 - 希少フラクションを対象にバスケット試験を実施する動機
- 全がん種をまとめて解析することが妥当でない場合
 - がん種ごとに**治療効果が異なる**ことが想定される
 - 仮に治療効果が同じだとしても、がん種によって有効/無効の判断規準（**閾値/期待値**）が異なる
 - 有効な**がん種**のスクリーニングが目的
- がん種ごとに解析しておけば無難だが、希少フラクションでサンプルサイズ確保が困難、という問題の解決策になっていない
 - **階層ベイズモデル**の適用

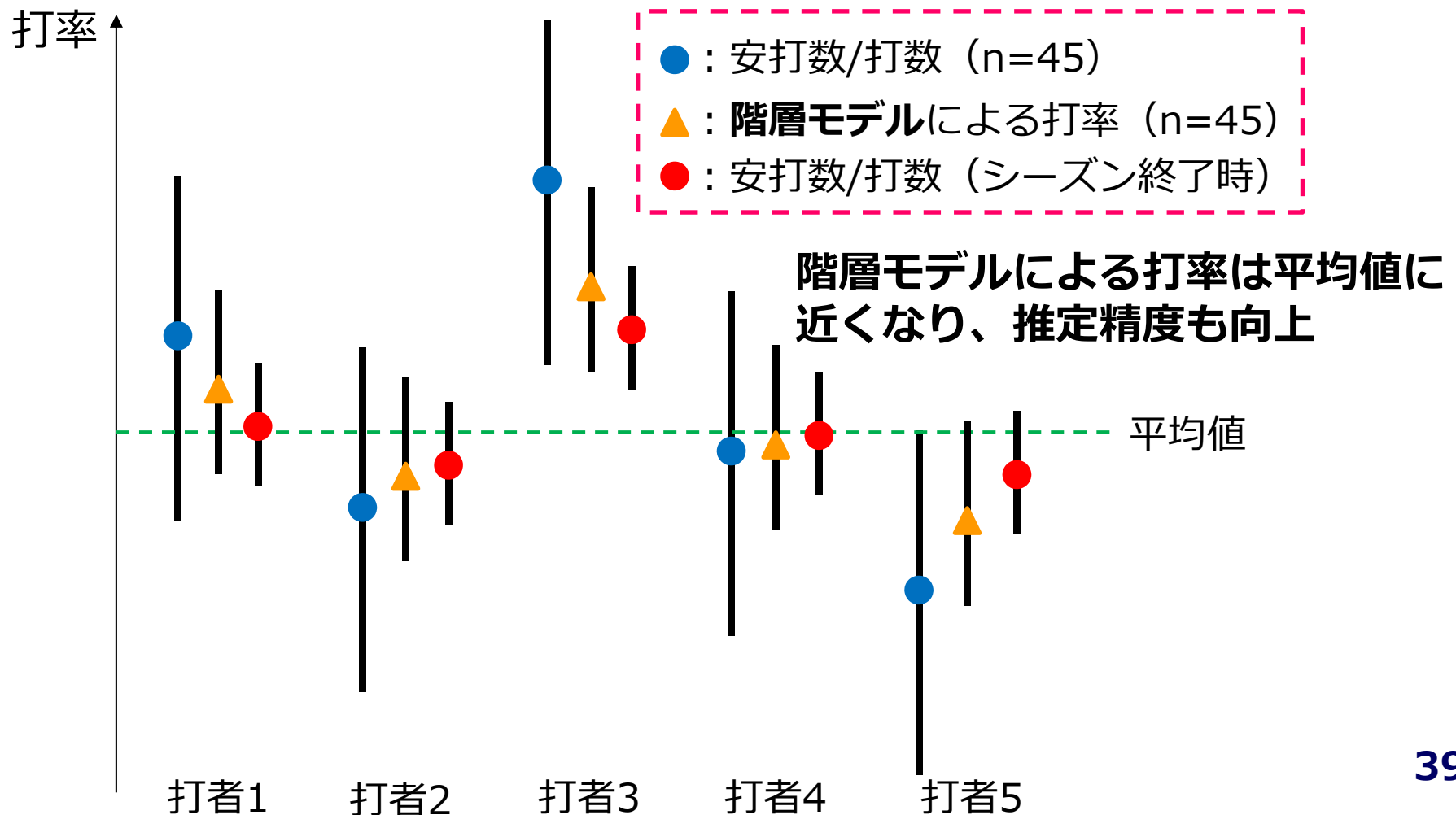
階層ベイズモデル

- あるがん種のアウトカムの推定に、他のがん種の情報を利用する
- 階層ベイズモデルのメリット
 - **がん種ごとに結論が出せる**
 - 推定精度・検出力を上げることができ、各がん種の**サンプルサイズ削減**が期待される

**がん種ごとのサンプルサイズ確保が困難な
希少フラクションでも、バスケット試験で結論が出せる**

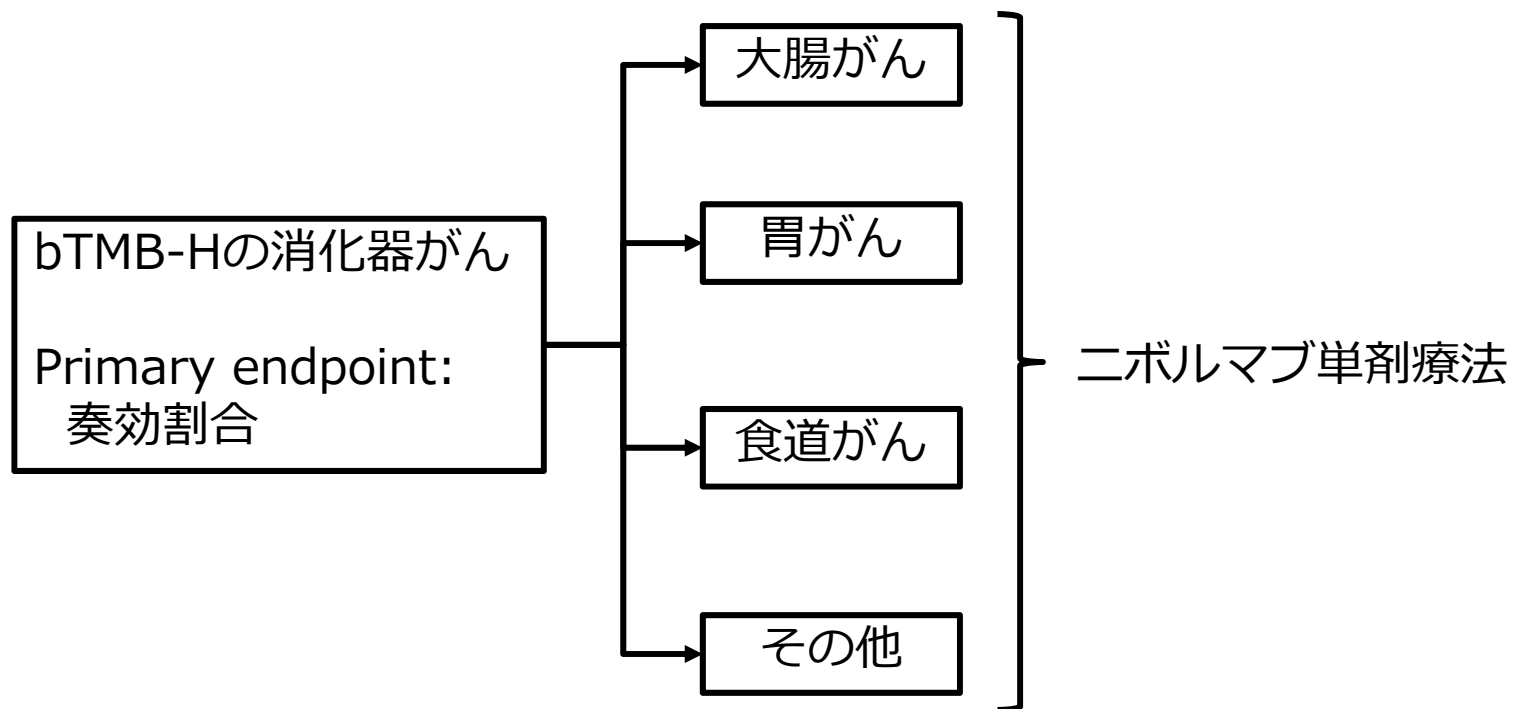
階層ベイズモデルのイメージ

- ある年のプロ野球選手の45打数での打率と、シーズン終了時（400打数以上）の打率（黒線は信頼区間）



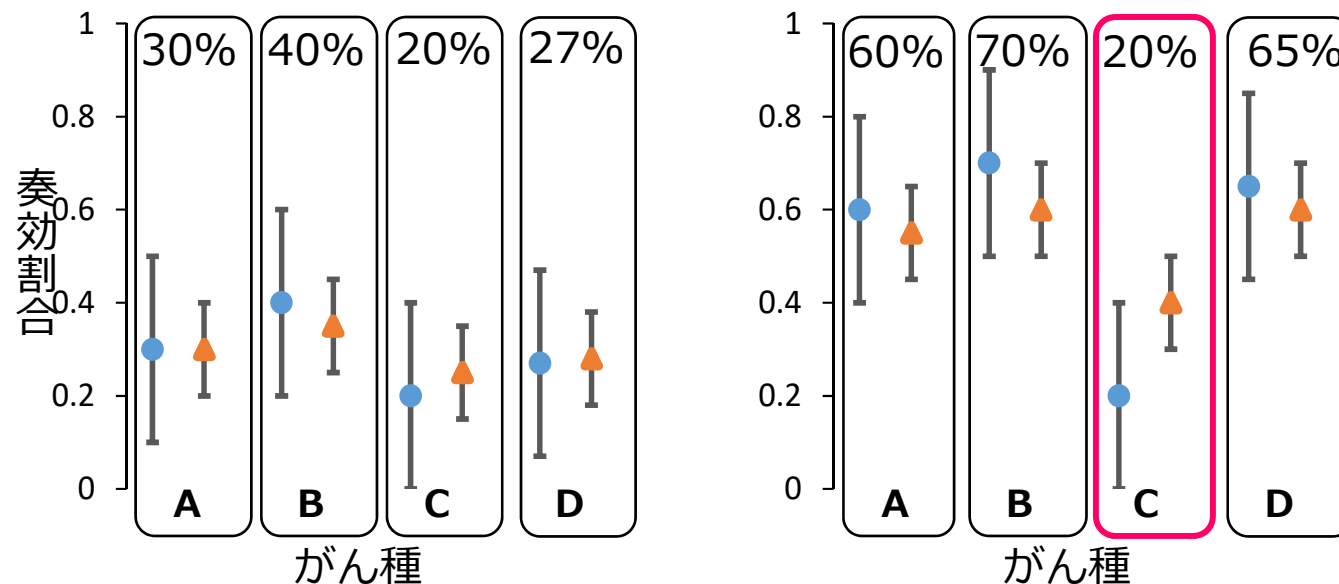
例 : TMB-H Basket

- TMB-Hの消化器がんを対象に、ニボルマブの有効性を検討
 - 消化器がんを4つのコホートに分類
 - 階層ベイズモデルを使用して、コホートごとに有効/無効の判断



階層ベイズモデルの留意点：強い仮定

- がん種間で効果が均一であるという強い仮定が必要
- 効果が均一（左図）、異質（右図）
 - 青：がん種ごとの解析、オレンジ：階層ベイズモデル



- 右図の状況で、がん種Cを誤って有効と判断する確率が上昇

階層ベイズモデルの利点・欠点

- 利点

- 推定精度・検出力の向上により**サンプルサイズを小さく**できる (?)
- **がん種ごと**に結論が出る
 - 希少フラクションにおいて、単一のがん種だけのサンプルサイズ確保が困難でも、各がん種について結論が下せる

- 欠点

- 複雑な解析法で、デザインパラメータの慎重な統計学的検討が必要
 - モデルの構造や判断規準をどう設定するか
- **治療効果のがん種によって異なると、誤った結論を導く危険性**がある
 - 無効ながん種の結果が、有効ながん種の結果に引っ張られて、誤って有効と判断してしまう (α エラー)
 - α エラーを制御する方法は存在するが、シンプルにがん種ごとに解析する場合と比べて検出力が変わらなくなるという報告もある
Freidlin, B. and E. L. Korn (2013). Clin Cancer Res 19(6): 1326-1334.

バスケット試験のまとめ

- 希少フラクションに対する治療開発アプローチの1つとして バスケット試験がある
- バスケット試験では、複数のがん種に対して、単一薬剤の有効性を評価
- 階層ベイズモデルによって、がん種ごとに結論を出し、がん種ごとのサンプルサイズを小さくできる
 - ただし、がん種間で効果が均一という強い仮定が必要