

がんを対象とした 単群Phase II試験の ベイズ流デザイン

国立がん研究センター 研究支援センター
生物統計部
佐立 峻

目次

1. がんを対象とした単群 Phase II 試験
2. 復習：頻度流の単群 Phase II デザイン
3. そもそも頻度流とは？ベイズ流とは？
4. ベイズ流の考え方に必要な事前知識
5. がんを対象としたベイズ流単群 Phase II 試験

典型的な単群 Phase II 試験

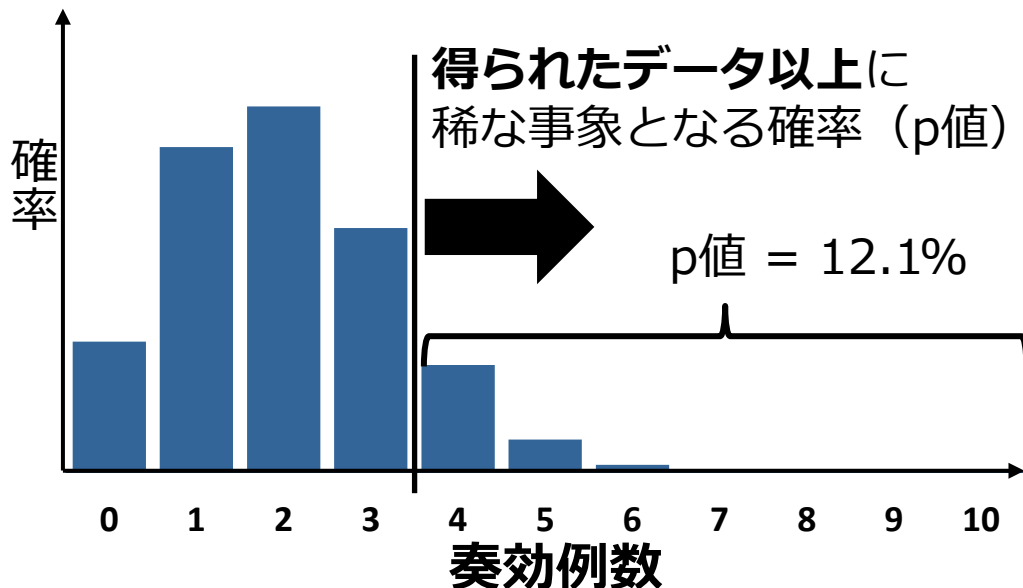
- 対象：がん種を限定
- 登録数：20-60 例
- デザイン：単群のオープンラベル（非盲検）
- Endpoint：RECIST による奏効割合
 - 長期ではなく短期的なエンドポイント
 - 分子標的薬の発展で時点 PFSを用いることもある
 - ただし多くの臨床試験では奏効割合が使われているのは事実
 - Phase III へ進むか否かをより早く決める
 - 生物学的な抗腫瘍活性の存在を確認する
 - 有効性によるスクリーニング

Phase II 試験の意思決定に必要な知識

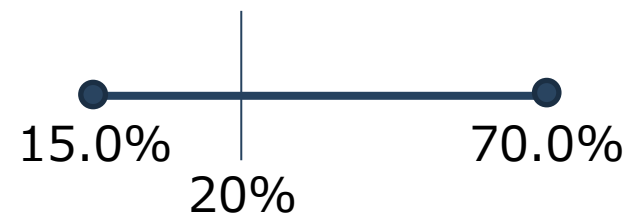
- 閾値奏効割合
 - これ以下の値であれば開発を中止を考慮する値
 - 「帰無仮説（効果が無い）」に対応
 - 標準治療と同等または少し低い値を設定することが多い
- 期待奏効割合
 - これ以上であれば開発を継続を考慮する値
 - 標準治療から改善が見込める高い値を設定することが多い
- 閾値と期待の差が大きいほど必要な症例数は少ない
 - 意志決定に直接関係するのは閾値奏効割合
 - 実際は試験の実施可能性から決めることもある

復習：頻度流の単群 Phase II デザイン

- 有意水準と p 値による意思決定
 - 帰無仮説の下で得られた結果以上の結果になる確率
- 仮想例
 - 閾値奏効割合：20%
 - 試験結果：10例中4例奏効（奏効割合40%）



信頼区間が検定と対応
片側5%の検定で有意 =
90%信頼区間の下限が20%以上



復習：頻度流の伝統的なデザイン

- 1-stage デザイン

- 途中で解析・登録停止することなく, 目標症例数まで一気に登録
- 目標症例数が集まった時点で解析してp値が有意水準以下で有意

- 2-stage デザイン

- Phase I しかしていない治療法を無効中止を考慮せず試験を続けるのは倫理的ではない
- 途中で1回解析して試験を中止するか, 継続するか意思決定

- 頻度流デザインの特徴

- 解析のタイミングと症例数が決まっている
- 症例数は α エラーと検出力によって決まることが多い

そもそも頻度流とは？ベイズ流とは？

- 統計学の2大流派
 - 何をバラつくものと見るかの違い
- 頻度流とベイズ流の違い
 - がんを対象とした単群Phase II試験の場合

	頻度流	ベイズ流
バラつくもの (意思決定に使用)	データ (奏効例数)	推定したいもの (奏効割合)
設定するもの (共通)	閾値奏効割合 期待奏効割合	
設定するもの (異なるもの)	有意水準 検出力	設定した奏効割合の曖昧さ 有効と判定する規準
必要な症例数	設定から算出された 固定症例数	(理論的には) 何例でも OK

何をバラつくものとするかの違いとは？

- 例. 対象が同じ試験を 2 回行ったとして…

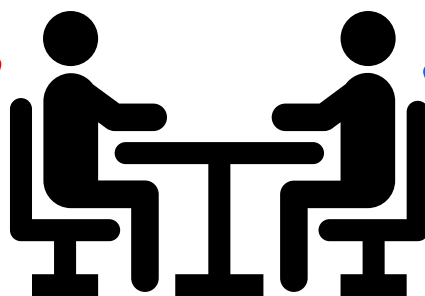
－ 結果

	試験 1	試験 2
治療患者数	10	10
奏効例数	4	6
奏効割合	40%	60%

頻度流で
バラつくもの
ベイズ流で
バラつくもの

結果が違うのは真の奏効割合
は同じだけれど**データがバラ
ついている**から！
コインを2回振っても常に表裏
が1回ずつ出る訳じゃない！
**帰無仮説の下で珍しい結果
であれば**有効な治療！
(帰無仮説棄却)

頻度流



ベイズ流

結果が違うのは**真の奏効割合
がバラついている**から！
測定できない患者背景まで考
えると常に真の奏効割合が同
じとは限らない！
得られた奏効割合が**閾値奏効
割合より高いことが確か**な
ら有効な治療！

意思決定に**奏効例数の分布**を使用

意思決定に**奏効割合の分布**を使用

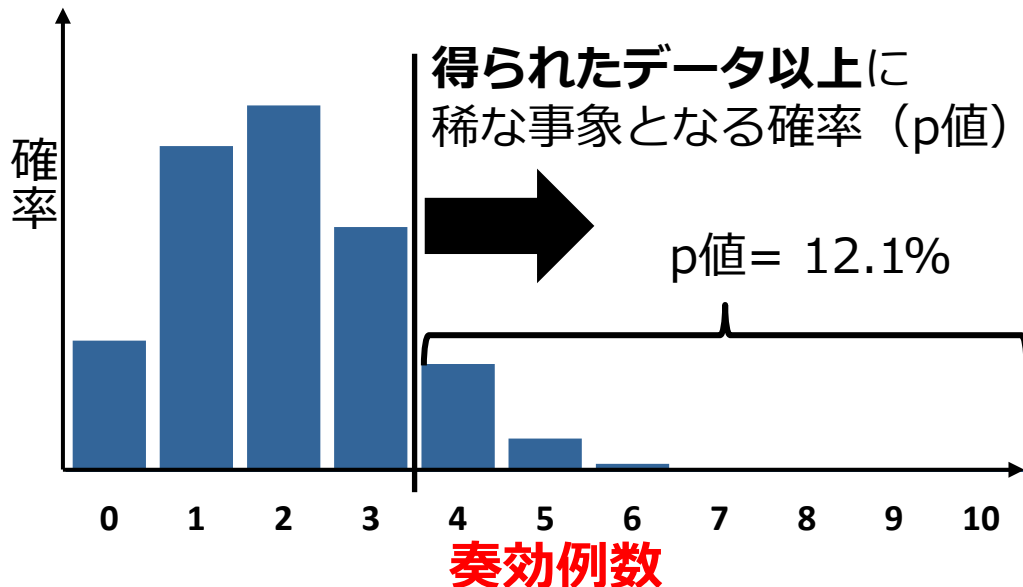
頻度流の有効性の判定

- 有意水準と p 値による意思決定
 - 帰無仮説の下で得られた結果以上の結果になる確率

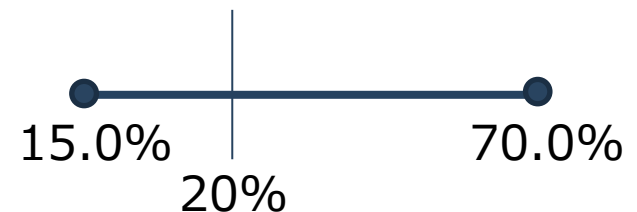
- 仮想例

- 閾値奏効割合：20%
- 試験結果：10例中4例奏効

試行回数10回，成功確率20%の
二項分布



信頼区間が検定と対応
片側5%の検定で有意 =
90%信頼区間の下限が20%以上



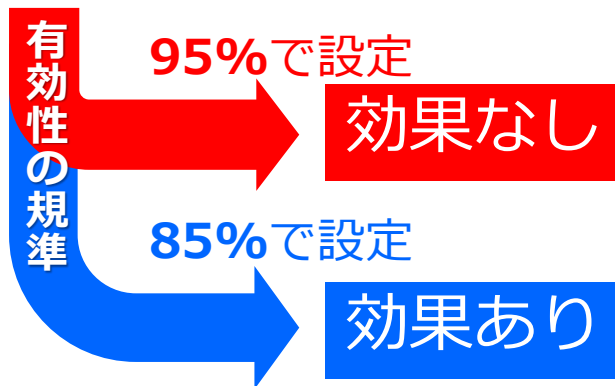
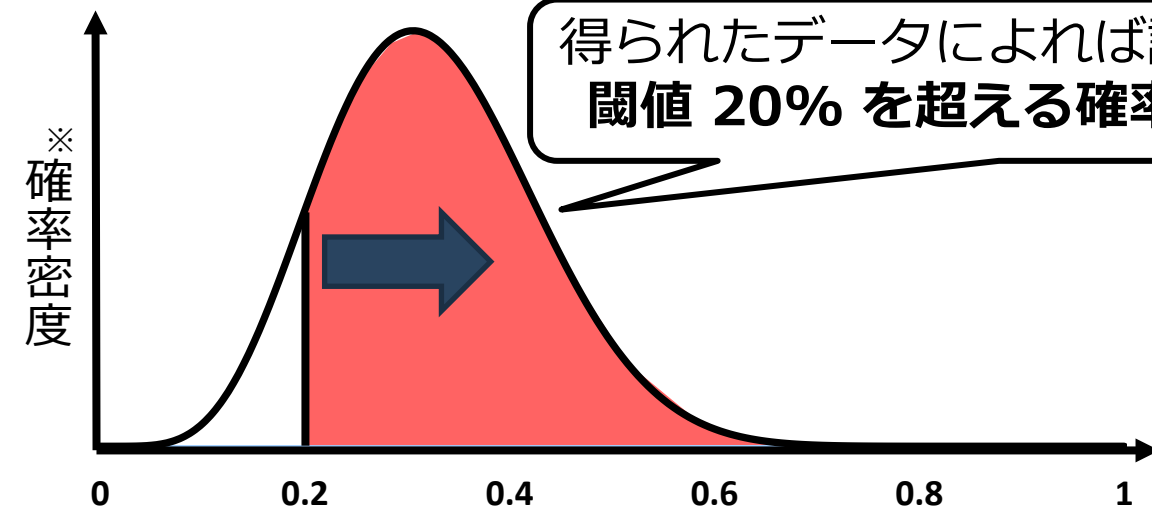
ベイズ流の有効性の判定

- 得られた結果から奏効割合が「**閾値より高くなる確率はどれくらいか**」で有効性判定
- 仮想例
 - 閾値奏効割合：20%
 - 試験結果：10例中4例奏効

設定した規準より**高ければ有効**
低ければ無効

試行回数10回，奏効例4例の
(どんな分布でも良いが一般的には) ベータ分布

得られたデータによれば試験薬の奏効割合が
閾値 20% を超える確率 (**赤色**) は 90%



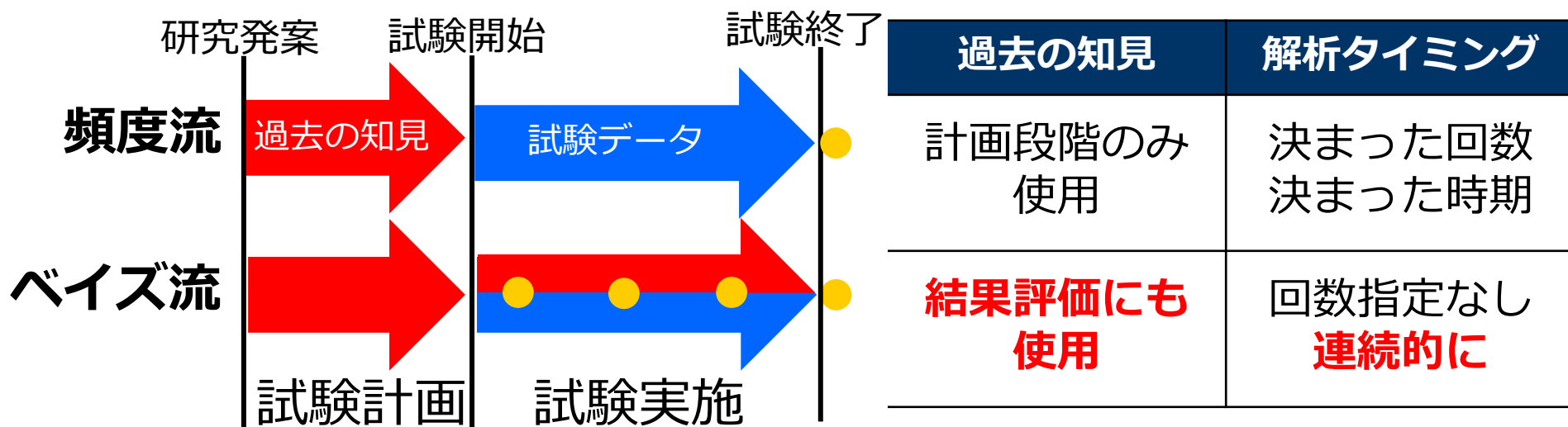
※スライド末の「統計の復習：
確率密度関数とは？」を参考

奏効割合

違うのは意思決定方法だけ？

• ベイズ流のメリット

- **過去の知見を活用する**ことで少ない症例数でも評価可能
- 連続的なモニタリングにより**柔軟な意思決定**が可能



• 最近ベイズ流が注目される理由

- 抗がん剤開発のメインターゲットが分子標的薬に
- 臨床試験の対象を遺伝子情報等で絞る必要→患者集積が困難

どうやって過去の知見と試験データを結びつける？

- 過去の知見と試験データをベイズの定理により統合
- 事前分布**：過去の知見を反映した事前情報の分布
 - 閾値奏効割合と期待奏効割合を**曖昧さ**をもって表現
 - 閾値奏効割合が20%であれば、平均 20%を中心にバラつく分布



Message 事前分布の設定に必要なもの

- どれぐらい情報を持たせるかを**症例数**によって分布を設定
 - 観察研究や過去の治験データからおおよその情報量を見積もる
 - 例：過去に200人ぐらいの治療経験があるならば200人分の情報を持ったもしくは少し割り引いた100人分の分布を設定

	治療	過去の治療経験	情報量	バラつき
閾値奏効割合	標準治療	ある	大	小
期待奏効割合	試験治療	ない	小	大

どんな事前分布を設定しますか？

- 国内Phase II臨床試験を計画したい
 - － 標準治療の臨床試験の結果

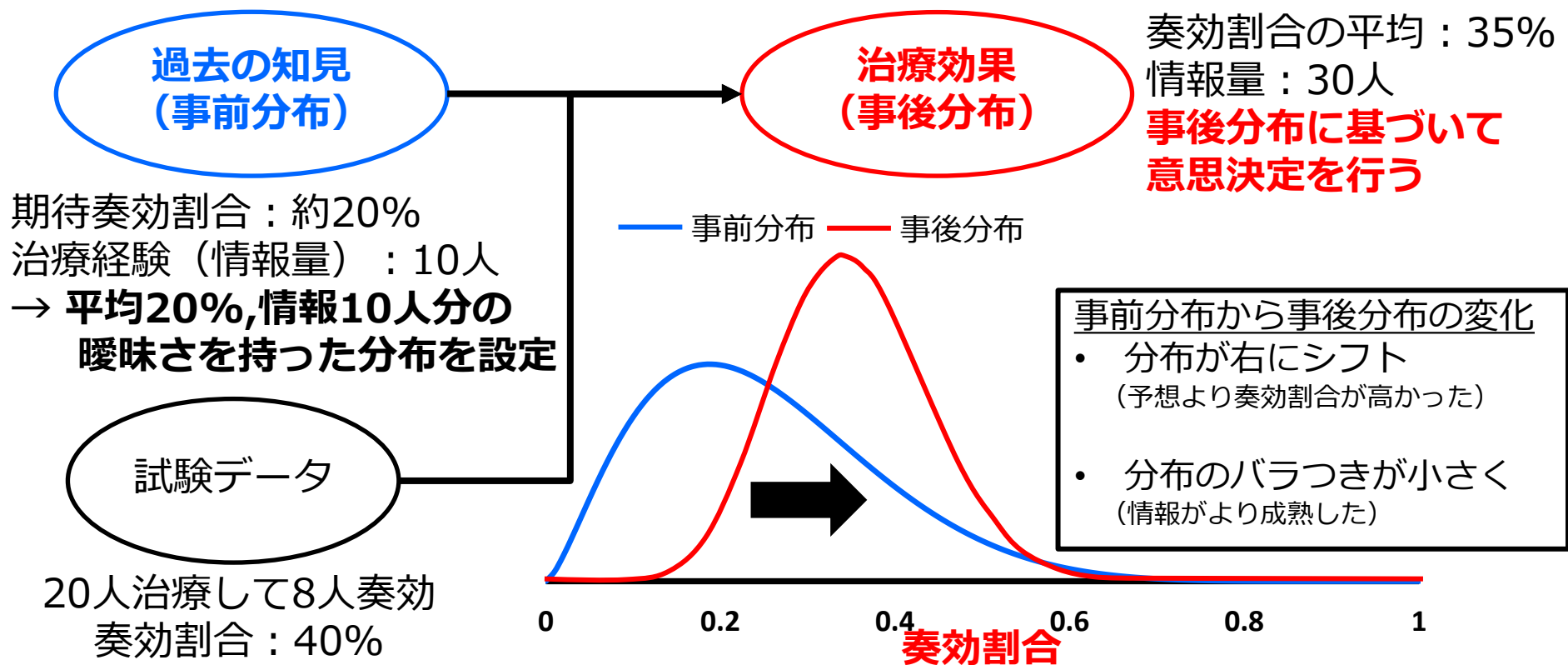
	試験開催国	症例数	奏効割合
試験1	国内	10	30%
試験2	国内	20	30%
試験3	海外	10	40%
試験4	海外	20	40%

- 標準治療の事前分布は？

- ①奏効割合：35%，情報量：60人 ②奏効割合：30%，情報量：30人
③奏効割合：40%，情報量：30人 ④奏効割合：30%，情報量：15人

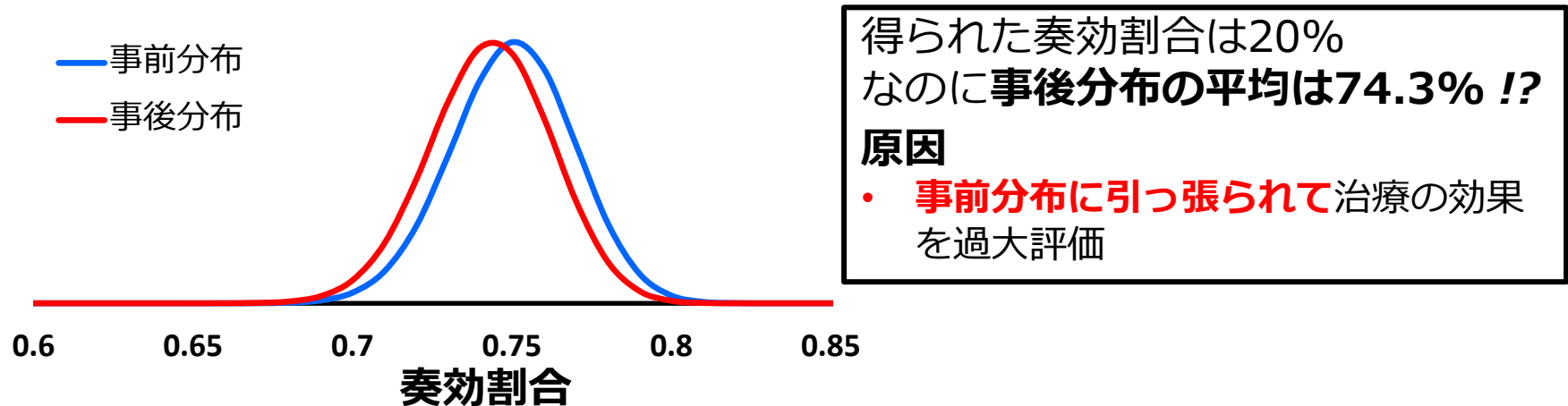
有効性の評価に使用する分布

- 事後分布：事前分布と試験データを統合した分布
 - 有効性の評価に使用する
- 事前分布と事後分布の関係



事前分布を設定するときの注意点

- 事前分布が効果の予測に影響する
 - 事前分布に情報を多く持たせると
 - 事前分布：500人分の情報で奏効割合を 75% と予想
 - 試験データ：10人治療して2人奏効（奏効割合20%）



- 特に期待奏効割合の事前分布の設定には注意
- 妥当な事前情報がない場合
 - 情報が無い事前分布を設定する（**無情報事前分布**）

ベイズ流単群 Phase II 試験の統計デザイン

- **事後分布に基づく方法** (Thall and Simon, 1994, Biometrics)
 - 閾値奏効割合と期待奏効割合両方の事前分布を設定
 - **期待奏効割合の事後分布と閾値奏効割合の事前分布を比較**して有効性を判定
 - NCCH 試験で適応試験あり
- **予測確率に基づく方法** (Lee and Liu, 2008, Clinical Trials)
 - 閾値奏効割合と期待奏効割合両方の事前分布を設定
 - **最終的な奏効例数の分布を予測し**, その下で治療が有効/無効となる予測確率を使って有効性を判定
 - NCCH 試験では過去に使用した試験はなし

今回説明するのは
事後分布に基づく方法

ベイズ流単群 Phase II 試験の進め方

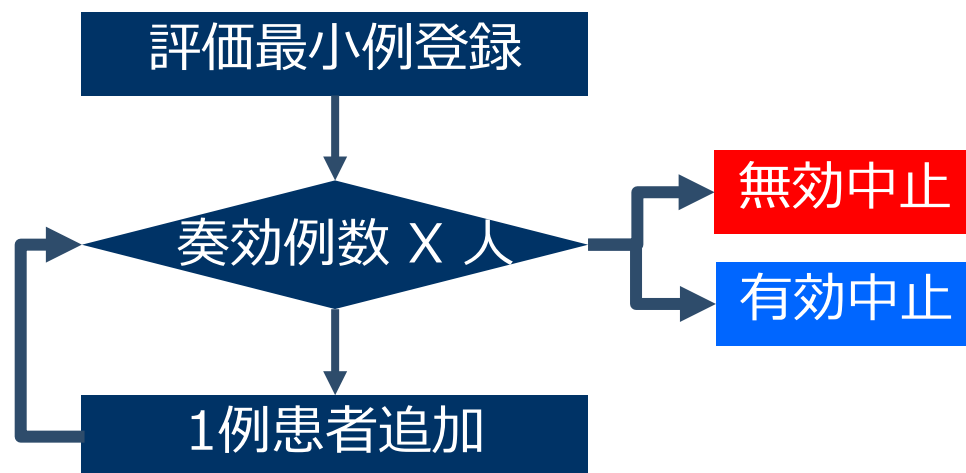
- 試験開始前に決めるもの
 - 閾値奏効割合と期待奏効割合の事前分布
 - 無効/有効と判定する規準
 - 無効中止のマージン
 - 症例数：何例から何例まで評価するか

これらを決めると
意思決定の表が作成できる

中止規準 (奏効例数)	症例数					
	10	11	12	13	14	15
有効中止 ≥	5	5	5	6	6	6
無効中止 ≤	0	0	1	1	2	2

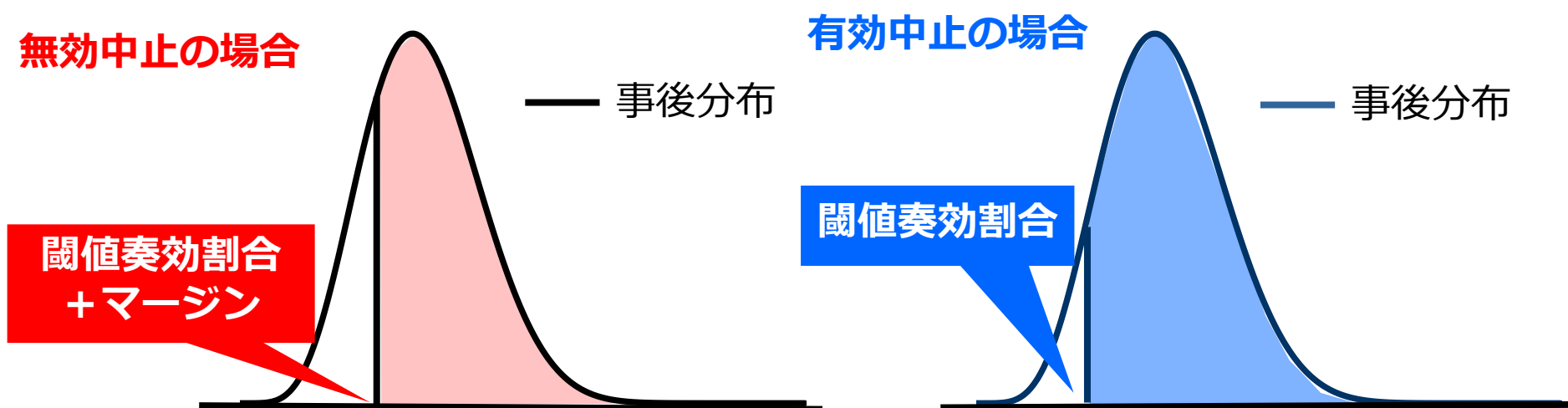
10例中奏効例数が5例以上なら有効中止
0例なら無効中止

表に従って1例ずつモニタリング



有効中止，無効中止の規準はどう決める？

- まず事後分布に基づいた中止規準を設定する
- 無効中止の規準 θ_L (0.01~0.05; 1%~5%が推奨)
(閾値奏効割合+マージン) 以上となる確率 $\leq \theta_L$
多くの場合マージンは0%
- 有効中止の規準 θ_U (0.95~0.99; 95%~99%が推奨)
閾値奏効割合以上となる確率 $\geq \theta_U$



有効中止，無効中止の奏効例数の決め方

- 治療例数 n 例のとき，奏効例数を $0 \sim n$ 例まで一例ずつ増やして無効/有効中止規準を満たす奏効例数を探索
 - 例．閾値奏効割合20%に設定し，10例のときの中止基準探索
 - $\theta_L = 5\%$ ，マージン 10%， $\theta_U = 95\%$

奏効例数が0例：無効中止

	奏効例数										
	0	1	2	3	4	5	6	7	8	9	10
20%+10%以上になる確率	0.7%	7.4%	25.5%	51.9%	76.2%	91.2%	97.7%	99.6%	99.9%	100%	100%
20%以上になる確率	3.2%	22.6%	53.3%	79.5%	93.4%	98.5%	99.8%	100%	100%	100%	100%

奏効例数が5例以上：有効中止

- 想定する症例数で同じ手順で規準を見つける
 - 10~15例で評価したい場合，同じように11例治療したとき，12例治療したとき…の中止規準というように探索する

ベイズ流デザインの適応例

NCCH1510 (OSCAR試験)

- 切除不能の明細胞肉腫 (CCS) , 胞巣状軟部肉腫 (ASPS) に対するニボルマブの医師主導治験
- 試験実施のハードル

頻度流で設定した症例数が試験終了までに集積可能か

- CCS, ASPSは希少がん
- 頻度流の設定では2年で18例の集積が必要
 - 全国骨・軟部腫瘍登録でCCSは8例/年, ASPSは10例/年
 - 頻度流では18例集積するまで試験を終了することができない
- ベイズ流デザインを適応することで実施可能となった

OSCAR 試験のデザイン

- 試験デザイン：事後分布に基づくベイズ流デザイン

- 閾値奏効割合：5%（情報量200）
- 期待奏効割合：30%（情報量2）
- 症例数：15~25例
- 有効中止の規準：95%
- 無効中止の規準：設けない

希少疾患であるため

- 少しでも有望であれば申請へ
- 早期で誤って無効と判定するリスクを避けた

- 意思決定表

中止規準 (奏効例数)	症例数										
	15	16	17	18	19	20	21	22	23	24	25
有効中止 ≥	3	3	3	3	3	3	3	4	4	4	4

- 登録期間終了時の症例数と奏効例数を意思決定表と比較

- 事前に規定された有効中止の症例数よりも多ければ有効
- 連続モニタリングではない方法での評価

ベイズ流デザイン適応時の注意点

- ベイズ流は恣意性の高い方法
 - 症例数, 事前分布, 無効/有効判定の規準...
 - どれも頻度流のようにGold standardは存在しない
- どのような頻度流の特性になるか検討する必要がある
 - 主に α エラーと検出力をシミュレーションを使って検証
 - FDA guidance : “Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials”

Because of the inherent flexibility in the design of a Bayesian clinical trial, a thorough evaluation of the operating characteristics should be part of the trial planning. This includes evaluation of:

- probability of erroneously approving an ineffective or unsafe device (type I error rate),
- probability of erroneously disapproving a safe and effective device (type II error rate),
- power (the converse of type II error rate: the probability of appropriately approving a safe and effective device),
- sample size distribution (and expected sample size),
- prior probability of claims for the device, and
- if applicable, probability of stopping at each interim look.

α エラーと検出力の確認（症例数を変更）

- OSCAR 試験の評価する症例数を変えると
 - 閾値奏効割合：5%（情報量200）
 - 期待奏効割合：30%（情報量2）
 - 有効中止の規準：95%

中止規準 (奏効例数)	症例数															
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
有効中止 $\geq$	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4

- α エラーと検出力

評価する症例数	α エラー	検出力
10 ～ 15	10.3 %	90.9 %
10 ～ 25	13.8 %	98.0 %
15 ～ 25	8.6 %	97.6 %
20 ～ 25	8.6 %	97.8 %

OSCAR試験の設定
希少疾患であるため
8.6%の α エラーは許容

α エラーと検出力の確認（期待奏効割合を変更）

- OSCAR 試験の期待奏効割合の事前分布を変えると
 - 閾値奏効割合：5%（情報量200）
 - 有効中止の規準：95%
 - 症例数：15~25例

• α エラーと検出力

頻度流と同じく期待奏効割合が大きいほど
 α エラーと検出力は大きくなる

平均	情報量	α エラー	検出力
20%	2	6.9 %	81.2 %
30%	2	8.6 %	97.6 %
35%	2	9.5 %	99.4 %
40%	2	10.4 %	99.9%
30%	10	60.1 %	99.9%
50%	1	8.6 %	99.9 %

注意 !!

事前情報が強すぎたため
効果が過大評価された

無情報事前分布

事前情報が無いことを仮定
感度解析によく使われる

ベイズ流デザインを検討したいときは

● 試験デザインの設定に注意が必要

- 想定される設定で α エラーと検出力等の頻度流の特性を評価する必要がある
- シミュレーションを用いた検討はかなり複雑
- **ベイズ流デザインを検討したい際は統計家まで！**



Message 現状ベイズ流が受け入れられる状況は**少ない**

- 事前分布の設定は**主観的な確率**に基づくもの（恣意性が高い）
- 希少疾患の試験等，頻度流で実施するハードルがある場合に限って用いられる傾向
- **まず頻度流**で検討をして，**困難であればベイズ流**を検討することを**推奨**

まとめ ~本セミナーで覚えておいていただきたいこと~

がんを対象とした単群Phase II試験のベイズ流デザイン

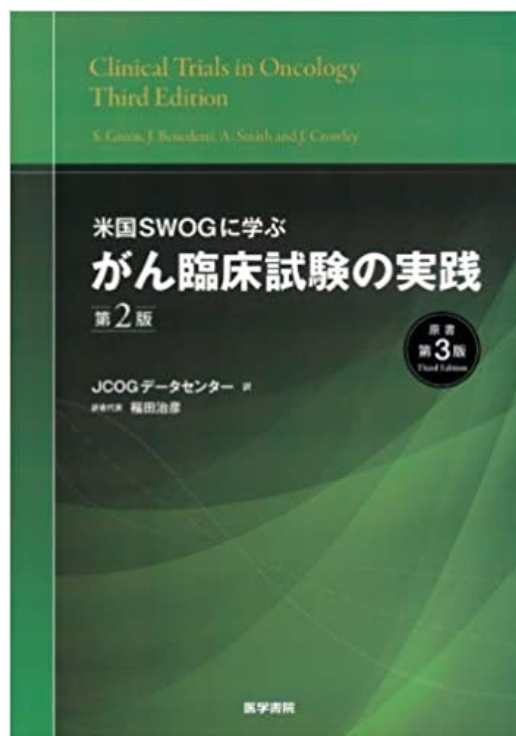
- 頻度流とベイズ流の違い
 - 意志決定のためにみるものが違う
- ベイズ流デザインに必要な基礎知識
 - 事前分布とは？事後分布とは？
 - 事前分布の設定方法
- ベイズ流のデザインを実施する際の注意点
 - あらゆる設定で頻度流の動作特性を確認すること
 - **頻度流での実施が厳しい場合に用いられる傾向**
 - まずは頻度流で検討し、実施困難であればベイズ流を検討する

参考図書

(数理を飛ばして)
ベイズ流の考え方



がんPhase II試験
全般

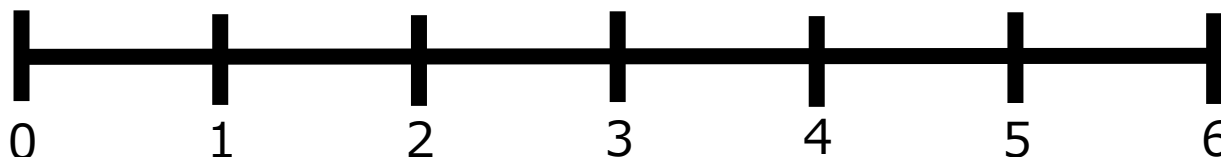


がんPhase II試験の
ベイズ流デザイン



統計の復習：確率密度関数とは？

- ある連続の値を取る X に関して, ある区間に X が収まる確率を計算するための関数 $f(X)$
 - X をとる確率でないことに注意 (x をとる確率は0)
 - X が $0 \sim 1$ に収まる確率は計算できる
- 例. $0 \sim 6$ の値を等しい確率でとる数直線
 - $0.00000000 \cdots 00001$ のような微小な値をとることもある
 - ある特定の値をとる確率：ほぼ0% ($= 1/\infty$)
 - $0 < x \leq 1$ の値をとる確率は計算できる： $1/6$

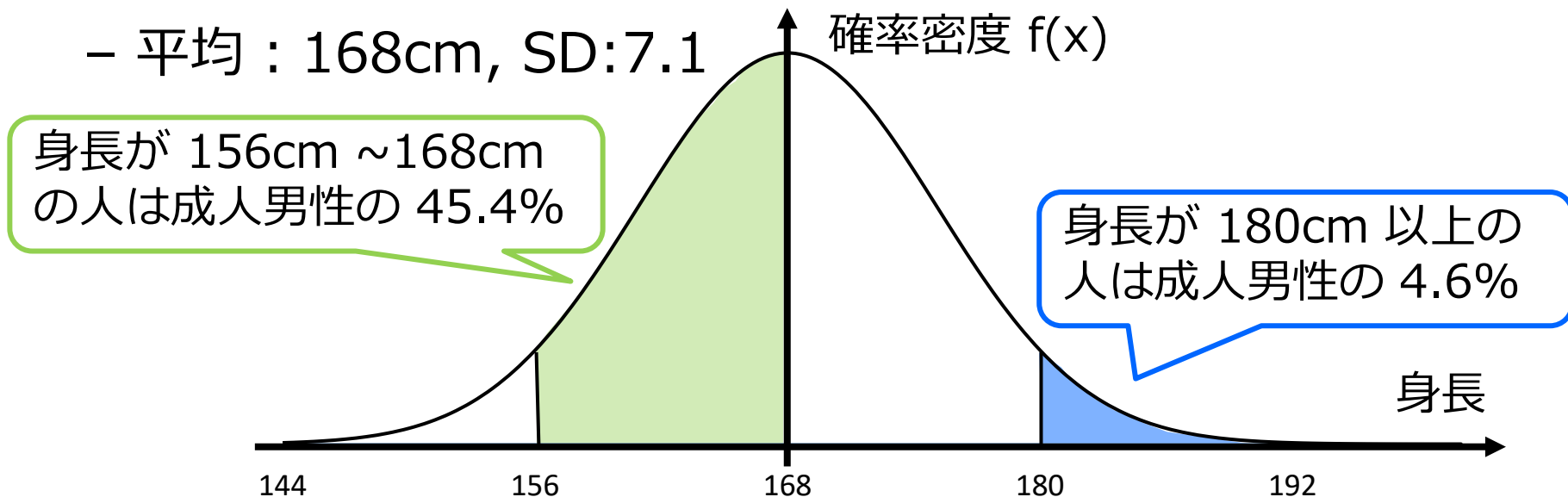


統計の復習：確率密度関数とは？

- すごく難しい概念に見えるけども
 - 検定のp値を計算するためにも使う
 - 正規分布の上側 5%: 平均 + 1.68*SD
 - これよりも z 統計量が大きければ片側 5% で有意

- 成人男性の身長分布（正規分布）

– 平均：168cm, SD:7.1



参考：なぜ平均と情報量が必要か

- 事前分布の設定に必要なのは平均と情報量
 - 単群 Phase II 試験ではよくベータ分布が使われる
 - 情報量： $\alpha + \beta$
 - 平均： $\alpha / (\alpha + \beta)$
- 連立方程式を解くために
平均と情報量を設定
例えば平均20%,情報10人分であれば $\alpha=2, \beta=8$

