

# がんの臨床開発における 単群第2相試験の症例数設計

長島 健悟(千葉大学)

藤井 陽介(ファイザー)

2016年2月29日

第3回 データサイエンスラウンドテーブル

はじめに

# 抗がん剤の第2相試験

- 目的
  - 代替エンドポイントを用いて抗腫瘍効果のある薬剤を選別
- 評価
  - 有効性の評価
  - 安全性の評価
- 評価項目(有効性)
  - 腫瘍縮小効果に基づく最良効果(2値データ)
  - PFS, OS
- 群
  - 単群
  - 多群

# 抗がん剤の第2相試験

- 目的
  - 代替エンドポイントを用いて抗腫瘍効果のある薬剤を選別
- 評価
  - 有効性の評価
  - 安全性の評価
- 評価項目(有効性)
  - 腫瘍縮小効果に基づく最良効果(2値データ)
  - PFS, OS
- 群
  - 単群
  - 多群

## 本セッションの趣旨

- 本会議の案内に掲載したもの

抗がん剤の第II相試験の目的は、代替エンドポイント(通常、奏効率)を用いて抗腫瘍効果のある薬剤を選別する事である。したがって、効果の期待できる薬剤をできるだけ早く効率的に選択し、第III相試験の実施を決定できるような情報を得ることが重要である。代表的なデザインとしてSimon(1989)の二段階デザインなどがあり、実際によく用いられている。また、近年、多くのデザインが提案されている。そこで、本テーマでは

- 比較的新しい第II相試験のデザインおよび症例数設計の適用可能性
- 抗がん剤の臨床開発における単群第II相試験の症例数設計の経験に対する参加者の情報共有(用いた手法、工夫した点など)

について議論を行う。

- とはいえ・・・

例数設計の背後にある試験デザインに関する統計的枠組みに関する議論も重要と判断し、そのトピックを取り上げています。

# 本セッションの進め方

## • 本セッションの流れ

1. グループディスカッション＋発表資料作成(60分)
  - 2. での発表資料は以下のアジェンダを進めつつ書記が発表資料作成
  - 書記は実行委員が行います
2. 各グループからの発表(60分, 1グループ最大10分)
  - 発表者は参加者から有志で(1. のときに書記系の隣に座る)

## • 上記1. のアジェンダ

- － はじめに(10分) ※ 2. での発表者の決定も含む
- － 頻度論流の代表的な方法(5分)
- － ベイズ流の代表的な方法(15分)
- － 議論(10分)
- － 発展的な方法(5分)
- － 議論(10分)
- － 総合的な議論・・・時間があれば

このセッション中にみなさまへ伺いたいこと

抗がん剤の臨床開発における第2相試験に関する  
みなさまの経験をお聞かせください。

他の方の経験でも構いません。

キーワード: 開発計画, 試験計画, 症例数設計

- 当該薬剤開発計画における第2相試験の位置づけは？
- その開発計画に基づいてどのような試験計画を立案したか？
- その試験計画に基づいてどのように症例数設計を行ったか？
- 以上に関連した質問をしながら議論を進めて参ります

## みなさまの経験は？【N=9】

---

頻度論流	単純な(1段階)デザイン【n=9】
------	-------------------

【n=】

---

Gehanの2段階デザイン【n=1】
--------------------

---

Simonの2段階デザイン【n=5】
--------------------

---

ベイズ流	事後確率によるデザイン【n=1】
------	------------------

【n=】

---

予測確率によるデザイン【n=1】
------------------

---

発展的なアプローチ	多重アウトカムを伴うデザイン【n=0】
-----------	---------------------

【n=】

---

2変量2値データデザイン(第1/2相)【n=0】
--------------------------

---

毒性および効果の発現までの時間を伴うデザイン(第1/2相)【n=0】
------------------------------------

---

- シームレス1/2【n=2】
- SWOGの標準デザイン【n=1】

# 頻度論流の代表的な方法

# Simonの2段階デザイン (1989)

---

## Optimal Two-Stage Designs for Phase II Clinical Trials

**Richard Simon, PhD**

*Biometric Research Branch, National Cancer Institute, Bethesda, Maryland*

---

**ABSTRACT:** The primary objective of a phase II clinical trial of a new drug or regimen is to determine whether it has sufficient biological activity against the disease under study to warrant more extensive development. Such trials are often conducted in a multi-institution setting where designs of more than two stages are difficult to manage. This paper presents two-stage designs that are optimal in the sense that the expected sample size is minimized if the regimen has low activity subject to constraints upon the size of the type 1 and type 2 errors. Two-stage designs which minimize the maximum sample size are also determined. Optimum and “minimax” designs for a range of design parameters are tabulated. These designs can also be used for pilot studies of new regimens where toxicity is the endpoint of interest.

# Simonの2段階デザイン (*Controlled Clin Trials* 1989)

- 仮説検定の枠組みを用いる
- 2つの段階から成る
  - 第1段階: 次相へのno-goを早期に検討する(futility stop)
  - 第2段階: 次相へのgo, no-goを検討する
- 2つの閾値
  - それ以下だと臨床的意義のない反応率:  $p_0$
  - それ以上だと臨床的意義のある反応率:  $p_1$
- 症例数の組み合わせは複数
  - 閾値が2つあることに由来
  - 症例数設計においては, 制約をつけて一意に定める
    1. 最適デザイン(optimal)
      - 帰無仮説のもとでの期待症例数を最小にする
    2. ミニマックスデザイン(minimax)
      - 最大症例数を最小にする

URL:

<http://cancer.unc.edu/biostatistics/program/ivanova/SimonsTwoStageDesign.aspx>


UNC Lineberger Comprehensive Cancer Center

**Anastasia Ivanova, Ph.D, University of North Carolina at Chapel Hill**

Continuous monitoring for toxicity | 
 Simon's two-stage design | 
 Fleming's two-stage design | 
 Simon's like design with relaxed futility stopping

Two-stage design for ordinal outcomes | 
 The Rapid Enrollment Design (RED) for Phase I trials | 
 Other programs

### Simon's Two-Stage design

This program generates Simon's optimal two-stage designs (Simon, 1989) and admissible designs from Jung et al. (2004) for Phase II single arm clinical trials.

- Simon R (1989). *Controlled Clinical Trials* 10: 1-10. [Click here to download Simon's \(1989\) article.](#)
- Jung SH, Lee TY, Kim KM, George S (2004). *Admissible two-stage designs for phase II cancer clinical trials*, *Statistics in Medicine* 23: 561-569.

Type I error rate,  $\alpha$  (one-sided):

Power:

Response probability of poor drug,  $p_0$ :

Response probability of good drug,  $p_1$ :

$n$	$n_1$	$r_1$	$r_2$	Type 1 Error	Power	$EN_0$	Probability of early stopping	Interval for $w$	Comment
33	18	4	10	0.0458	0.8011	22.3	0.7164	[0.1683,1]	Minimax
38	14	3	11	0.0495	0.8154	21.2	0.6982	[0.1172,0.1682]	
43	13	3	12	0.0496	0.8002	20.6	0.7473	[0,0.1171]	Optimal

Calculated in 4 milliseconds

$n$  is the total number of subjects  
 $n_1$  is the number of subjects accrued during stage 1  
 $r_1$ , if  $r_1$  or fewer responses are observed during stage 1, the trial is stopped early for futility  
 $r_2$ , if  $r_2$  or fewer responses are observed by the end of stage two, then no further investigation of the drug is warranted  
 $EN_0$  is the expected sample size for the trial when response rate is  $p_0$   
**Interval for  $w$**  is the set of values  $w$  such that the design minimizes  $w * n + (1 - w) * EN_0$

**Recommended write up for a protocol**  
*Simon's two-stage design (Simon, 1989) will be used. The null hypothesis that the true response rate is  $[p_0]$  will be tested against a one-sided alternative. In the first stage,  $[n_1]$  patients will be accrued. If there are  $[r_1]$  or fewer responses in these  $[n_1]$  patients, the study will be stopped. Otherwise,  $[n - n_1]$  additional patients will be accrued for a total of  $[n]$ . The null hypothesis will be rejected if  $[r_2 + 1]$  or more responses are observed in  $[n]$  patients. This design yields a type I error rate of  $[Type I error rate]$  and power of  $[power]$  when the true response rate is  $[p_1]$ .*

The development of this software was supported by funds from the National Institutes of Health [RO1 CA120082-01A1].  
 For comments, questions and suggestions e-mail to Anastasia Ivanova at [aivanova@bios.unc.edu](mailto:aivanova@bios.unc.edu)



UNC Lineberger Comprehensive Cancer Center  
 Campus Box 7295, Chapel Hill, NC 27599  
 Appointments: 1-800-800-1856  
 Copyright © 1999-2012. All rights reserved.  
 UNC Home | UNC Health Care | UNC School of Medicine



# ベイズ流の代表的な方法

# ベイズ流のアプローチ

- 特徴

- 連続的なモニタリング(第一種の過誤の確率の制御という観点がない)

- 2つのアプローチ

1. 事後確率によるアプローチ

- コンセプト: それまでに蓄積されたデータに基づいて反応率の事後分布を求め, その分布から得られる事後確率に基づいて判断を下す

2. 予測確率によるアプローチ

- コンセプト: それまでに蓄積されたデータに基づいて反応率の事後分布を求め, その後のデータがこの分布に従うと仮定したうえで, 有効と判断される確率(予測確率: PP, Predictive Probability)に基づいて判断を下す

## 事後確率による方法(1/2)

- **Thall and Simon (*Biometrics* 1994)**

- $n_{\max}$ : サンプルサイズの上限
- $p_E \sim \text{Beta}(\alpha_E, \beta_E)$ : 試験治療の反応率(事前分布)
- $p_S \sim \text{Beta}(\alpha_S, \beta_S)$ : 標準治療の反応率(事前分布)
- $\alpha_E, \beta_E, \alpha_S, \beta_S$ : ハイパーパラメータ
- $Y_n \sim \text{Binomial}(n, p_E)$ :  $n (\leq n_{\max})$  人目までの試験治療のデータ
- $p_E | y_n \sim \text{Beta}(\alpha_E + y_n, \beta_E + n - y_n)$ :  $p_E$ の事後分布

- **事後確率**

- 最低限必要な群間差を  $\delta (\in (0, 1))$  とし、事後分布において  $p_E > p_S + \delta$  なる確率を考える

$$\begin{aligned} & \Pr(p_E > p_S + \delta \mid Y_n = y_n) \\ &= \int_0^{1-\delta} \{1 - F(p; \alpha_E + y_n, \beta_E + n - y_n)\} f(p; \alpha_S, \beta_S) dp \end{aligned}$$

$F$ : ベータ分布の分布関数,  $f$ : ベータ分布の密度関数

※ 数値積分で簡単に計算できる

## 事後確率による方法(2/2)

- 境界値

- 優越性の境界値 $\theta_U$ を通常 0.95~0.99 に設定 / 設定しない場合あり
- 無益性の境界値 $\theta_L$ を通常 0.01~0.05 に設定
- $\Pr(p_E > p_S \mid Y_n = y) \geq \theta_U$  を満たす最小の  $y$  を  $U_n$  とする
- $\Pr(p_E > p_S + \delta \mid Y_n = y) \leq \theta_L$  を満たす最大の  $y$  を  $L_n$  とする

- 決定規則

- $Y_n \geq U_n$  の場合, 見込みありと判断し, 試験を中止
- $Y_n \leq L_n$  の場合, 見込みなしと判断し, 試験を中止
- $L_n < Y_n < U_n$  かつ  $n < n_{\max}$  の場合, 試験を継続
- $n_{\max}$  に到達して中止に至らない場合, 結論を保留

- 症例数の見積もり

- シミュレーションを実施して検討

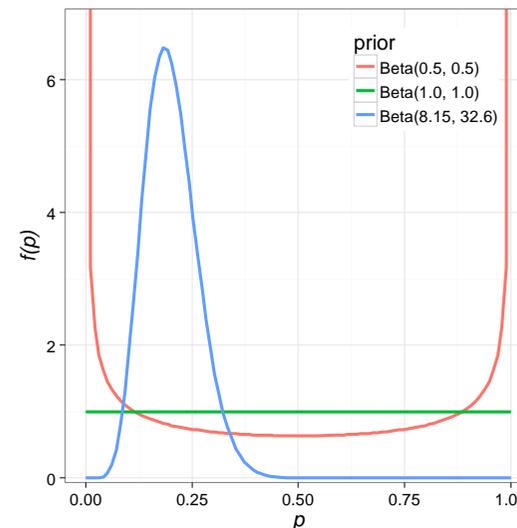
- 適用例: Chang et al. (Lancet Oncol. 2009)

- (単群試験ではないが)脳腫瘍の術式のランダム化比較試験で二群の試験に応用(結果, 有効中止となった)

# 事後確率による方法(補足)

## • 事前分布

- 試験治療には $\alpha + \beta$ が小さく ( $\leq 10$ ) 情報量の少ない事前分布  
無情報: Beta(0.5,0.5), Beta(1,1)  
情報:  $E[p_E] = E[p_S] + \delta$  or  $E[p_S] + \delta/2$ ,  $\alpha + \beta \leq 10$ として求める
- 標準治療には既知の結果等からある程度の情報量の事前分布
  - e.g.)  $p_E \sim \text{Beta}(1,1)$ ,  $p_S \sim \text{Beta}(8.15, 32.6)$ ,  $\alpha + \beta$ は例数に相当
- ベータ分布の期待値 =  $\alpha/(\alpha + \beta)$ : 反応率に相当
  - e.g.)  $E[p_E] = 1/(1 + 1) = 50\%$ ,  $E[p_S] = \frac{8.15}{8.15+32.6} = 20\%$
- $W_{90}$ : 分布の5%~95%タイル  
までの幅: 情報量に相当
  - e.g.)  $p_E : W_{90} = 0.9$ ,  
 $p_S : W_{90} = 0.2$



## 予測確率による方法(1/2)

- **Lee and Liu (*Clinical Trials* 2008)**

- $Y_n = y_n$ を得たもとで, 試験終了までの  $n_{\max} - n$  人の事後予測分布  $X$  はベータ二項分布に従う; この確率関数を  $P(X = x | y_n)$  とおく

$$X | y_n \sim \text{Beta-Bin}(n_{\max} - n, \alpha_E + y, \beta_E + n - y)$$

- 最後まで登録があった時の事後分布は

$$p_E | y_n, x \sim \text{Beta}(\alpha_E + y + x, \beta_E + n_{\max} - y - x)$$

- 仮に試験が最後まで到達したときに, 試験治療が標準治療よりも反応率が高くなる確率を考える

$$\Pr(p_E > p_S | Y_n = y_n, X = x)$$

※  $p_S$  には事前分布を仮定せず, 標準治療の有効確率を直接与える

- 十分高い  $\theta_T$  を考え (通常, 0.85~0.95),  $\Pr(p_E > p_S | Y_n = y_n, X = x) \geq \theta_T$  を満たすならば試験治療に見込みがあると考えられる

- 実際は  $X$  が観測できず, 出現確率で平均化した以下の指標を考える

$$PP = \sum_{x=0}^{n_{\max}-n} P(x | y_n) I\{\Pr(p_E > p_S | y_n, x) \geq \theta_T\}$$

※  $I\{ \}$  は indicator function

## 予測確率による方法(2/2)

- 境界値

- 事後分布の場合と同様に, 優越性の境界値 $\theta_U$ を通常高い値に, 無益性の境界値 $\theta_L$ を通常低い値に設定

- 決定規則

- $PP > \theta_U$ の場合, 見込みありと判断し, 試験を中止
- $PP < \theta_L$ の場合, 見込みなしと判断し, 試験を中止
- その他の場合は $n_{\max}$ に至るまで試験を継続

- 症例数の見積もり

- シミュレーションを実施して検討

- 適用例: Tap et al. (JCO 2012)

- 線維形成性小細胞腫瘍のPh 2試験の無益性のモニタリングに応用

- 補足

- R package {ph2bayes}

<https://cran.r-project.org/web/packages/ph2bayes/>

# 議論

## みなさまの経験は？【N=9】

---

頻度論流 単純な(1段階)デザイン【n=9】

---

【n=】 Gehanの2段階デザイン【n=1】

---

Simonの2段階デザイン【n=5】

---

ベイズ流 事後確率によるデザイン【n=1】

---

【n=】 予測確率によるデザイン【n=1】

---

発展的なア 多重アウトカムを伴うデザイン【n=0】

---

プローチ 2変量2値データデザイン(第1/2相)【n=0】

---

【n=】 毒性および効果の発現までの時間を伴うデザイン(第1/2相)【n=0】

---

- シームレス1/2【n=2】
- SWOGの標準デザイン【n=1】

2段階デザインにおいて最適デザイン, ミニマックスデザイン,  
どちらを利用しますか? その根拠は?

- ミニマックスデザインを選択したケース
  - 集積が厳しく例数を抑えたかった(60例程度で5例~10例程度違った)
  - 大きな差は出てこない; 大体30~50例ぐらい, 二つのデザインの違いは5例程度?
- 両方とも実現できなかったケース
  - バイオマーカー=ポジの患者に限定した試験で, feasibility的に困難
  - 期待奏効率と閾値奏効率の設定が困難
    - この領域ではそれほど情報が集まっていない段階で設定しなければならない
- 集積スピード
  - 速い: 中間解析が不要だった(単純デザイン)
  - 遅い: 2段階デザイン(途中でチェックできる)

なぜ, Simonの2段階デザインが頻用されるのでしょうか？

- 少数例では有効中止は難しいことから無効中止のSimonが頻用
- 過去の事例に沿って, Simonを頻用

- 統計解析計画書
  - オペレーショナルな観点
  - 開発計画との関係性
    - グローバル開発戦略
- など

ベイズ流の方法を使用した経験は？その方法を使用した背景は？

- 希少疾患での事例

- 背景： 頻度論では過誤確率が大きくなるので、推定の枠組みで
  - 計画上例数が小さくて済んだ(頻度論流よりも10例程度小)
  - しかし、最終的にはfeasibilityの観点から
- 方法： Whiteheadの方法(有効性・安全性の基準を設定)
  - 中間解析は入れなかった
  - ベータ二項分布の共役
- 他の開発チームからは「任せた」と言われた(ベイズは複雑だから？)

- バイオマーカー＝ポジを対象とした事例

- 背景： 奏効率の情報が未だ明確ではない。また、例数が多くない
- 方法： 予測確率に基づくデザインのアプローチ
  - シームレス1/2

ベイズ流の方法にはどのようなメリット・デメリットがありますか？

- メリット

- (統計家には)直観的にわかりやすい
- 閾値を設定しなくてもいい
- 症例数を低く抑えられる

- デメリット

- 先生や試験運営にはわかりにくい
- 経験が少ない(論文数も少ない)
- Bayes流方法論のメリットを理解できていない

## ベイズ流の方法を使いますか？

- 使う
  - Bayesian
  - 海外(本社)チームが使って行こうという流れ
    - 開発戦略としてadaptive designを取り入れる
- 使わない(使うことができない)
  - ベイズを理解できる・適用できるメンバーが少ない
- 要検討事項
  - プロトコール, CSRの記載

# 発展的な方法

# 多重アウトカムを伴うデザイン

[有効確率および毒性確率を同時に観察するデザイン]

- **Thall, Simon, and Estey (StatMed 1995)**

- 単群で有効確率と毒性確率を同時評価
- 評価変量が二次元になるため尤度関数を多項分布, 事前分布と事後分布にディリクレ分布を仮定

- **推測**

- 事後分布の周辺分布がベータ分布に従う事を用いた推測を考える
- 標準治療における有効確率と毒性確率の事前分布をベータ分布で与え, Thall and Simon (*Biometrics* 1994) と同様の方法を用いる

- **中止境界値**

- 拡張に伴い, 優越性と無益性の境界値に加え, 毒性の境界値が追加する
- 毒性の境界値に基づいて中止が可能になる

# 2変量2値データデザイン (Ph 1/2)

[有効確率および毒性確率を同時に観察するデザイン]

## ● 背景

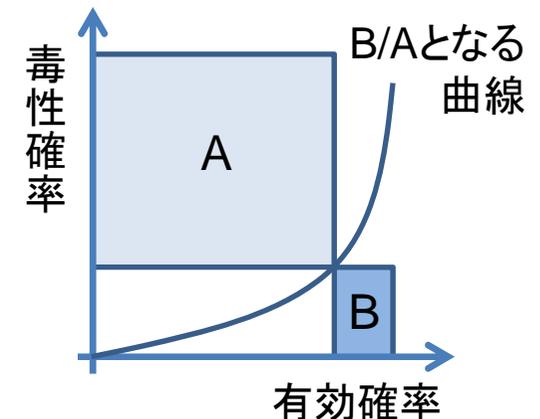
- Ph 1の目的: MTDの推定。少数の被験者が過剰な毒性をもつ用量や有効でない用量で治療される
- Ph 2の目的: 探索したMTDで短期の有効性を評価
- MTD推定に関する問題点: 各phaseは別々に実施されるが, Ph 1の被験者数には限りがあるため, MTDの推定は不正確となることがあり, 後相でその影響が生じる可能性あり

## ● メリット

- 試験の間が無くなることによる開発期間の節約
- 有効かつ毒性をコントロール可能な用量の十分な探索
- データのプーリングによる精度の向上

## ● モデル

- 各用量における有効確率と毒性確率の多項分布に, 毒性確率の用量単調性の仮定を入れたモデル
- 有効および毒性確率によって構成される平面上で, これらの値に基づく面積比(B/A)が等しくなる等高線を構成し, 用量選択基準(面積比小が良)とする
- 毒性と効果の相関を考慮することもできる(上記平面を立体に拡張)



# 毒性および効果の発現までの時間を伴うデザイン(Ph 1/2)

## • 背景

- どのぐらいの早さで有効性や毒性が生じたか加味する
  - e.g.) 反応を示すまでの時間, 毒性発現までの時間(便宜上, 生存時間)
  - 有効性・毒性発現の経時推移も検討に入れることができる

## • モデル

- 以下のパラメトリックモデルの組み合わせ
  - イベント時間: 比例ハザード性を仮定したワイブル分布
  - 有効性のモデル: 無反応・反応者を考慮するためにcure model
  - 2つのイベント時間の相関: 周辺分布のコピュラ
- Kaplan-Meier曲線のAUSC(制約付き平均生存時間)の比が指標の一つとして提案されている

## • 用量探索アルゴリズム

- 初めは毒性確認コホートを設けて様子を見てからベイズ流に切り替え
  - e.g.)  $n=3$ で毒性発現例数を確認
- 毒性イベントの数に応じて用量を変更したり, 有効性および毒性に関して生存確率のカットオフ値を設定し, それに基づいて探索を行うもの

# 議論

## みなさまの経験は？【N=】

### 頻度論流

単純な(1段階)デザイン【n=】

Gehanの2段階デザイン【n=】

Simonの2段階デザイン【n=】

### ベイズ流

事後確率によるデザイン【n=】

予測確率によるデザイン【n=】

### 発展的なアプローチ

多重アウトカムを伴うデザイン【n=】

2変量2値データデザイン(第1/2相)【n=】

毒性および効果の発現までの時間を伴うデザイン(第1/2相)【n=】

- その他【n=】
- その他【n=】

# 発展的な方法にはどのようなメリット・デメリットがありますか？

- Ph 1/2

- コホートエクспанション

- Ph 1的に用量を選択し, Ph 2として n 例を追加する。

- 期間の短縮はメリットだが, 有効性の評価は難しい

- デイジジョンをするにはデータが乏しいか？

- 統計解析計画書
  - オペレーショナルな観点
  - 開発計画との関係性
    - グローバル開発戦略
- など