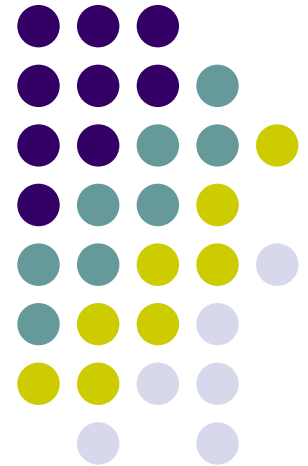


中間解析における 症例数再推定のための PH法とLi法の性能比較

2011.7.29 SASユーザー総会

塩野義製薬(株) 解析センター
馬場裕子, 多田由希夫
長谷川貴大, 渡辺秀章





はじめに

- 近年、「臨床試験の効率化」のための新しい開発ツールの一つとして、Adaptive designが注目されている。
- Adaptive designとは
 - 試験の科学的適切性を損なうことなく、試験中に実施する中間解析の結果に基づき、試験計画の一部について変更を許容する試験デザイン
- 具体的には
 - 試験中止(有効性, 無益性)
 - 治療群の中止
 - 症例数の変更
 - 割付比の変更 など



本報告の構成

1. 条件付き検出力に基づく症例数の再推定
 - 1.1 Proschan and Hunsbarger (PH) 法
 - 1.2 修正PH (Li) 法
2. シミュレーション比較
 - 2.1 PH法とLi法の性能比較
 - 2.2 Li法のより効率的な使い方を検討
3. まとめ

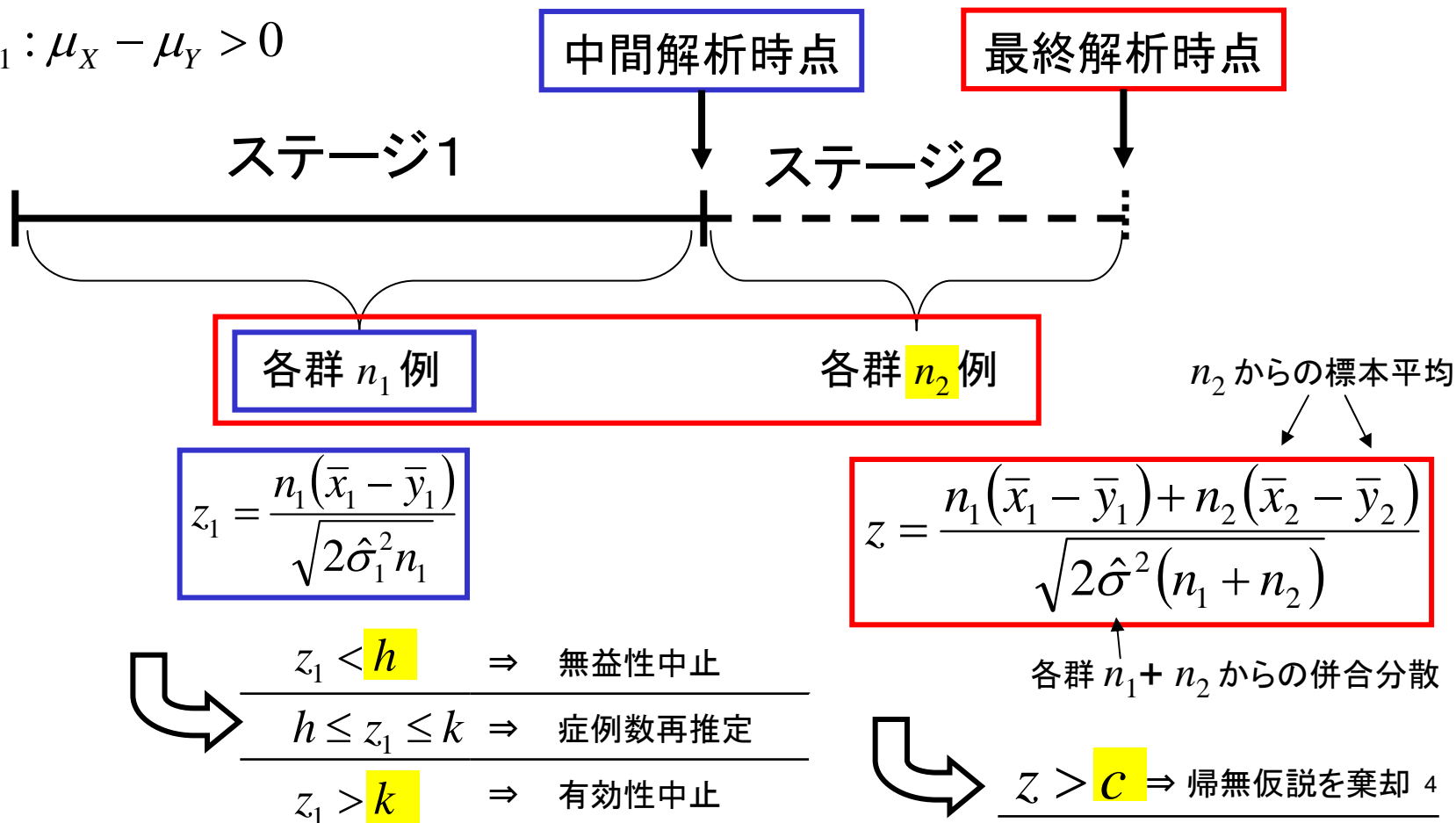
1. 条件付き検出力に基づく症例数の再推定



2群比較試験
正規分布
平均差の検定

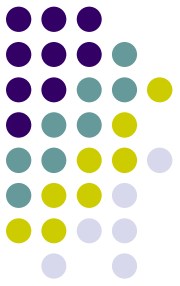
$$H_0 : \mu_X - \mu_Y = 0$$

$$H_1 : \mu_X - \mu_Y > 0$$



条件付き検出力

ステージ1が終了した(中間解析)段階で, 群間比較を行って得られる情報を用いて, ステージ2の終了(最終解析)時での検定結果を予測する.



● $Z_1=z_1$ の条件付き検出力

$$CP_\delta(n_2, c | z_1) \equiv P(z > c | z_1, \delta) \\ = 1 - \Phi \left[\frac{c\sqrt{n_1 + n_2} - z_1\sqrt{n_1} - n_2\delta/\sqrt{2}}{\sqrt{n_2}} \right]$$

- $\delta = (\mu_X - \mu_Y)/\sigma$: 仮説上の効果サイズ
- Φ : 標準正規分布関数
- $\hat{\sigma}_1^2 \approx \hat{\sigma}^2 \approx \sigma^2$: 中間解析と最終解析の併合分散が想定と等しい

● 帰無仮説のもとでの $Z_1=z_1$ の条件付き確率

- $\delta = 0$ とする

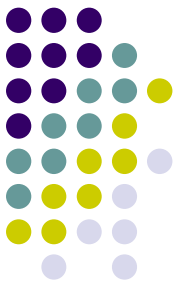
$$CP_0(n_2, c | z_1) = 1 - \Phi \left[\frac{c\sqrt{n_1 + n_2} - z_1\sqrt{n_1}}{\sqrt{n_2}} \right] \equiv A(z_1)$$

PH法では,
条件付エラー
関数という

- 試験全体でのタイプ I エラー

$$\int_0^\infty A(z_1)\phi(z_1)dz_1 = \alpha$$

片側有意水準



1.1 PH法

- ステージ2の必要症例数

$$n_2 = \frac{2(z_{A(z_1)} + z_{\beta_1})^2}{\hat{\delta}^2} = n_1 \frac{(z_{A(z_1)} + z_{\beta_1})^2}{z_1^2}$$

最終解析で想定する
条件付き検出力

中間解析での
効果サイズ

- ステージ2の棄却限界値

$$c = \frac{z_1^2 + (z_{A(z_1)} + z_{\beta_1})z_{A(z_1)}}{\sqrt{z_1^2 + (z_{A(z_1)} + z_{\beta_1})^2}}$$

条件付きエラー関数 $A(z_1)$ に依存している。

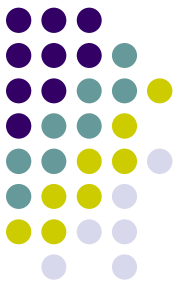
中間解析の統計量 z_1 に依存している。

- 条件付きエラー関数 (Circular conditional error functionの場合)

- $\int_0^\infty A(z_1)\phi(z_1)dz_1 = \alpha$ を満たし、値域が $[0, 1]$ である任意の増加関数

$$A(z_1) = \begin{cases} 0 & \text{if } z_1 < h \\ 1 - \Phi(\sqrt{k^2 - z_1^2}) & \text{if } h \leq z_1 < k \\ 1 & \text{if } k \leq z_1 \end{cases}$$

無益性中止基準 h と有効性中止基準 k が独立に規定できない。



1.2 Li法

- ステージ2の必要症例数

$$n_2 = \left[\left(\frac{c + z_{\beta_1}}{z_1} \right)^2 - 1 \right] n_1$$

条件付きエラー関数
とは無縁

- 条件付き検出力

$$1 - \beta_1 = 1 - \Phi \left(\frac{c\sqrt{n_1 + n_2} - z_1\sqrt{n_1} - n_2\hat{\delta}/\sqrt{2}}{\sqrt{n_2}} \right)$$

- 条件付き検出力は、少なくとも $1 - \beta_1$ となればよいので、分母の $\sqrt{n_2}$ を $\sqrt{n_1 + n_2}$ に置き換えて解く

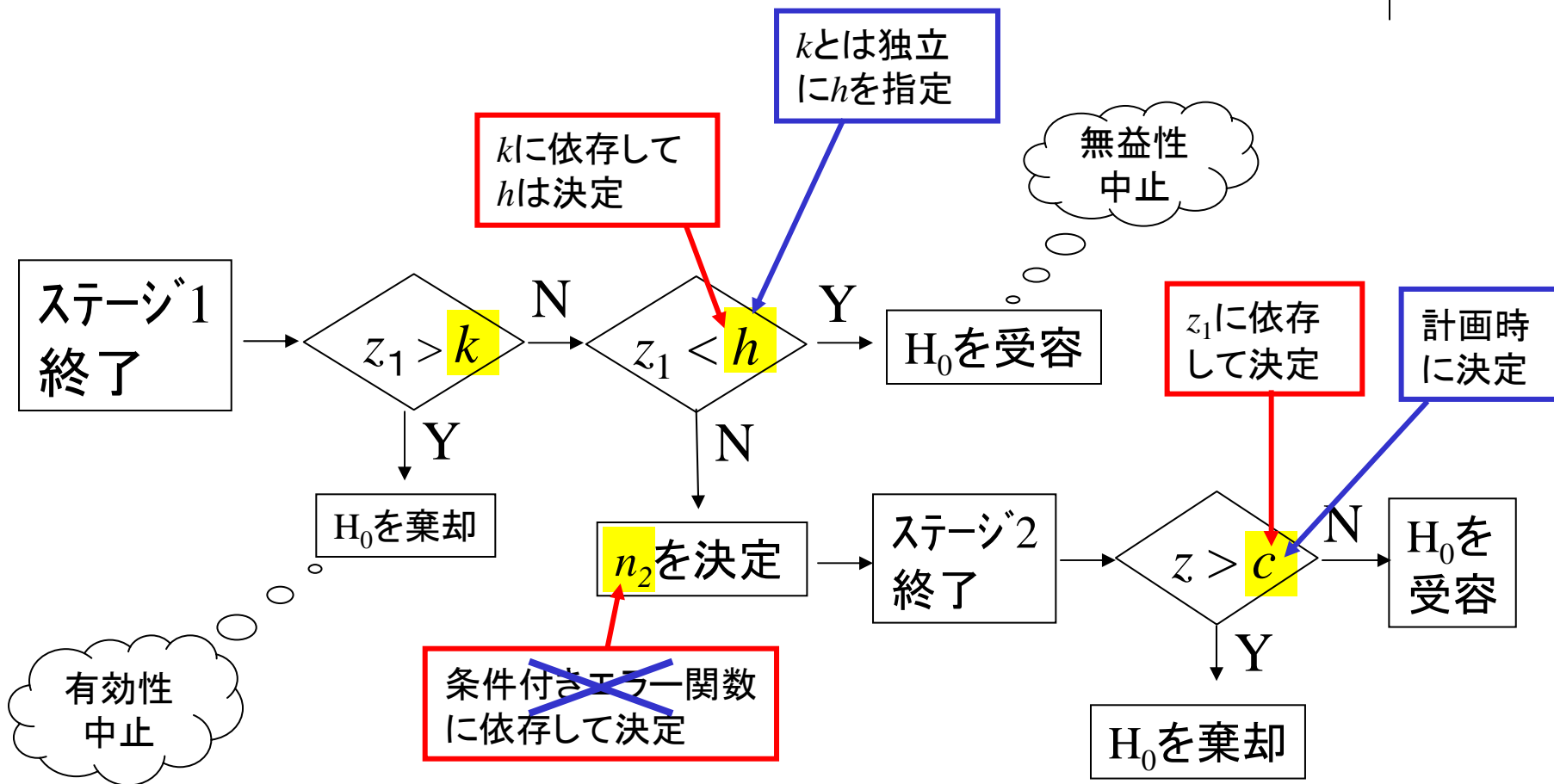
- ステージ2の棄却限界値

$$p^* - \alpha = \int_h^k \Phi \left[\frac{c(c + z_{\beta_1}) - z_1^2}{\sqrt{(c + z_{\beta_1})^2 - z_1^2}} \right] \phi(z_1) dz_1$$

k, h, α, β_1 が決まれば、 c が決まる。
 c は z_1 に依存しない。

を満たす c を数値積分で求める。ただし、 $p^* = 1 - \Phi(h)$ とする。

PH法とLi法の違い





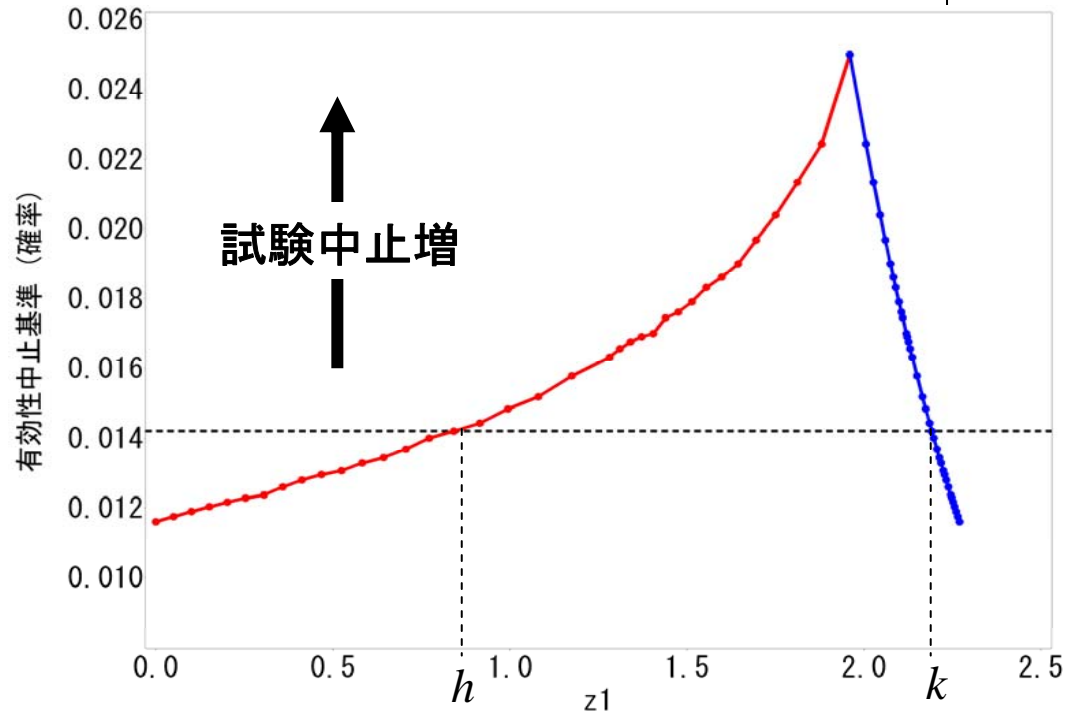
PH法よりもLi法が柔軟で有用

- PH法の中止基準値

- $\alpha = 0.025$ (片側)
- $\beta = 0.2$

$h = 0.842$ (0.2000) ,
 $k = 2.192$ (0.0142)

$h = 0.000$ (0.5000) ,
 $k = 2.270$ (0.0116)



- PH法は、中間解析時の有効性中止基準 k が大きい場合(試験中止を積極的にしない場合)に利用できない。

- 50%時点で中間解析をする場合、O'Brien-Flemingデザインでは、 $k = 2.7965$ (0.0026)



2. シミュレーション比較

- 目的

- PH法の中止基準を用いた場面に、PH法とLi法で、条件付き検出力および期待症例数の違いを評価する。
 - 中間解析時点
 - 最大症例数

- 仮定

計画時より真値
が低いとする

ある程度中止
しやすい設定

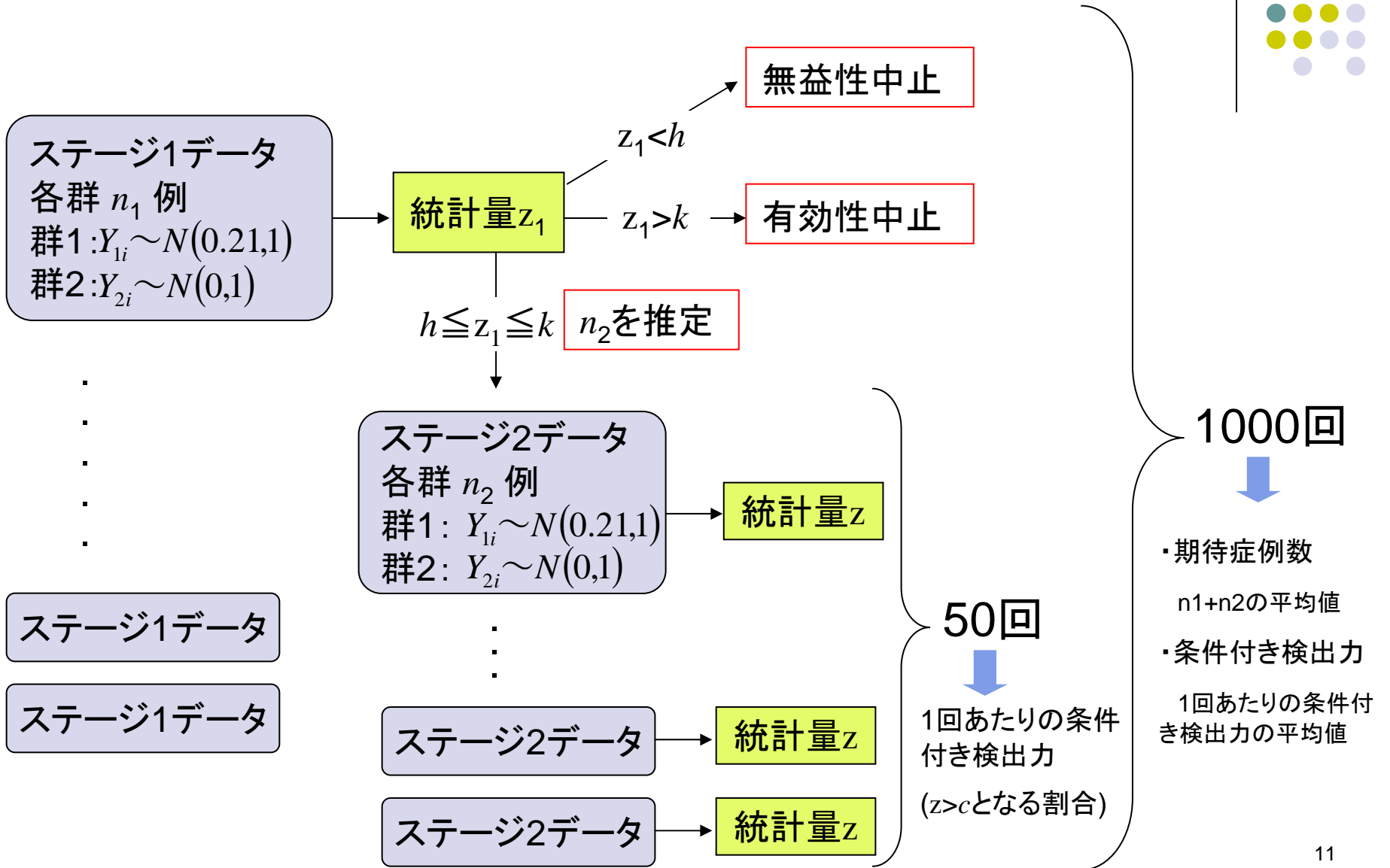
群間差 (計画時)	群間差 (真値)	標準 偏差	片側有意 水準	検出力	無益性 中止	有効性 中止
0.30	0.21	1.00	0.025	0.8	0.842 (0.2000)	2.192 (0.0142)

統計量ベース
(確率ベース)

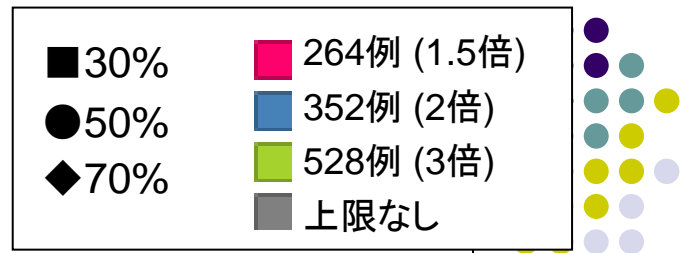
⇒ 計画時の想定での必要症例数: 176例/群

- 中間解析時点: 30%, 50%, 70%
- 最大症例数: 想定必要症例数の1.5倍(264例), 2倍(352例), 3倍(528例), 上限なし

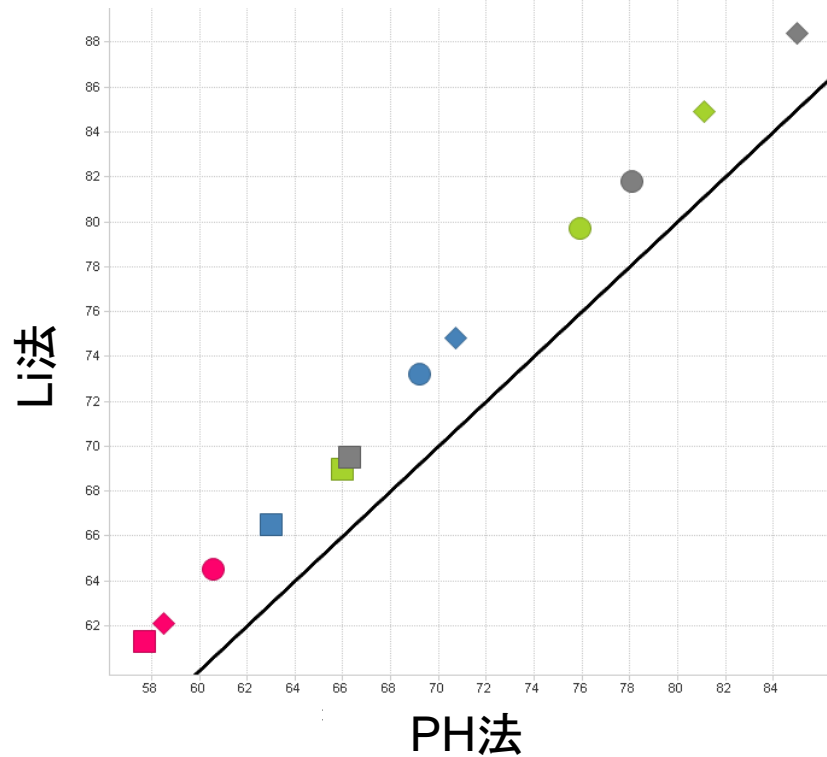
シミュレーションの流れ



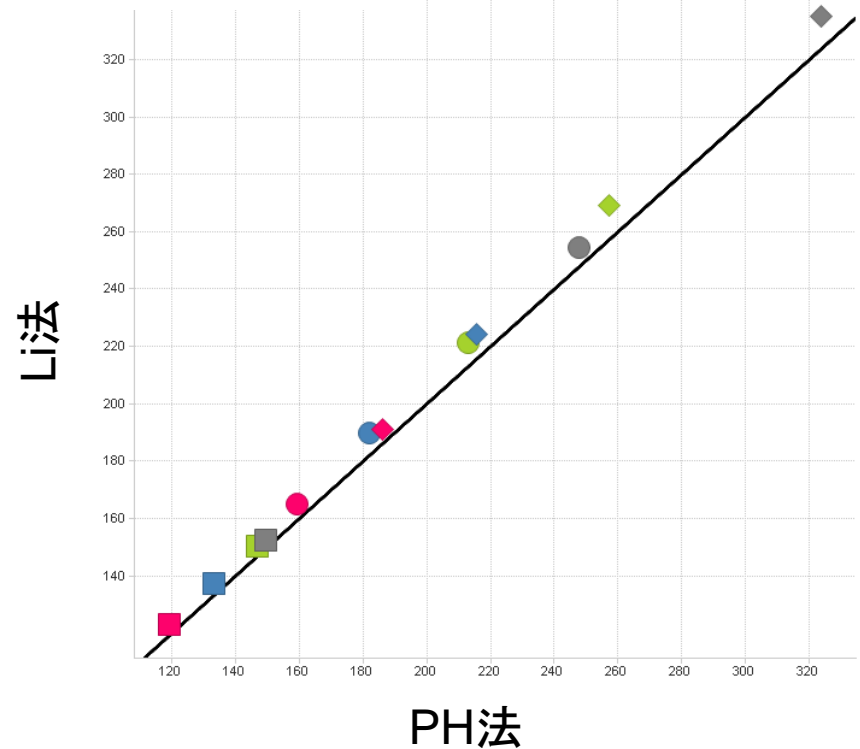
2.1 PH法とLi法の比較



条件付き検出力



期待症例数



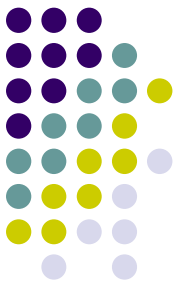
- PH法とLi法で大きな違いはない
 - Li法の方が条件付き検出力は3~4%高い
 - Li法の方が3~12例多い

PH法は $CP_{\hat{\delta}}(n_2, c | z_1) = 80$
 Li法は $CP_{\hat{\delta}}(n_2, c | z_1) \geq 80$
 となるように n_2 を決定



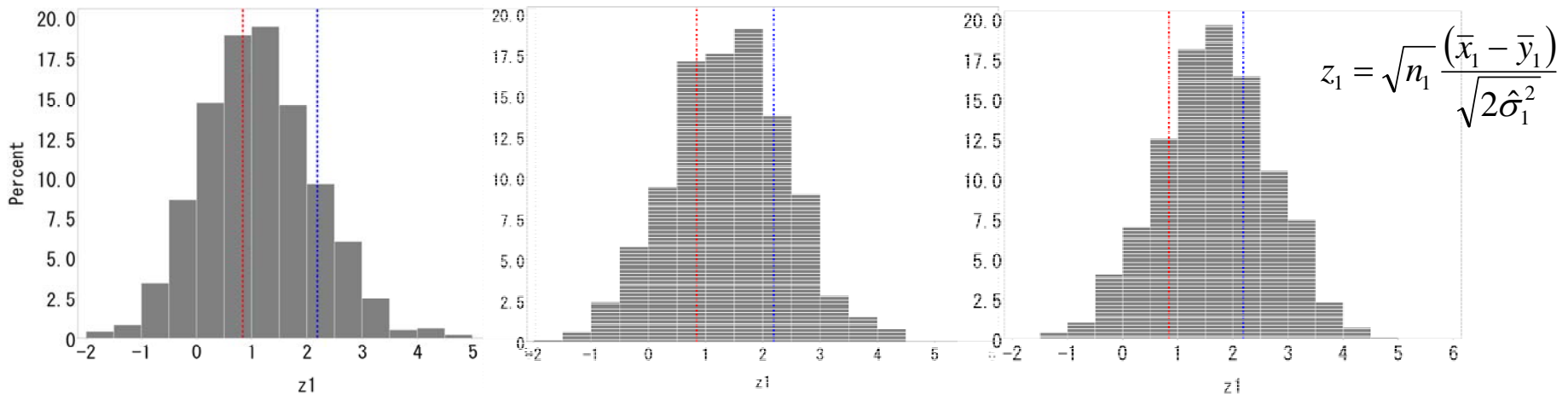
2.2 Li法により効率的な使い方を検討

- 中間解析時点, 最大症例数の違いによる影響を
検討する.
- 解析時点と無効中止率の関係
- 解析時点とステージ2での症例数の関係



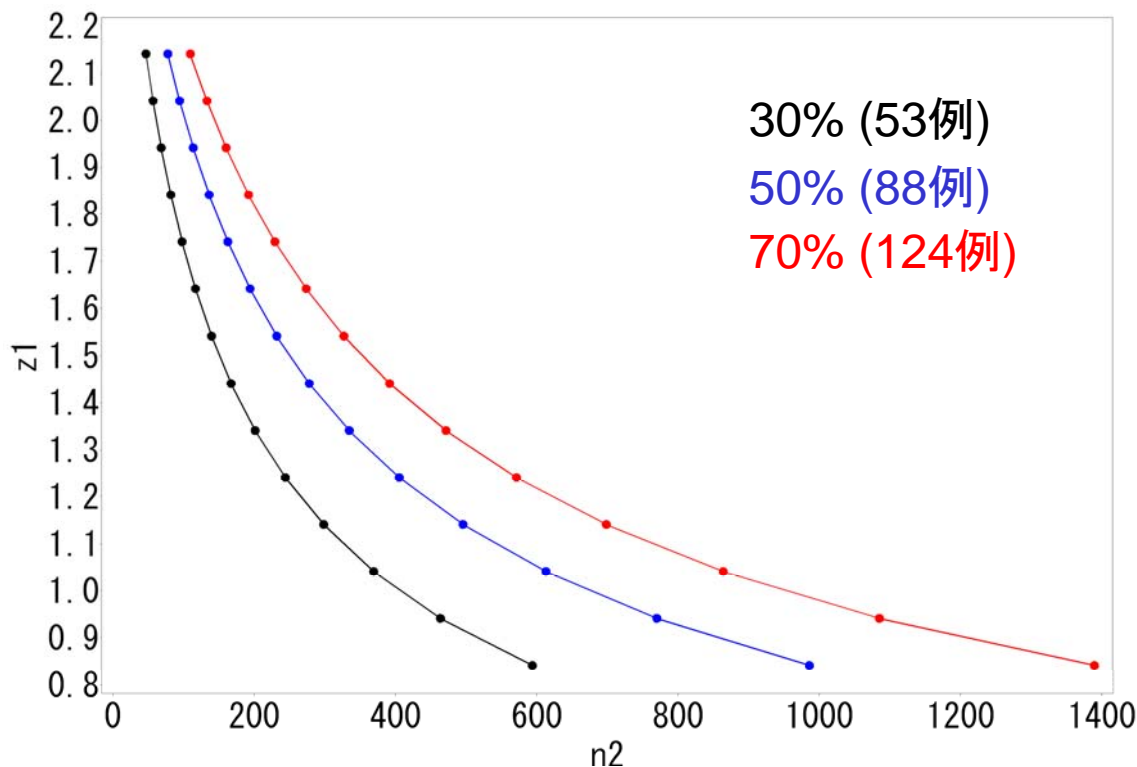
解析時点と無効中止率の関係

中間解析時点	中間解析時症例数	z_1 の 平均値	無益性中止率	有効性中止率	ステージ2への 移行率
30%	53	1.12	40.3	15.2	44.5
50%	88	1.40	28.6	22.4	49.0
70%	124	1.68	20.3	30.8	48.9



- 中間解析時点が早いと z_1 は小さくなる。
- 無益性中止を少なくするには、中間解析時点を遅くする。

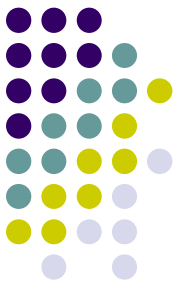
解析時点とステージ2での症例数の関係



Li法でのステージ2の例数

$$n_2 = \left[\left(\frac{c + z_{\beta_1}}{z_1} \right)^2 - 1 \right] n_1$$

- 中間解析時点が遅いと n_2 は大きくなる。
- n_2 を少なくするには、中間解析時点を早くする。



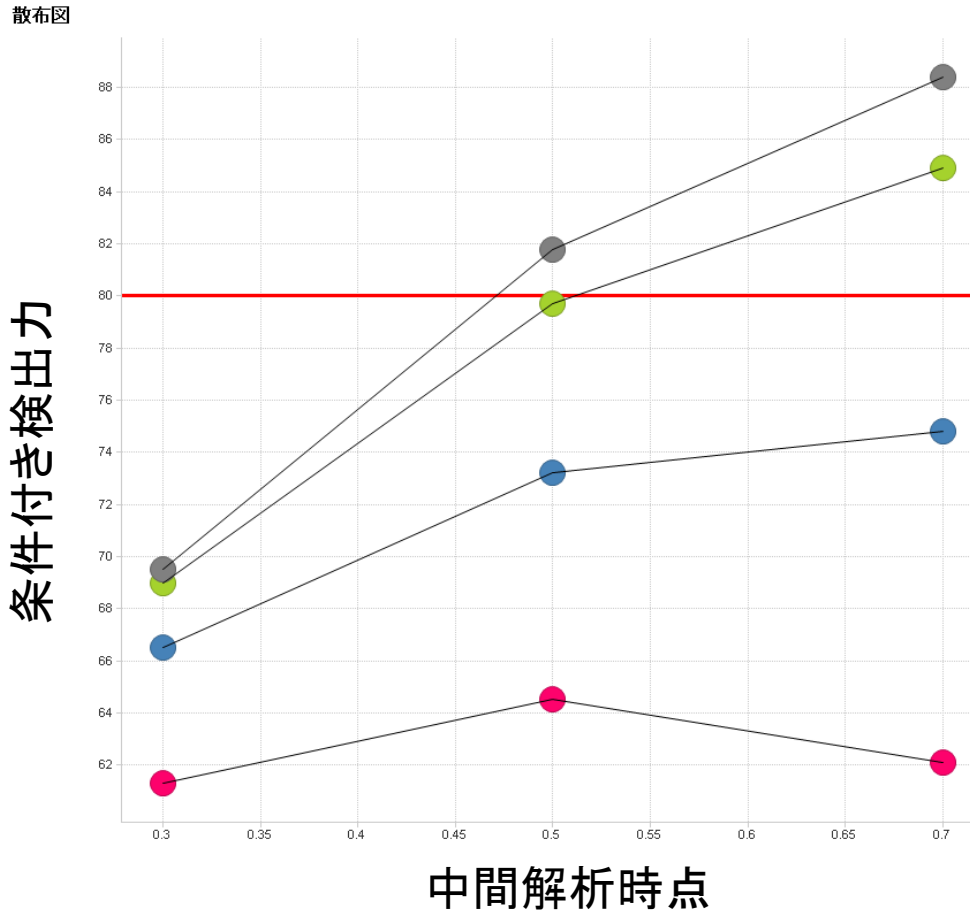
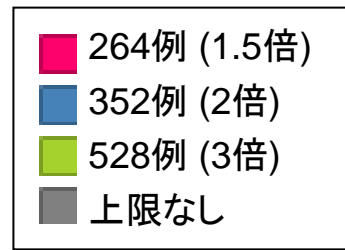
解析時点，最大症例数の違い

- 解析時点と無効中止率の関係
⇒ 解析時点は遅い方がよい
- 解析時点とステージ2での症例数の関係
⇒ 解析時点は早い方がよい

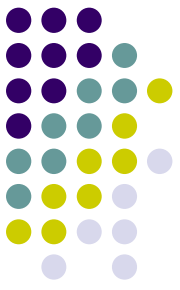


- 条件付き検出力から最適な解析時点および最大症例数を見つける！

シミュレーション結果



- 上限なしの場合と比べて、最大例数を528例(3倍)以上と設定すれば、条件付き検出力は大きく変わらない。



3. まとめ

- 条件付き検出力に基づく症例数の再推定方法として、PH法とLi法の手順を整理した.
- PH法よりもLi法が有用である.
 - 無益性中止基準と有効性中止基準を独立に規定できる.
 - シミュレーションでPH法とLi法を比較した結果, 大きな違いはなかった.
- z_1 の推定精度を上げ, 期待症例数を抑えるために, 適切な解析時点および最大症例数を設定する必要がある.
 - 今回の想定のもとでは解析時点は50%, 最大症例数は3倍程度であった.



参考文献

- [1] Proschan, M. A and Hunsbarger, S. A. (1995). Designed extension of studies based on conditional power. *Biometrics* 51, 1315-1324.
- [2] Gang Li, Weichung J. Shih, Tailiang Xie and Jiang Lu (2002). A sample size adjustment procedure for clinical trial based on conditional power. *Biostatistics* 3, 2, 277-287.