

群逐次デザインの再考

浜口 和人¹⁾、土屋 佳英¹⁾、田崎 武信¹⁾、森川 馨²⁾

1) 塩野義製薬株式会社 解析センター

2) 国立医薬品食品衛生研究所 安全情報部

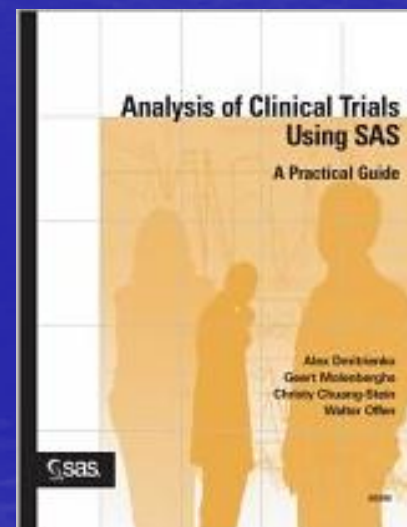
発表にあたって

SAS Pressより出版されている

Alex Dmitrienko, Geert Molenberghs, Christy Chuang-Stein, Walt Offen
(2005). *Analysis of Clinical Trials using SAS: A practical guide*.
SAS Institute Inc. Cary, NC, USA
Chapter 4 : Interim Data Monitoring(pp.177-267)

の第4章(群逐次デザイン) は中間解析を学ぶための入門書として非常に優れている。

中間解析の基本を丁寧に紹介
実践的なマクロプログラムの紹介
事例が豊富である



中間解析の実演

Dmitrienko et al.(2005) で紹介されてる複数の事例のうち

有効性または無益性のための早期中止
無益性の評価(確率的打ち切り検定)

を目的とした中間解析の一連の流れをたどり、紹介されているSASマクロをどのように使うかの例を示す。

事例1：重症敗血症患者を対象にした臨床試験

主要エンドポイントは28日後の生存率(生存率には2項分布を仮定)
プラセボを対照としたDBT試験(2群)
優越性を示すことが目的

1回目(20%集積時):主として無益性
2回目(66%集積時):有効性と無益性

使用するマクロ

%EffFutDesign

計画時の有意水準などの算出

%EffFutMonitor

解析時の有意水準などの算出

%EffFutDesignによって、例数の見積もり、計画段階における(大枠の)有意水準を算出できる。

	試験薬群 [76%を想定]	プラセボ群 [70%を想定]	有意水準 Efficacy	有意水準 Futility
	合計例数	合計例数		
中間解析1 (20%集積)	216	216	0.0000	0.4469
中間解析2 (66%集積)	711	711	0.0108	0.1004
最終解析 (100%集積)	1078	1078	0.0309	0.0309

有効性の有意水準 O'Brien Fleming法
無益性の有意水準 Pocock法としている

ところが、中間解析は計画どおりには進まない。

- 症例のエントリーは不規則
- 有意水準の再見積もり(消費関数に基づく算定)

	試験薬群	プラセボ群	有意水準 Efficacy (計画時)	有意水準 Futility (計画時)
	合計例数	合計例数		
中間解析1 (20%集積)	220(216)	216(216)	0.0000 (0.0001)	0.4469 (0.4428)
中間解析2 (66%集積)	-	-	-	-
最終解析 (100%集積)	-	-	-	-

%EffFutMonitorにより
解析段階の症例数に基づき
有意水準を算出

()内 計画段階での症例数と有意水準
()外 解析段階での症例数と有意水準

1回目(約20%)の段階で中間解析を実施したところ・・・

	試験薬群	プラセボ群	検定結果 p値	有意水準 Efficacy	有意水準 Futility
	生存数/合計例数	生存数/合計例数			
中間解析 1	164/220 (74.5%)	152/216 (70.4%)	0.1646	0.0001	0.4428

判 定 方 法

検定p値 < 有意水準Efficacy : 有効中止決定！

有意水準Efficacy 検定p値 < 有意水準Futility : 試験継続

検定p値 有意水準Futility : 無効中止決定！

2回目(約66%)の中間解析時の症例集積状況

	試験薬群	プラセボ群	有意水準 Efficacy (計画時)	有意水準 Futility (計画時)
	合計例数	合計例数		
中間解析1 (20%集積)	220(216)	216(216)	0.0000 (0.0001)	0.4469 (0.4428)
中間解析2 (66%集積)	715(711)	715(711)	0.0109 (0.0108)	0.0988 (0.1004)
最終解析 (100%集積)	-	-	-	-

2回目 (約66%) の段階で中間解析を実施した。

	試験薬群	プラセボ群	検定結果 p値	有意水準 Efficacy	有意水準 Futility
	生存数 / 合計例数	生存数 / 合計例数			
中間解析 1	164/220 (74.5%)	152/216 (70.4%)	0.1646	0.0001	0.4428
中間解析 2	538/715 (75.2%)	526/715 (73.6%)	0.2336	0.0109	0.0988

判 定 方 法

検定p値 < 有意水準Efficacy : 有効中止決定！

有意水準Efficacy 検定p値 < 有意水準Futility : 試験継続

検定p値 有意水準Futility : 無効中止決定！

確率的打ち切り検定

条件付き検出力検定 (頻度論)

Lan-Simon-Halperin検定

Pepe-Anderson検定

Betensky検定

予測確率検定 (Bayes論)

予測検出力検定 (頻度論とBayes論)

検出力や予測確率がある基準値を下回った場合に
無効中止と判断する。

確率的打ち切り検定による無益性のモニタリングを実施 条件付検出力・予測確率 < 20%で中止とすれば

試験薬群 [76%を想定]			プラセボ群 [70%を想定]			条件付検出力 予測確率			
例数	生存者	生存率	例数	生存者	生存率	LSH	PA	Bayes1	Bayes2
55	33	60.0%	45	29	64.4%	55.5%	20.6%	11.1%	33.5%
79	51	64.6%	74	49	66.2%	47.4%	17.3%	10.9%	30.1%
101	65	64.4%	95	62	65.3%	37.8%	13.2%	8.6%	20.2%
117	75	64.1%	115	77	67.0%	16.4%	4.3%	1.8%	6.5%
136	88	64.7%	134	88	65.7%	14.3%	4.2%	2.4%	5.8%
155	99	63.9%	151	99	65.6%	3.5%	0.9%	0.4%	0.9%

LSH・・・Lan-Simon-Halperin法

PA・・・Pepe-Anderson法

Bayes1 生存率の事前分布に無情報事前分布(一様分布)を用いた

Bayes2 生存率の事前分布にベータ分布を用いた

■ 中止決定時

期待症例数の増分に及ぼす因子の探索

— %EffDesignマクロを用いた検討 —

目 的

%EffDesignを適用し、群逐次デザインでの期待症例数における固定標本デザインでの症例数からの増分に対して影響を及ぼす因子を探索する

期待症例数における固定標本デザインでの症例数からの増分[減少分]の定義

$$\delta \equiv \frac{\overbrace{\left\{ \sum_{k=1}^m N_k p_k + N_{\max} \left(1 - \sum_{k=1}^m p_k \right) \right\}}^{\text{期待標本サイズ (期待症例数)}} - 2 \overbrace{\left(\frac{Z_{\alpha/2} + Z_{\beta}}{\Delta} \right)^2}^{\text{固定標本サイズ}}}{2 \underbrace{\left(\frac{Z_{\alpha/2} + Z_{\beta}}{\Delta} \right)^2}_{\text{固定標本サイズ}}}$$

:効果サイズ

N_k : k回目の中間解析時の評価症例数

N_{\max} : 最大症例数

p_k : 対立仮説のもとでk回目の中間解析で試験が停止する確率

期待標本サイズと固定標本サイズの差を固定標本サイズで割ったものを百分率表示

%Eff F Designの概要: 計画時の有意水準などの算出

マクロの引数

FRACTION : 中間解析のタイミング (症例の集積率で与える)

EFFSIZE : 効果サイズ

POWER : 検出力

ALPHA : 片側の第1種の過誤

RHO : Wang-Tsiatis族デザインの停止境界の形状パラメータ

BOUNDARY : 計画した中間モニタリングにおける停止境界と停止確率
を保存するためのSASデータセット名

SIZEPOWER : 指定した効果サイズに対する逐次検定の検出力と平均
症例数を保存するためのSASデータセット名

%EffDesignの概要: 計画時の有意水準などの算出

得られる結果 (一部抜粋)

検定統計量の尺度上での片側停止境界、すなわちz統計量の停止境界

p値の尺度上での片側停止境界、すなわち片側p値の停止境界

指定した効果サイズに対応する対立仮説の下で、その時点の解析で試験が中止される確率

指定した効果サイズに対応する対立仮説の下で、その時点までの解析で試験が中止される累積確率

指定した効果サイズに対応する帰無仮説の下で、その時点の解析で試験が中止される確率

指定した効果サイズに対応する帰無仮説の下で、その時点までの解析で試験が中止される累積確率

%EffMonitor (解析時の有意水準などの算出) もほぼ同様

試験デザインの設定

2群の並行群間比較試験

群逐次デザインを検討

中間解析では有効中止のみを検討

デザイン時の想定効果サイズとモニタリング時の効果サイズが等しい

主要評価項目は連続量で正規分布に従う

検出力は80%、有意水準は片側2.5%

効果サイズは、0.20、0.30、0.40、0.50の4通り

停止境界はO'Brien-Fleming型、 Pocock型

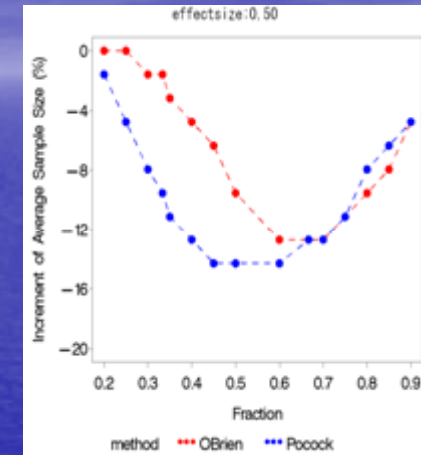
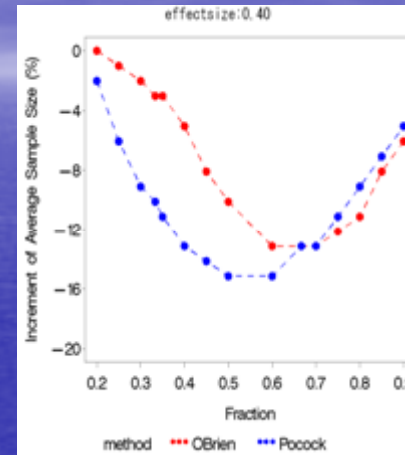
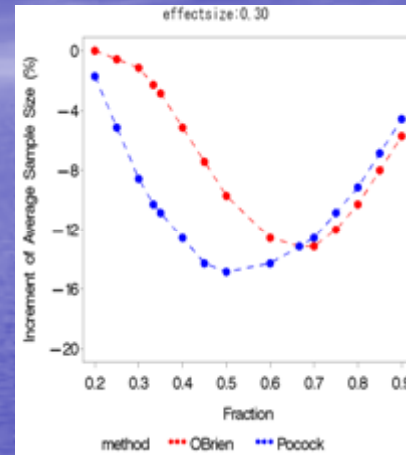
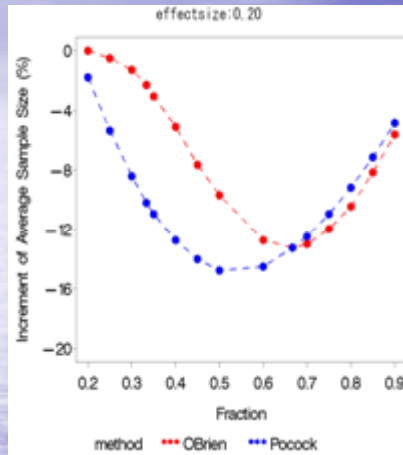
【検討1】1度の中間解析を計画するデザインにおける 中間解析のタイミング

1度の中間解析を計画するデザイン
中間解析を実施するタイミングを変化させ、タイミン
グが期待症例数の増分に及ぼす影響を検討。

【中間解析を実施するタイミング】

最大症例数に対する集積症例の割合が、
20%、25%、30%、33%、40%、45%、
50%、60%、67%、70%、80%、90%
の時点

【検討1】1度の中間解析を計画するデザインにおける 中間解析のタイミング



効果サイズ:0.20

0.30

0.40

0.50

O'Brien-Fleming法

Pocock法

効果サイズに関わらず、いずれも一峰性の傾向を示した。
O'Brien-Fleming法では70%の集積段階、Pocock法では50%
の集積段階で、減少分が最も大きかった。

【検討2】2度の中間解析を計画するデザインにおける 中間解析のタイミング

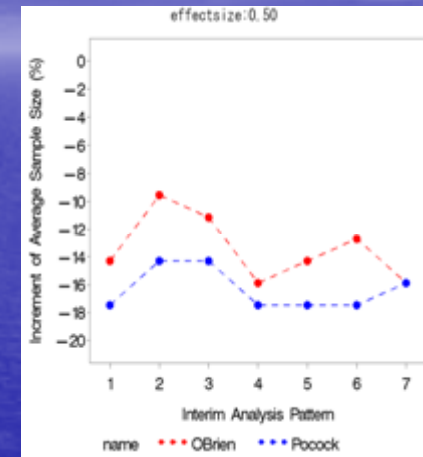
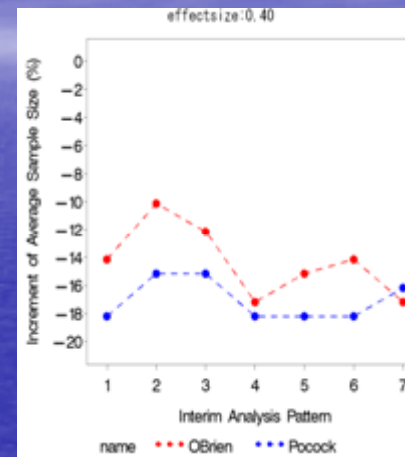
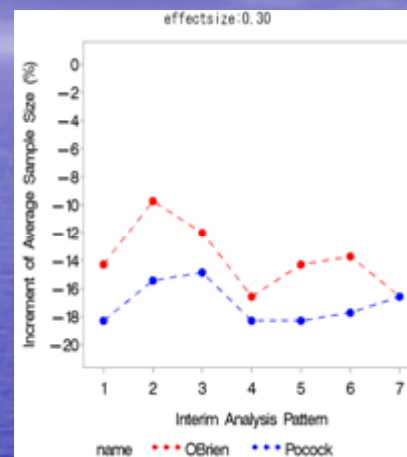
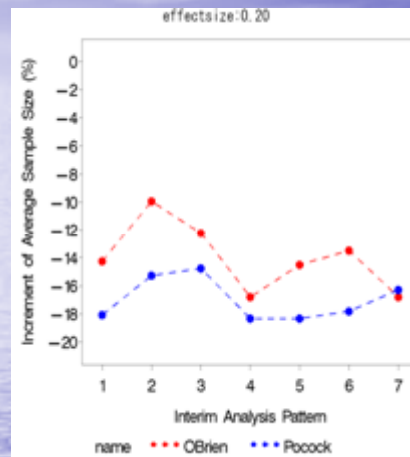
2度の中間解析

中間解析を実施するタイミングが期待症例数の増分に
及ぼす影響を検討する。

【中間解析を実施するタイミング】

最大症例数に対数する集積症例の割合が、
(第1回中間解析計画時、第2回中間解析計画時)の順に、
(33%、67%)、(25%、50%)、(25%、75%)、(50%、75%)、
(40%、60%)、(40%、80%)、(60%、80%)

【検討2】2度の中間解析を計画するデザインにおける 中間解析のタイミング



効果サイズ:0.20

0.30

0.40

0.50

パターン1: (33%, 67%), パターン2: (25%, 50%), パターン3: (25%, 75%), パターン4: (50%, 75%)
パターン5: (40%, 60%), パターン6: (40%, 80%), パターン7: (60%, 80%)

O'Brien-Fleming法

Pocock法

O'Brien-Fleming法とPocock法のいずれにおいてもパターン4すなわち集積率が(50%, 75%)の場合で減少分が最も大きくなった。

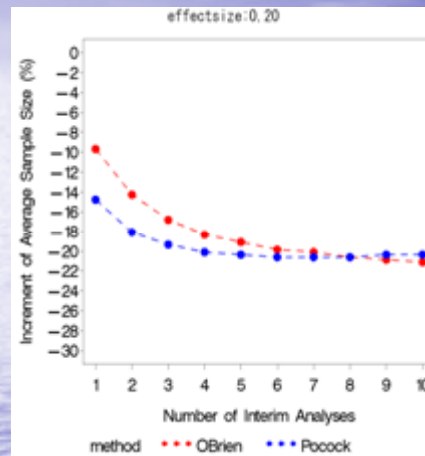
【検討3】 n 回の中間解析を等間隔で計画するデザイン における適切な n の値

n 回の中間解析

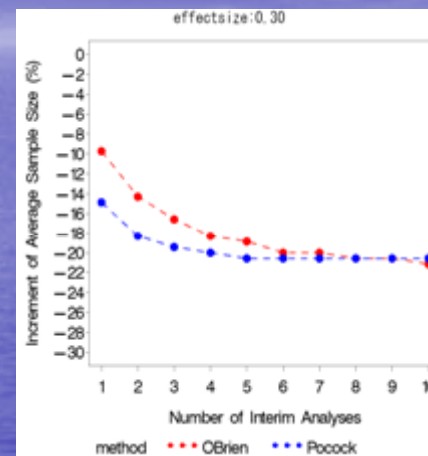
各中間解析は症例数において等間隔の時点で実施
 n が期待症例数の増分に及ぼす影響を検討する。

$$\{n \in \mathbb{Z} \mid 1 \leq n \leq 10\}$$

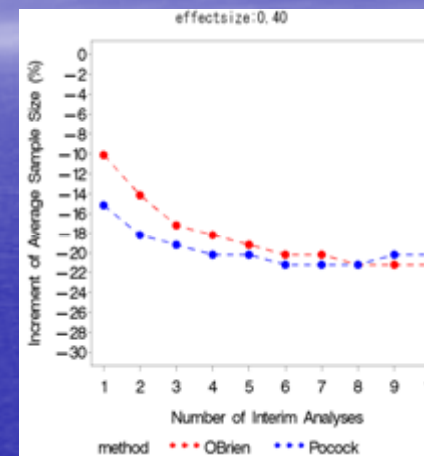
【検討3】 n 回の中間解析を等間隔で計画するデザインにおける適切な n の値



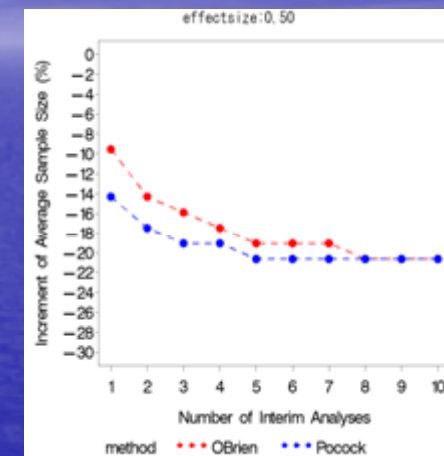
効果サイズ:0.20



0.30



0.40



0.50

O'Brien-Fleming法

Pocock法

O'Brien-Fleming法、Pocock法ともに中間解析の回数を5回以上実施してもそれほど減少分が大きくなるわけではない。

まとめ

1回の中間解析を計画した場合にはO'Brien-Fleming法では70%の集積段階Pocock法では50%の集積段階での減少分が最も大きかった。

2回の中間解析を計画した場合には、O'Brien-Fleming法 Pocock法ともに集積率が(50%, 75%)の場合で減少分が最も大きくなった。

等間隔で中間解析を行う際にはO'Brien-Fleming法、Pocock法ともに中間解析の回数は最大4回でよいのではないだろうか。

中間解析の回数を5回以上にしても、それほど減少分が大きくなるわけではない。

紹介されているSASマクロ(8種類)

群逐次検定

デザイン(計画)段階

%EffDesign %EffFutDesign

モニタリング(解析)段階

%EffMonitor %EffFutMonitor

確率的打ち切り検定

条件付き検出力

%CondPowerLSH %CondPowerPAB

予測検出力、予測確率

%BayesFutilityCont %BayesFutilityBin

URL: support.sas.com/companionsites

Dmitrienko et al.(2005)の 第4章 で紹介されてる内容

中間解析の目的と評価方法

有効性のための早期中止
群逐次検定

無益性のための早期中止
群逐次検定

確率的打ち切り検定

有効性または無益性のための早期中止
群逐次検定

群逐次検定

デザイン(計画)段階: Wang-Tsiatisファミリー

O'Brien-Fleming法

Pocock法

モニタリング(解析)段階: 過誤消費関数

群逐次検定

デザイン(計画)段階: Wang-Tsiatisファミリー

O'Brien-Fleming法

棄却域の端点を等しく設定

$$a_1 = a_2 = \Lambda = a_K = a$$

$$Z_{\alpha_k'/2} = \frac{a}{\sqrt{k}}$$

K : 中間解析の最大回数
k : k回目の中間解析時

初期の解析は保守的に実施

後期の解析は名目水準により近い有意水準で評価

群逐次検定

デザイン(計画)段階: Wang-Tsiatisファミリー

Pocock法

すべての中間解析での有意水準を等しく設定

$$\alpha'_1 = \alpha'_2 = \Lambda = \alpha'_K = \alpha'$$

K : 中間解析の最大回数

群逐次検定

デザイン(計画)段階: Wang-Tsiatisファミリー

=0.0 のとき O'Brien-Fleming法のデザイン

=0.5 のとき Pocock法のデザイン

$$Z_{\alpha_k/2} = c_{WT}(\alpha, \rho, K) k^{\rho-1/2}$$
$$[0 \leq \rho \leq 0.5]$$

$c_{WT}(\alpha, \rho, K)$ 各値はWang, Tsiatisの原著を参照

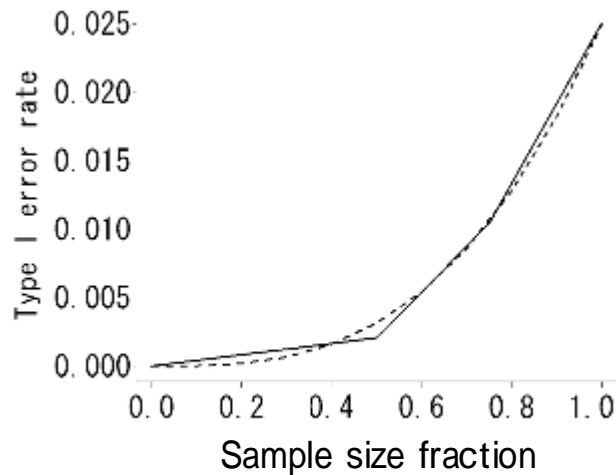
群逐次検定

モニタリング(解析)段階: 過誤消費関数

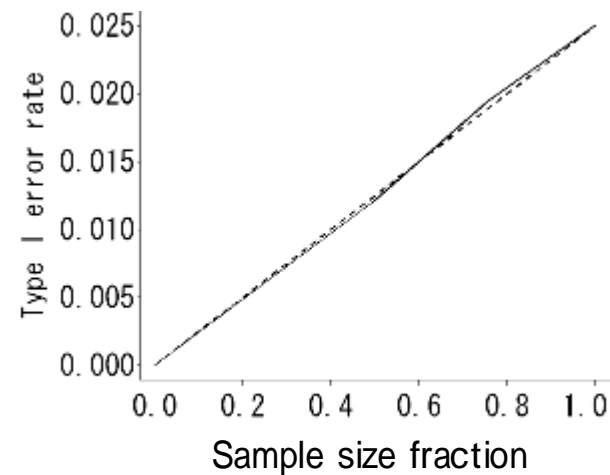
Lan-DeMets族

$$\alpha(t) = 2 - 2\Phi\left(z_{1-\alpha/2}/\sqrt{t}\right) \quad (O'Brien - Fleming)$$

$$\alpha(t) = \alpha \ln(1 + (e - 1)t) \quad (Pocock)$$



O'Brien-Fleming法



Pocock法

群逐次検定

モニタリング(解析)段階: 過誤消費関数

Jennison-Turnbull族

$$\alpha(t) = \alpha t^{\rho}$$

が 1 : Pocock法

が 3 : O'Brien-Fleming法

Hwang-Shih-DeCani族

$$\alpha(t) = \begin{cases} \frac{\alpha(1 - e^{-\rho t})}{(1 - e^{-\rho})} & \rho \neq 0 \text{ のとき} \\ \alpha t & \rho = 0 \text{ のとき} \end{cases}$$

が 1 : Pocock法

が -4, -5 : O'Brien-Fleming法

打ち切りのある指数分布型

確率的打ち切り検定

条件付き検出力検定 (頻度論)

Lan-Simon-Halperin法

n例の集積時点

N例の集積時点(最終時点)



Z_n が与えられた場合の Z_N の分布

$$\text{平均 } \sqrt{\frac{n}{N}}Z_n + \frac{\delta(N-n)}{\sigma\sqrt{2N}} \quad \text{分散 } \frac{N-n}{N}$$

: 対立仮説における効果の差

Z_n が与えられた場合の Z_N の条件付き検出力

に依存

$$P_n(\delta) = \Phi\left(\frac{\sigma\sqrt{n}Z_n + (N-n)\delta/\sqrt{2} - \sigma\sqrt{N}z_{1-\alpha}}{\sigma\sqrt{N-n}}\right)$$

確率的打ち切り検定

条件付き検出力検定 (頻度論)

Pepe-Anderson法 ($c = 1$)

Betensky法 ($c = 2.326$: 標準正規分布の99%点)

δ の代わりに $\hat{\delta}_n + cSE(\hat{\delta}_n)$ を用いる

$$\hat{\delta}_n = \sigma\sqrt{2/n}z_n$$

$$SE(\hat{\delta}_n) = \sigma\sqrt{2/n}$$

Z_n が与えられた場合の Z_N の条件付き分布

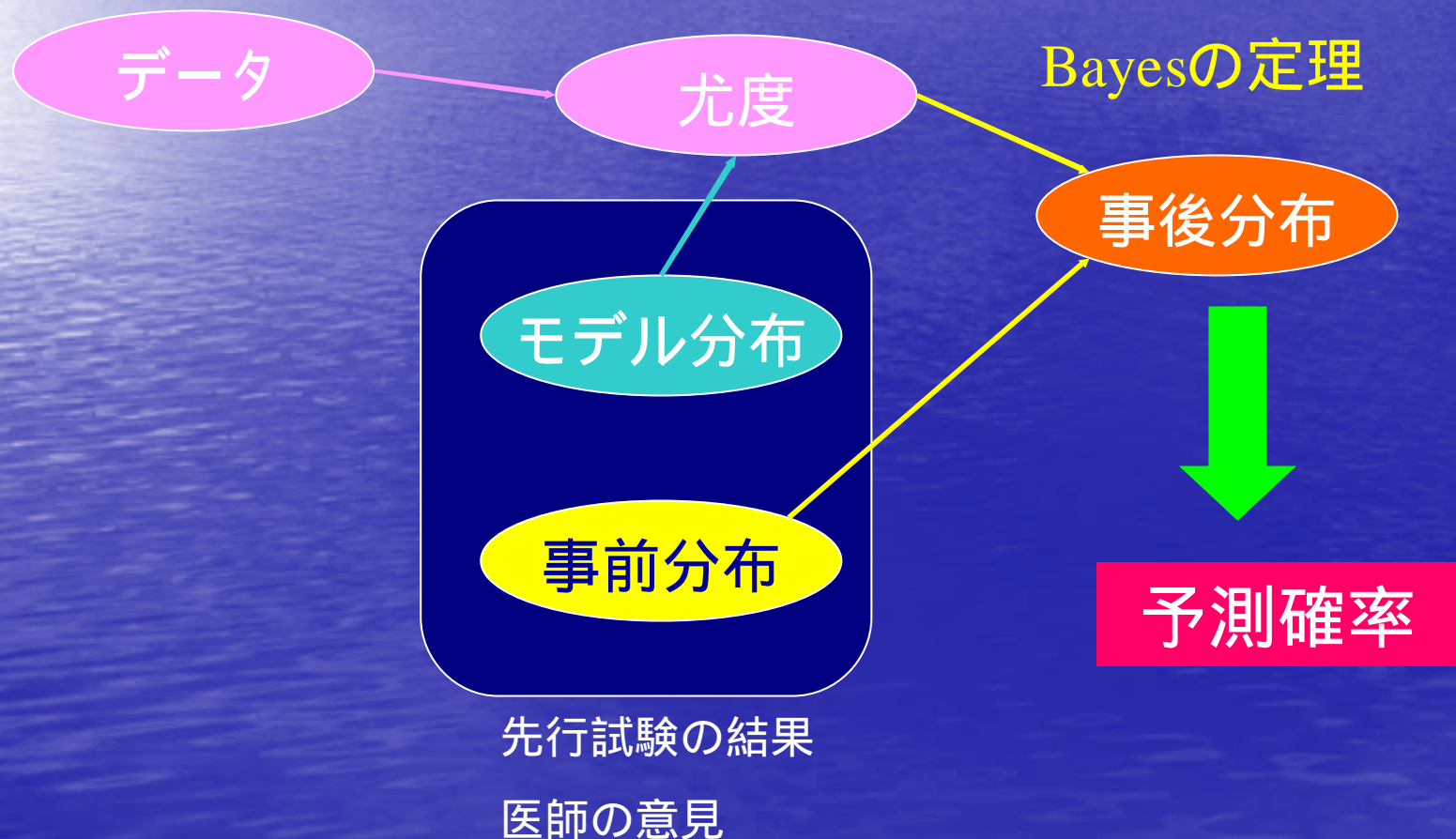
$$P_n(\delta) = \Phi\left(\frac{\sqrt{n}Z_n + (N-n)(Z_n + c)/\sqrt{n} - \sqrt{N}z_{1-\alpha}}{\sqrt{N-n}}\right)$$

にも
しない

にも依存

確率的打ち切り検定

予測確率検定 (Bayes論)



確率的打ち切り検定

予測検出力検定 (頻度論とBayes論)

$$P_n = \int P_n(\delta) f(\delta | Z_n) d\delta$$

条件付
検出力

事後分布

$$P_n(\delta) = P(\text{治療効果が}\delta\text{に等しいとき } Z_N > z_{1-\alpha} | Z_n)$$

$$= \Phi\left(\frac{\sigma\sqrt{n}Z_n + (N-n)\delta/\sqrt{2} - \sigma\sqrt{N}z_{1-\alpha}}{\sigma\sqrt{N-n}}\right)$$

条件付検出力

$f(\delta | Z_n)$ は、観測されたデータが与えられた場合での
治療差 の事後密度関数