

臨床試験における中間解析手法

α -spending function 法と Indifference-zone アプローチ

総合研究大学院大学

数物科学研究科統計科学専攻

小山 暢之

目次

1	序文	2
1.1	α -spending function 法による中間解析	3
1.2	Indifference-zone アプローチによる試験デザイン	5
2	仮説検定に基づく中間解析手法と indifference-zone アプローチ	6
2.1	中間解析時の検定統計量とブラウン運動	6
2.1.1	ブラウン運動	7
2.1.2	ブラウン運動の性質	8
2.1.3	中間解析時の検定統計量とブラウン運動	9
2.2	仮説検定に基づく中間解析手法	12
2.2.1	Pocock の方法	14
2.2.2	O'Brien and Fleming の方法	15
2.2.3	Pocock 法と O'Brien-Fleming 法の一般化	17
2.2.4	α -spending function 法	18
2.2.5	α -spending function の例	20
2.3	正確な棄却限界値	22
2.3.1	逐次 t -検定	22
2.3.2	二項確率の比較	24
2.4	α -spending function 法の適用上の問題点	27
2.4.1	Lan and DeMets のルール	27
2.4.2	Proschan らのルール	29

2.5	α -spending function 法の最適性	30
2.5.1	α -spending function としての最適性	31
2.5.2	Li and Geller の提案する α -spending function	33
2.5.3	α -spending function 法と解析時点	34
2.5.4	Pocock 法と解析回数	35
2.5.5	中間解析時点と被験者数	39
2.6	選択問題としての方法	44
2.6.1	Indifference-zone approach	45
2.6.2	Randomized Phase II 試験	46
2.6.3	対照群がある場合の試験デザイン	47
2.7	Appendix	54
2.7.1	定理 2.1 の証明	54
2.7.2	Armitage らによる第一種の過誤確率の評価法	56
2.7.3	O'Brien-Fleming 型の α -spending function の導出	58
2.7.4	2.3.1 節の π_k の算出方法 (Jennison and Turnbull)	59
2.7.5	定理 2.2 の証明 (Thall ら)	60
3	データに依存した中間解析計画への対応と解析時点の最適性	62
3.1	中間解析結果に依存した解析計画に対する棄却限界値	63
3.1.1	中間解析結果と解析時点の独立性	63
3.2	データに依存した解析計画に対する棄却限界値の算出方法	65
3.2.1	棄却限界値の算出例	72

3.3	解析時点と期待被験者数	73
3.3.1	最大被験者数を固定した場合	74
3.3.2	検出力一定のときの最大被験者数と期待被験者数	79
3.3.3	数値計算による検討結果	80
4	対照に対し許容的な治療法をすべて選択するための試験デザイン	83
4.1	Indifference-zone	84
4.2	One-stage デザイン	85
4.3	Two-stage デザイン	87
4.3.1	Two-stage デザインと選択確率	88
4.3.2	期待被験者数	90
4.4	数値計算による試験デザイン例と被験者数の比較	91
4.4.1	試験デザイン例	92
4.5	Weighted assignment デザイン	93
4.5.1	One-stage デザイン	94
4.5.2	Two-stage デザイン	94
4.5.3	Weighted assignment デザインでの選択ルールと被験者数	95
5	まとめ	96
5.1	仮説検定を基にした中間解析手法	97
5.2	Indifference-zone アプローチ	100

1 序文

臨床試験では、倫理的配慮から試験の途中でデータを統計学的に評価し、その後の試験の継続や中止、あるいは計画の変更などが行えるような試験デザインがしばしば必要となる。これは、有害反応や劣った治療法を受ける被験者を最小限に抑える一つの手段であり、可能な限り試験に参加している被験者の安全性および利益を確保することが大きな目的である。特に、試験治療に強い毒性が予想される場合や致命的疾患に対する治療法の試験、あるいは長期間にわたり追跡調査が必要な試験では多くの場合、定期的にデータを評価しながら被験者への利益や安全性を確認した上で試験が進められる。一般に、試験途中におけるデータの統計学的な評価を中間解析 (interim analysis) と呼ぶ。中間解析は、被験者側だけでなく、試験を行う研究者側にとっても利点がある。例えば、中間解析を行うことで試験が実際に計画通りに進められ、質の高いデータが得られているかどうかを確認することができる。また、期待した結果を得る見込みのない試験を不必要に長く継続せずにすみ、試験に必要な資源を有効に活用することが可能となる。

ただし、試験の途中で医師や被験者あるいは製薬企業などの試験依頼者が中間解析結果を知ることにより、その後の試験にバイアスが混入する可能性も指摘されている。例えば、治験担当医師の被験者のエントリー方法や効果の評価基準などが意識的かどうかに関わらず変化してしまったり、被験者からの副作用の申告頻度や服薬遵守状況が変わるなどの可能性がある。このため、中間解析を実施する際には、直接試験に関係しない独立な第三者で構成されたデータモニタリング委員会を組織し、その委員会内で非公開にデータを評価することが強く望まれる

1. データモニタリング委員会は試験が完全に終了するまで中間解析の結果を外部に開示せず、試験関係者には試験の継続や中止といった必要最小限の情報のみを提供することになる。一方、試験の途中で繰り返しデータを評価するということはそれだけ偶然の変動により誤った判断をしてしまう可能性も大きくなる。このため、統計学的方法により偶然変動による過誤を抑えることが要求されることになり、これまで頻度論や Bayes 理論の立場から様々な研究がなされてきた。

本研究では、仮説検定としての中間解析手法である α -spending function 法と統計学的選択問題としての indifference-zone アプローチを取り上げ、以下に示すような側面から、効率的に中間解析を実施するための指針を与えることを目的とする。

1.1 α -spending function 法による中間解析

臨床試験において中間解析を行う最も大きな目的は、その時点までに得られているデータに基づいて試験治療の効果に関する推測を行い、試験継続の可否を決定するということである。試験継続の可否を決定するための統計学的方法はいくつか考えられるが、一般に仮説検定の結果に基づいて行われることが多い。ただし、仮説検定を中間解析時点ごとに繰り返し行うと検定の多重性の問題が生じ、試験全体での第一種の過誤確率は各解析時の名義的な有意水準よりも大きくなる。このため、第一種の過誤確率をあらかじめ定めた有意水準以下に抑えるためには何らかの調整が必要となる。その一つの方法が α -spending function 法である²。この方法は、中間解析を行う時点や解析回数をあらかじめ固定する必要がなく、各解析時点ごとに棄却限界値を算出できるため実用的かつ柔軟な方法として現在でも広く用いられている。しかし、その柔軟さゆえにこの方法を適用する上での最適

性や問題点に関していくつかの論点がある。その一つが解析時点の最適性である。 α -spending function 法を用いた場合、有意水準は α -spending function の形状や解析時点に依存せずに制御できるが、検出力や試験終了時の期待被験者数(期待情報量)についてはこれらに依存して変化する。従って、検出力や試験の期待終了時点を考慮して効率的に中間解析を実施するためには、用いる α -spending function や解析時点も試験計画時に十分検討する必要がある。

また、解析時点の選択に関連する問題として α -spending function 法の「乱用」がある。 α -spending function 法は、解析時点が中間解析結果とは独立に決められることを前提としているため、解析時点が中間解析結果、つまり得られたデータに依存して選択される場合には必ずしも第一種の過誤の確率が事前に定めた有意水準以下にならない可能性が指摘されている^{3,4}。このため、中間解析結果に依存して解析回数や解析時点を変えることを α -spending function 法の「乱用」と呼ぶ。これまで乱用により第一種の過誤確率が極端に増大する例は報告されていないが、乱用に対する対処法も十分に検討されているとはいえず、実際どのように解析時点を決定すべきかの一般的な指針は与えられていない。

本稿では、 α -spending function 法を用いた場合の解析時点の最適性と乱用への対処方法について考察する。これらの問題に関し、これまで得られている結果を第2章(2.1~2.5節)で詳しく述べる。その後、第3章3.1節において、1度だけ中間解析結果に依存して解析時点が変更されるという単純な状況を考え、この状況で試験全体での有意水準を維持するための棄却限界値の算出方法を提案する。この方法により中間解析結果に依存して解析時点を決めても有意水準を保った棄却限界値を算出でき、「乱用」への一つの対処法となる。また、第3章3.2節では代表的

な α -spending function (O'Brien-Fleming 型, Pocock 型, 直線型) を用いた場合の解析時点の最適性について, 期待被験者数の観点から検討する. これらの結果は試験計画時および試験開始後のデータモニタリング委員会において, その後の中間解析時点を決定するための一つの指針を与えられられる.

1.2 Indifference-zone アプローチによる試験デザイン

医薬品開発の過程では複数の治療法の中から有効性もしくは安全性に関して一定の基準を満たす治療法を選択したいという状況がしばしば生じる. このような場合, それぞれの治療群に無作為に被験者を割り付けて臨床試験を行い, あらかじめ選択ルールとして定めた基準以上の有効性を示した治療法を選択するという方法が一般的である. 臨床試験では選択ルールとして仮説検定を用いることが多いが, 仮説検定よりも単純な選択ルールを定め, その選択ルールを用いたときに正しく選択できる確率が十分大きくなるように試験をデザインすることも考えられる. このような試験方法を選択問題 (selection problem) や選択確率 (selection probability) に基づく方法という. 選択問題では, 目的の条件を満たす治療法を選択できる確率が 0.8 や 0.9 といった大きな値になるように選択ルールを定めるが, この確率は興味あるパラメータに依存する. このため, 選択すべき群と選択すべきでない群のパラメータ値にはある一定値以上の差があるという仮定の下で選択確率の最小値を評価する方法がよく用いられる. これを indifference-zone アプローチという⁵.

本稿では, 臨床試験において試験治療の有効性が「有効」, 「無効」のような二値データで表される場合の選択問題について検討する. 二値データに対する選択問題を応用した臨床試験デザイン法はこれまで Simon⁶ や Thall ら⁷ などによりいく

つか提案されている(本稿 2.6 節). しかしながら, これらの方法は複数の試験治療の中から最も良いものを一つだけ選択することを前提としていることが多い. 一方, 新治療の開発過程においては複数の治療法の中から有望なものを見出すためのスクリーニングを目的とした臨床試験もしばしば実施され, このような場合は必ずしも一つの治療法のみを選択しなければならないわけではない. むしろ少数例で基準を満たすものをすべて選択できるような試験デザインが望ましいこともある.

このため, 本稿では第 4 章において, 対照治療に対し有効率が許容範囲以上にある試験治療を過不足なく選択するための one-stage および two-stage での試験デザイン法を提案する. Two-stage デザインでは, 試験の途中で一度中間解析を行い, その時点で対照治療よりも劣ると判断された試験治療があれば, その治療群は試験から削除することにする. これにより劣った治療を施される被験者数を少なくすることが期待でき, one-stage デザインよりも平均的に少ない被験者数で試験を実施できる可能性がある. 本稿では, one-stage デザインと two-stage デザインに対し必要な被験者数と選択ルールを数値計算により与え, two-stage デザインの利点について考察する.

2 仮説検定に基づく中間解析手法と indifference-zone アプローチ

2.1 中間解析時の検定統計量とブラウン運動

本節では, はじめに各中間解析時点ごとに算出される検定統計量の推移が確率過程として解釈できることについて述べ, その確率過程が独立増分構造を満たす

つか提案されている(本稿 2.6 節). しかしながら, これらの方法は複数の試験治療の中から最も良いものを一つだけ選択することを前提としていることが多い. 一方, 新治療の開発過程においては複数の治療法の中から有望なものを見出すためのスクリーニングを目的とした臨床試験もしばしば実施され, このような場合は必ずしも一つの治療法のみを選択しなければならないわけではない. むしろ少数例で基準を満たすものをすべて選択できるような試験デザインが望ましいこともある.

このため, 本稿では第 4 章において, 対照治療に対し有効率が許容範囲以上にある試験治療を過不足なく選択するための one-stage および two-stage での試験デザイン法を提案する. Two-stage デザインでは, 試験の途中で一度中間解析を行い, その時点で対照治療よりも劣ると判断された試験治療があれば, その治療群は試験から削除することにする. これにより劣った治療を施される被験者数を少なくすることが期待でき, one-stage デザインよりも平均的に少ない被験者数で試験を実施できる可能性がある. 本稿では, one-stage デザインと two-stage デザインに対し必要な被験者数と選択ルールを数値計算により与え, two-stage デザインの利点について考察する.

2 仮説検定に基づく中間解析手法と indifference-zone アプローチ

2.1 中間解析時の検定統計量とブラウン運動

本節では, はじめに各中間解析時点ごとに算出される検定統計量の推移が確率過程として解釈できることについて述べ, その確率過程が独立増分構造を満たす

場合にブラウン運動で近似できることを示す。試験の進行に伴う検定統計量の推移がブラウン運動で近似できることにより、各解析時点での棄却限界値を比較的容易に計算することができる。また、それぞれの棄却限界値に対する検出力も評価することが可能になる。さらに具体的な棄却限界値の算出方法として Pocock 法、O'Brien and Fleming 法および Lan and DeMets の α -spending function 法について述べ、それぞれの方法の特徴や相違点について示す。

2.1.1 ブラウン運動

まず、確率過程とブラウン運動について定義する。

定義 2.1 確率過程

ある実数値の集合 T に対し、確率変数の集合 $\{W(t), t \in T\}$ を確率過程という。つまり、任意の $t \in T$ に対し、 $W(t)$ が常に確率変数であるとき、 $W(t)$ を T における確率過程という。

T が可算集合であれば、 $W(t)$ を離散過程といい、 T が連続集合であれば $W(t)$ を連続過程という。

定義 2.2 定常増分構造と独立増分構造

連続過程 $\{W(t), t \in T\}$ において、

$$t_0 < t_1 < t_2 < \cdots < t_n$$

を満たすすべての $t_0, t_1, t_2, \dots, t_n \in T$ に対し、

$$W(t_1) - W(t_0), W(t_2) - W(t_1), \dots, W(t_n) - W(t_{n-1})$$

が互いに独立であるとき, $\{W(t), t \in T\}$ は独立増分構造 (independent increments) を満たすという.

また, $t+s \in T$ となる任意の s に対し,

$$W(t+s) - W(t)$$

の分布がすべての t に対して同じであるとき, $\{W(t), t \in T\}$ は定常増分構造 (stationary increments) を満たすという.

定義 2.3 ブラウン運動

次の条件を満たす確率過程 $\{W(t), t \geq 0\}$ を標準ブラウン運動という.

1. $W(0) = 0$
2. $\{W(t), t \geq 0\}$ は定常かつ独立増分構造を満たす.
3. すべての $t > 0$ に対し, $W(t)$ は平均 0, 分散 t の正規分布に従う.

また, 次の条件を満たす確率過程 $\{W(t), t \geq 0\}$ をドリフト μ をもつブラウン運動という.

1. $W(0) = 0$
2. $\{W(t), t \geq 0\}$ は定常かつ独立増分構造を満たす.
3. すべての $t > 0$ に対し, $W(t)$ は平均 μt , 分散 t の正規分布に従う.

2.1.2 ブラウン運動の性質

標準ブラウン運動 $W(t)$ について以下の性質が成り立つ.

1. 密度関数

$W(t)$ は t を固定すると平均 0, 分散 t の正規分布に従うため, $W(t) = x$ における密度関数 $f_t(x)$ は, 次のように表される.

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

また, 定常かつ独立増分構造を満たすことから, $(W(t_1), W(t_2), \dots, W(t_n))$ に対する同時密度関数 $f(x_1, x_2, \dots, x_n)$ は以下を満たす.

$$f(x_1, x_2, \dots, x_n) = f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \cdots f_{t_n-t_{n-1}}(x_n - x_{n-1})$$

つまり, $(W(t_1), W(t_2), \dots, W(t_n))$ は多変量正規分布に従う.

2. 共分散

任意の $s < t$ に対し,

$$\text{Cov}(W(s), W(t)) = s$$

となる.

次節ではこのブラウン運動と中間解析時の検定統計量の関係について述べる.

2.1.3 中間解析時の検定統計量とブラウン運動

臨床試験が継続されているある時点で試験治療の効果の有無や試験治療間の効果の差などを検証するため, それまでに得られているデータを用いて仮説検定を行うとする. 各解析時点 t_i , ($i = 1, 2, \dots$) における検定統計量を $X(t_i)$ とすると, $X(t_i)$ は確率変数であり, $\{X(t_i), i = 1, 2, \dots\}$ は確率変数の集合となる. 従って, 定義 2.1 から $\{X(t_i), i = 1, 2, \dots\}$ は t_1, t_2, \dots に対する一つの確率過程と考えられる. 一般に仮説検定を用いて中間解析を行う場合は, 検定統計量の値が棄却限界値を越えた時点で試験を終了し, 帰無仮説を棄却することになる. このため, 各解

析時点 t_i における棄却限界値の算出には、帰無仮説の下での $\{X(t_i), i = 1, 2, \dots\}$ の分布が必要となるが、ここでは、各解析時点間の情報量の増加がそれぞれ独立であるという独立増分構造を満たす場合に $\{X(t_i), i = 1, 2, \dots\}$ の分布がブラウン運動で近似できることを示す。

時間を表す変数を t とする。この t に対し、

$$X(t_1), X(t_2), \dots, X(t_i), \dots$$

を時間 t の経過とともに離散的に観察される確率変数列とし、それぞれ独立に平均 μ 、分散 1 の同一分布に従うとする。時間 t における確率変数の個数を $n(t)$ で表す。このとき、 $t < t'$ ならば $n(t) \leq n(t')$ が成り立つ。

ここで、

$$S(t) = \sum_{i=1}^{n(t)} X(t_i)$$

$$Z(t) = \sum_{i=1}^{n(t)} X(t_i) / \sqrt{n(t)}$$

と定義すると、 $S(t)$ 、 $Z(t)$ はそれぞれ以下の性質を満たす。

$$E[S(t)] = n(t) \cdot \mu$$

$$\text{Var}[S(t)] = n(t)$$

$$\text{Cov}[S(t), S(t')] = \min(n(t), n(t'))$$

$$E[Z(t)] = \sqrt{n(t)} \cdot \mu$$

$$\text{Var}[Z(t)] = 1$$

$$\text{Cov}[Z(t), Z(t')] = \begin{cases} [n(t')/n(t)]^{1/2}, & t > t' \\ [n(t)/n(t')]^{1/2}, & t \leq t' \end{cases}$$

一般に $S(t)$, $Z(t)$ のような性質を持つ確率過程を S -過程, Z -過程と呼ぶ⁸.

時間 t に対する試験期間を $[0, T]$ で表すと,

$$N = n(T)$$

は, N は試験期間内に集積される総データ数を表す. このとき,

$$B(n(t)/N) = S(t)/\sqrt{N}$$

を B -値と呼び, 次の性質を満たす⁹.

$$\begin{aligned} E[B(n(t)/N)] &= \frac{n(t) \cdot \mu}{\sqrt{N}} = \sqrt{N} \mu \left(\frac{n(t)}{N} \right) \\ \text{Var}[B(n(t)/N)] &= \frac{n(t)}{N} \\ \text{Cov}[B(n(t)/N), B(n(t')/N)] &= \min \left(\frac{n(t)}{N}, \frac{n(t')}{N} \right) \end{aligned}$$

従って, $B(n(t)/N)$ は上記の定義で μ を $\sqrt{N}\mu$, $n(t)$ を $n(t)/N$ に置き換えた S -過程と考えられる.

この $B(n(t)/N)$ を用い $\tilde{t} \in [0, 1]$ に対し,

$$B(\tilde{t}) = B(n(t)/N), \quad \tilde{t} \in \left[\frac{n(t)}{N}, \frac{n(t)+1}{N} \right)$$

と定義すると, この $B(\tilde{t})$ は $[0, 1]$ 上の連続な確率過程と考えることができる. Lan and Zucker は, この $B(\tilde{t})$ が近似的にブラウン運動に従うことを示している⁸.

定理 2.1 (Lan and Zucker)

$\mu = 0$ ならば, $B(\tilde{t})$ は $N \rightarrow \infty$ のとき標準ブラウン運動に従う. また, $\mu \neq 0$ ならば, $B(\tilde{t})$ は $N \rightarrow \infty$ のときドリフト $\sqrt{N}\mu$ をもつブラウン運動に従う. (2.6.1 節 参照)

この定理 2.1 から、標本平均など独立な観測値の和など S -過程から構成される検定統計量の推移の分布は適切な変数変換によりブラウン運動で近似できることがわかる。また、このブラウン運動における時間は次のように定義できる。

定義 2.4 Information fraction

定理 2.1 において各解析時点での $\tilde{t} = n(t)/N$ を information fraction(情報分数)もしくは information time(情報時間)と呼ぶ。一般に、検定統計量に対し各解析時点での Fisher 情報量と試験が最後まで行われ、すべてのデータが収集された(と仮定した)時点での Fisher 情報量の比が information fraction である¹⁰。

以上の結果から、中間解析時の検定統計量が S -過程から導かれる場合には各解析時の棄却限界値をブラウン運動の性質を用いて近似的に求めることができる。

次節ではこれらの結果を利用した仮説検定に基づく中間解析手法について述べる。なお、今後、時間を表す変数 t は information fraction を表すものとする。

2.2 仮説検定に基づく中間解析手法

仮説検定を用いて中間解析を行う場合に最も問題となることは検定の多重性である。データの蓄積に応じて検定を繰り返し、有意になった時点で試験を終了すると試験全体での第一種の過誤確率は各検定時の名義的な有意水準よりも大きくなる。このため、第一種の過誤確率が事前に定めた値以下になるように各検定時の名義的な有意水準の調整が必要となる。

仮説検定の枠組みで逐次的なデータの評価方法を臨床試験に最も早く取り入れようと試みたのは Armitage¹¹ である。Armitage は、Wald の sequential analysis¹² を二つの治療法を比較する臨床試験に応用し、それぞれの治療法をペア単位で被

験者に割り付けていき、一例のデータが得られるごとに解析を行う方法を提案した。しかしながら、治療後すぐに効果が現れるという状況は非常に少なく、データを一例ごとに解析するという Armitage の方法は実際の応用範囲が狭いため、普及するには至らなかった。その後、 Pocock¹³ や O'Brien and Fleming¹⁴ は、事前に定めた一定の例数のデータが集積されるごとに解析を行う group sequential デザインを提案した。彼らは2つの治療法を比較する試験において解析回数及び解析間隔を事前に固定することを前提に、試験全体での有意水準をあらかじめ設定した値に保つ棄却限界値の算出方法を示した。以後、臨床試験において group sequential デザインという概念が普及し、実際の試験デザインや解析に数多く用いられるようになった¹⁵。

さらに Lan and DeMets^{2, 16} は、データの蓄積に伴う検定統計量の推移が information fraction(情報分数)と呼ばれる量を時間とする確率過程に従うことを利用し、試験の経過に対する有意水準の消費率を示す α -spending function を指定するだけで、解析回数や解析間隔をあらかじめ固定することなく棄却限界値を算出できる方法を提案した。この α -spending function 法は、データの蓄積が時間に対して一定ではない臨床試験においては非常に柔軟な方法として受け入れられ、現在でも広く用いられている手法である。また、この方法は理論的に Pocock や O'Brien and Fleming が提案した方法も含むことから、仮説検定による中間解析手法の包括的なアプローチ法といえる。

本節では、これら代表的な中間解析による検定の多重性の調整方法について述べる。

2.2.1 Pocockの方法

Armitage¹¹は2つの治療法を比較する際に、各治療法から1例ずつ観測値が得られるごとに対をつくり、その対ごとに解析を行う方法を提案している。しかしながらPocockは、一例一例のデータが得られるごとに解析をするという方法が、臨床試験では非常に限られた状況でしか適用できないと指摘し、ある一定の情報量が集積されるごとに解析を行う group sequential 法を提案した¹³。このPocockの方法は解析回数を固定し、等しい例数のデータが集積されるごとに検定を行うことを前提とし、各検定時の名義的な有意水準を実際の有意水準 α よりも小さくすることで、試験全体での第一種の過誤確率を α 以下に抑えるという方法である。

分散 σ^2 が既知で等しい2つの正規母集団の平均 μ_1, μ_2 の比較を考える。

$$H_0: \mu_1 = \mu_2 \quad (\text{帰無仮説})$$

$$H_1: \mu_1 \neq \mu_2 \quad (\text{対立仮説})$$

とし、

$$(X_1, X_2, \dots), (Y_1, Y_2, \dots)$$

をそれぞれの母集団からの標本とする。 n 例ずつの標本が得られるごとに仮説検定を行い、事前に定めた最大 K 回の検定のいずれかで有意差が認められれば、その時点で帰無仮説を棄却して試験を終えるとする。また、 K 回目の検定で有意差が認められなければ、帰無仮説を棄却できずに試験を終える。

\bar{X}_k, \bar{Y}_k ($k = 1, 2, \dots, K$) を $k - 1$ 回目の解析終了時から k 回目の解析時点までに得られた n 例のデータの標本平均とする。このとき、

$$d_k = \sum_{j=1}^k (\bar{X}_j - \bar{Y}_j) / k$$

は、平均 $\delta = \mu_1 - \mu_2$ 、分散 $2\sigma^2/(kn)$ の正規分布に従う。

$$Z_k = (\sqrt{kn})d_k/(\sqrt{2}\sigma), k = 1, \dots, K$$

を検定統計量とし、この Z_k に対し、毎回一定の棄却限界値 $z(K, \alpha)$ を用いるとする。つまり、

$$Z_k > z(K, \alpha)$$

のとき、帰無仮説 H_0 を棄却して試験を中止する。このとき、試験全体での第一種の過誤確率を α 以下にするためには、

$$\alpha = \Pr \{Z_k > z(K, \alpha) \text{ for some } k = 1, \dots, K \mid H_0\}$$

となるように $z(K, \alpha)$ を定めればよい。 $z(K, \alpha)$ の算出には多重積分が必要であるが、 Pocock は Armitage らが提案した Newton-Cotes 法により第一種の過誤確率の算出方法¹⁷を利用して $z(K, \alpha)$ を算出している。

Pocock 法における各中間解析時点の棄却限界値は Z_k に対して常に一定である。このため、最終解析時点での棄却限界値は中間解析を行わなかった場合に比べ、かなり保守的になるという特徴がある (表 1)。この性質は、対立仮説が正しい場合の試験の期待停止時間は小さくなる一方、多重性の調整による検出力の低下も大きいことが知られている。

2.2.2 O'Brien and Fleming の方法

O'Brien and Fleming は 2 つの母集団を特徴づける二項確率 p_1, p_2 の比較において、 Pocock 法と同様、あらかじめ解析回数を定め、それぞれの母集団から同数の標本が集積されるごとに繰り返し仮説検定を行う場合の棄却限界値の算出法を

提案している¹⁴。O'Brien-Fleming 法は

$$H_0: p_1 = p_2 \quad (\text{帰無仮説})$$

$$H_1: p_1 \neq p_2 \quad (\text{対立仮説})$$

に対し、各解析時点での検定統計量として χ^2 統計量をその時点までの解析回数と計画した総解析回数の比で重み付けたものを用いて棄却限界値を算出する方法であり、この点が Pocock の方法と異なる。

事前に定めた総解析回数を K とし、 k 回目の解析時点における帰無仮説のもとでの χ^2 統計量を $\chi^2(k)$ とする。検定統計量として、

$$W_k = \frac{k}{K} \chi^2(k)$$

を考えると、

$$W_k \sim \frac{k}{K} \left(\sum_{i=1}^k \frac{U_i}{\sqrt{k}} \right)^2 = \left(\sum_{i=1}^k U_i \right)^2 / K$$

となる。ここで、 U_i は標準正規分布 $N(0, 1)$ に従う確率変数である。

定理 2.1.1 より、 $\max \{W_k\}$ の分布はブラウン運動、 $\{X(t) \mid 0 \leq t \leq 1\}$ に対する、

$$\max \{[X(k/K)]^2\}, 1 \leq k \leq K$$

に等しい。このため、有意水準 α の棄却限界値 $b(K, \alpha)$ は、

$$\begin{aligned} \alpha &= \Pr \{W_k > b(K, \alpha) \text{ for some } k\} \\ &= \Pr \{[X(k/K)]^2 > b(K, \alpha) \text{ for some } k\} \end{aligned}$$

を解くことにより得られる。Pocock 法と同様、この棄却限界値 $b(K, \alpha)$ は Armitage らの積分計算法を利用して求めることも可能であるが、O'Brien-Fleming は 10000 回のシミュレーションにより $b(K, \alpha)$ を求めている (表 1)。

O'Brien-Fleming 法では、データ数が少ない時点での棄却限界値は非常に大きく、結果を保守的に解釈しなければならない。しかし、解析回数が増えるにつれて検定統計量に対する重みが増加し、最終的な解析時点での棄却限界値は多重性を調整しない場合の値に近くなることが知られている。このため多重性の調整による検出力への影響も少なく、最終的な解析結果が中間解析を行わなかった場合と異なる可能性も小さいという利点がある。

表 1: Pocock 法と O'Brien-Fleming 法の棄却限界値 (Geller¹⁸ から引用)

両側 $\alpha = 0.05$				
K	k	Pocock	OB-F	
3	1	2.289	3.471	
	2	2.289	2.454	
	3	2.289	2.004	
4	1	2.361	4.049	
	2	2.361	2.863	
	3	2.361	2.338	
	4	2.361	2.024	
5	1	2.413	4.562	
	2	2.413	3.226	
	3	2.413	2.634	
	4	2.413	2.281	
	5	2.413	2.040	

ただし、値は標準正規分布の%点である。

2.2.3 Pocock 法と O'Brien-Fleming 法の一般化

Pocock 法と O'Brien-Fleming 法は、それぞれ正規分布の平均値および二値データの発現率の二群比較において、解析回数を固定し、同数のデータが集積されるごとに仮説検定を行うための棄却限界値の算出法である。それぞれ検討した統計量は異なるが、両者の本質的な違いは、検定の際に用いる統計量の分布を確率過

程としてどのようにとらえるかということである。Pocock 法は検定統計量として標準正規分布の z 値を用いるため、その分布に 2.1.2 節で示した Z -過程 $Z(t)$ が仮定されており、O'Brien-Fleming 法は χ^2 統計量をその時点での情報量と最終的な情報量の比で割ったものを検定統計量とすることから B-値を検定統計量として用いていることになる。それぞれ、 z 値、B-値、に対して棄却限界値が一定となる方法である。従って、2.1.2 節で示したように検定統計量の推移を確率過程としてとらえブラウン運動 $B(t)$ で近似すると、それぞれの方法は次の条件式から棄却限界値の列 b_p, b_o を求めることに拡張できる。

$$\alpha = \Pr\{B(t_i)/\sqrt{t_i} > b_p \text{ for some } i = 1, \dots, K \mid H_0\} \quad (\text{Pocock 法})$$

$$\alpha = \Pr\{B(t_i) > b_o \text{ for some } i = 1, \dots, K \mid H_0\} \quad (\text{O'Brien-Fleming 法})$$

ここで、 H_0 は帰無仮説を表す。これらの確率の評価は Armitage らの積分計算法で可能であり、解析間隔や解析回数はあらかじめ定める必要はあるが、解析間隔は必ずしも同数の標本が得られるごとでなくてもよい。

2.2.4 α -spending function 法

Pocock や O'Brien and Fleming の方法は臨床試験に group sequential 法という柔軟な試験デザインと解析方法を広め、その後の大規模試験や長期試験の試験デザインや解析方法に大きな影響を与えた。しかしながら、これら 2 つの方法はいずれも解析回数や解析間隔(時点)を事前に定めなければならないという制約がある。臨床試験では、被験者のエントリーが不定期にならざるを得ないことや、中止や脱落の発生も避けられない。このため、統計学的な情報量の増加と物理的な時間の経過は必ずしも比例せず、常に一定の被験者数のデータを集積するごとに中

間解析を行うことは非常に難しいといえる。また、何らかの理由で解析回数を増減させることもこれら2つの方法では対応できない。このため、Lan and DeMets はこれらの方法をさらに発展させ、解析時点や解析回数に依存せずに第一種の過誤確率が調整可能な方法として α -spending function 法を提案した²。

2.1.2節で示したように検定統計量が独立な確率変数の部分和で構成される場合、その過程は information fraction を時間とする $[0, 1]$ 上のブラウン運動で近似することができる。この性質を利用し、次のような手順で各中間解析時点ごとの棄却限界値の列 $\{b(t_i), i = 1, 2, \dots\}$ を定めることが可能である。

帰無仮説の下で検定統計量は、information fraction t に対し確率過程 $B(t)$ に従うとする。はじめに、次の条件を満たす $\alpha^*(t)$ を事前に一つ任意に定める。

- $\alpha^*(t) : t \in [0, 1]$
- $\alpha^*(0) = 0, \alpha^*(1) = \alpha$
- $\alpha^*(t)$ は単調非減少

この $\alpha^*(t)$ を用い、次の関係式を満たすように各解析時点 t_i における棄却限界値 $\{b(t_i), i = 1, 2, \dots\}$ を $b(t_1)$ から逐次定める。

$$\Pr\{B(t_j) \leq b(t_j), j = 1, \dots, i-1, B(t_i) > b(t_i)\} = \alpha^*(t_i) - \alpha^*(t_{i-1})$$

ただし、 $t_0 = 0$ とする。

このとき、全体での第一種の過誤確率は、

$$\Pr\{B(t_i) > b(t_i) \text{ for some } i\}$$

$$\begin{aligned}
&= \sum_i \Pr \{B(t_j) \leq b(t_j), j = 1, \dots, i-1, B(t_i) > b(t_i)\} \\
&= \sum_i [\alpha^*(t_i) - \alpha^*(t_{i-1})] \\
&= \alpha
\end{aligned}$$

となる。従って、 $\{b(t_i), i = 1, 2, \dots\}$ は有意水準を α に保つ一つの棄却限界値の列となっていることがわかる。

さらにこの方法は、各解析時点 i での棄却限界値は $B(t)$ の分布と、 $\alpha^*(t)$ およびそれまでの解析時点 $j = 1, \dots, i-1$ にのみ依存し、その後の解析回数や解析時点には依存しない。このため、事前に解析回数や解析時点 (解析間隔) を定める必要がなく、どの時点で解析を行ってもよい。 $\alpha^*(t)$ を α -spending function と呼び、この手順で各中間解析時点での棄却限界値を定める方法を α -spending function 法と呼ぶ。

一般に、 α -spending function 法では、検定統計量の推移が必ずしもブラウン運動に従う必要はなく、その分布が特定できればよい²。また、 $\alpha^*(t)$ は単調非減少で $\alpha^*(0) = 0$, $\alpha^*(1) = \alpha$ という条件さえ満たせば任意に設定することができる。このため、 $\alpha^*(t)$ や解析回数、解析時点を変えることにより様々な棄却限界値の列を与えることができ、有意水準は常に α に保たれる。

2.2.5 α -spending function の例

前節で述べたように α -spending function は非常に緩い条件の下で任意に選択できる。 α -spending function として最も簡単な関数は直線 αt であるが、代表的なものとして Pocock や O'Brien-Fleming が提案した方法に近い棄却限界値を与える関数がよく知られている。これらの関数は以下の式で表すことができる。

- O'Brien-Fleming 型

$$\alpha_1^*(t) = 2 \left[1 - \Phi \left(z_{\alpha/2} / \sqrt{t} \right) \right]$$

- Pocock 型

$$\alpha_2^*(t) = \alpha \log \{ 1 + (e - 1)t \}$$

- 直線型

$$\alpha_3^*(t) = \alpha t$$

$\alpha_1^*(t)$ は O'Brien-Fleming 法の棄却限界値が B -値に対して一定になることから導くことができる。(2.6.2 節 参照)

一方, Pocock 型は Z -過程に対して棄却限界値が一定になる方法であり, 近似的に $\alpha_2^*(t)$ がそれに近い棄却限界値を与える. $\alpha_1^*(t)$ と $\alpha_2^*(t)$ はそれぞれ凸関数, 凹関数であり, 関数としては相反する性質を持つ. また, 算出される棄却限界値と検定統計量との関連は明確ではないが, 凸関数, 凹関数という観点から考えると直線 $\alpha_3^*(t)$ は凸関数でもあり, 凹関数でもある.

これらの関数以外にも α -spending function はいろいろ考えられる. 例えば, Hwang らは 1 パラメータの指数分布族,

$$\alpha(\gamma, t) = \alpha \left[\frac{1 - e^{-\gamma t}}{1 - e^{-\gamma}} \right], (\gamma \neq 0)$$

を α -spending function として利用することを提案している¹⁹. この場合, γ の値により異なった棄却限界値を構成できるが, その性質についての理論的な裏付けはない.

2.3 正確な棄却限界値

Pocock 法では観察されるデータの分散は既知であることを前提としているが、分散が未知の場合も同様に棄却限界値を算出することが可能である。また、O'Brien-Fleming 法では二項確率の比較に χ^2 近似を用いているが、 α -spending function 法を用いれば Fisher の直接確率法を用いることが可能である。本節ではこれらの方法について述べる。

2.3.1 逐次 t -検定

Jennison and Turnbull は分散が未知の正規分布の平均値の仮説検定に対して group sequential 法で正確な棄却限界値を算出する方法を示している²⁰。

X_1, X_2, \dots は互いに独立で平均 μ , 分散 σ^2 の正規分布に従うとする。ただし、分散 σ^2 は未知である。このとき μ に対し次のような仮説に対する検定を group sequential 法で行うことを考える。

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

このとき、 $X_{n_{k-1}+1}, \dots, X_{n_k}$ を $k-1$ 回目の解析以降、 k 回目の解析時点までに得られた g_k 個の観察データとし、

$$n_k = g_1 + \dots + g_k$$

とする。

$$\bar{X}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} X_i, \quad s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (X_i - \bar{X}_k)^2$$

と定義すると t -統計量は、

$$T_k = \sqrt{n_k}(\bar{X}_k - \mu_0)/s_k$$

である。2.2節の α -spending function 法で示したように試験全体での有意水準を α に保つためには、次の関係式を満たすように棄却限界値の列 $\{c_1, c_2, \dots\}$ を定めればよい。

$$\alpha = \sum_k \pi_k$$

$$\pi_k = \Pr \{T_1 \leq c_1, \dots, T_{k-1} \leq c_{k-1}, T_k > c_k\}, \quad k = 1, 2, \dots$$

従って、 $\{c_1, c_2, \dots\}$ を与えるためには π_k を評価しなければならないが、これは以下の手順で帰納的に行うことができる。

$\bar{X}(k) = (X_{n_{k-1}+1} + \dots + X_{n_k})/g_k$ とすると、簡単な式変形により、

$$(n_{k+1} - 1)s_{k+1}^2 = (n_k - 1)s_k^2 + \sum_{i=n_k+1}^{n_{k+1}} \{X_i - \bar{X}(k+1)\}^2 + \frac{n_k n_{k+1}}{g_{k+1}} (\bar{X}_{k+1} - \bar{X}_k)^2$$

が成り立つ。ここで、

$$Z_k = n_k (\bar{X}_k - \mu_0) / \sigma, \quad R_k = (n_k - 1)s_k^2 / \sigma^2$$

と定義すると、 $\{Z_1, R_1, Z_2, R_2, \dots\}$ の同時分布は次のような条件付き分布から得られる。まず、 Z_1 と R_1 は互いに独立で、

$$Z_1 \sim N\{n_1(\mu - \mu_0)/\sigma, n_1\}, \quad R_1 \sim \chi_{n_1-1}^2$$

であることから、 $k \geq 1$ に対し、

$$\begin{aligned} \Pr \{Z_{k+1} \mid Z_1, R_1, \dots, Z_k, R_k\} &= \Pr \{Z_{k+1} \mid Z_k, R_k\} \\ &\sim N \{Z_k + g_{k+1}(\mu - \mu_0)/\sigma, g_{k+1}\} \end{aligned}$$

となる。従って、前述した式変形を用いると、

$$\Pr \{R_{k+1} \mid Z_1, R_1, \dots, R_k, Z_{k+1}\} = \Pr \{R_{k+1} \mid Z_k, R_k, Z_{k+1}\}$$

$$\sim R_k + \frac{n_k n_{k+1}}{g_{k+1}} \left(\frac{Z_k}{n_k} - \frac{Z_{k+1}}{n_{k+1}} \right)^2 + \chi_{g_{k+1}-1}^2$$

となり、これらの条件付き確率から $\{Z_1, R_1, \dots, Z_k, R_k\}$ の同時分布を構成できる。

$$T_k = \sqrt{n_k - 1} Z_k / \sqrt{n_k R_k}$$

であることから、 $\{T_1, \dots, T_k\}$ の同時分布も $\{Z_1, R_1, \dots, Z_k, R_k\}$ の同時分布から求めることが可能である。(2.7.4 節参照)

構成された同時分布は μ, μ_0, σ に依存するが、 $(\mu - \mu_0)/\sigma$ としてのみ依存するため、帰無仮説 $\mu = \mu_0$ の下での π_k は μ_0 や σ に関係なく計算可能である。つまり、 $\{c_1, c_2, \dots\}$ を与えれば、 π_k を求めることが可能であり、逆に $\{\pi_1, \pi_2, \dots\}$ を与えれば、 $\{c_1, c_2, \dots\}$ を求めることが可能である。Jennison and Turnbull はこの方法を用いて、 Pocock や O'Brien-Fleiming 型の各解析時の名義的有意水準を求め、その有意水準に対応する標準正規分布の%点を示している。表 2 は、Pocock 法で求めた各解析時の棄却限界値と Jennison らの方法で $g_k = 10$ (一定) のときに求めた棄却限界値 Z_p である。両者ともほぼ同じ値になるが、 Z_p の方がやや大きい値となることがわかる。

2.3.2 二項確率の比較

2.2.2 節で述べたように O'Brien and Fleming¹⁴ は反応が有効もしくは無効といった二値である場合の 2 治療間の有効率の比較を group sequential 法で行う方法を示しているが、その際、棄却限界値の算出には χ^2 近似を用いている。しかしながら、 α -spending function 法などを用いて各解析時点での名義的な有意水準を与えれば、Fisher の直接確率計算法を用いて正確な棄却限界値を算出することができる。

表 2: Pocock 法と正確な棄却限界値 (Jennison ら)

両側 $\alpha = 0.05$		
K	Pocock	Z_p
3	2.289	2.303
4	2.361	2.376
5	2.413	2.428
6	2.453	2.468
7	2.485	2.500
8	2.512	2.527
9	2.535	2.550
10	2.555	2.569

表 2 で K は総解析回数で, 棄却限界値は標準正規分布の%点である.

X_{ij} , $i = 1, 2$, $j = 1, 2, \dots$ を治療群 i の j 番目の被験者の反応を示す確率変数とする. $X_{ij} = 0$ (無効), $X_{ij} = 1$ (有効) であり,

$$\Pr\{X_{1j} = 1\} = p_1, \Pr\{X_{2j} = 1\} = p_2$$

と定義する. また, n_i をそれぞれの治療群に割り付けられた被験者数とし,

$$S_i = \sum_{j=1}^{n_i} X_{ij}$$

と定義する.

表 3: 2×2 表

	有効	無効		有効	無効	
治療 1	p_1	$1 - p_1$	治療 1	S_1	$n_1 - S_1$	n_1
治療 2	p_2	$1 - p_2$	治療 2	S_2	$n_2 - S_2$	n_2
	p	$1 - p$		c_1	c_2	

表 3 のような 2 つの治療群と「有効」「無効」の 2 つのカテゴリから構成され

る 2×2 表を考え、その周辺和を固定すると、

$$\Pr\{S_1 = s\} = \frac{\binom{c_1}{s} \binom{c_2}{n_1 - s}}{\binom{N}{n_1}}$$

である²¹。従って、有意水準を α とすると Fisher の直接確率法での S_1 に対する棄却限界値 q_1 は、

$$\sum_{s=q_1}^{n_1} \Pr\{S_1 = s\} \leq \alpha$$

を満たす最小の q_1 である。

解析回数を K 回とし、

$$\sum_{k=1}^K \alpha_k = \alpha$$

となるように各中間解析時の名義的有意水準 α_i , $i = 1, \dots, K$ を与える。 k 回目の解析時点での S_1 , q_1 をそれぞれ S_1^k , q_1^k で表すことにし、

$$\Pr\{S_1^1 \leq q_1^1, \dots, S_1^{k-1} \leq q_1^{k-1}, S_1^k > q_1^k\} = \alpha_i$$

となるように $k = 1$ から帰納的に q_1^k を求めれば、 $\{q_1^k, k = 1, \dots, k\}$ が各名義的な有意水準 $\{\alpha_k, k = 1, \dots, K\}$ に対する正確な棄却限界値となる。実際、上式の左辺は、

$$(S_1^1, S_1^2 - S_1^1, \dots, S_1^K - S_1^{K-1})$$

が互いに独立であることを利用することにより計算可能である。

α_k の列を α -spending function 法で与える場合、information fraction を定義する必要があるが、この場合、 χ^2 統計量を用いた場合の information fraction を用いればよい。 k 回目の解析時までには治療群 1 および 2 に割り付けられた被験者数がそ

れぞれ n_{1k}, n_{2k} であるとする、帰無仮説の下で $p_1 - p_2$ の分散は、

$$V_k(p_1 - p_2) = \left(\frac{1}{n_{1k}} + \frac{1}{n_{2k}} \right) p(1-p)$$

であるから、 k 回目の解析時点での information fraction t_k は、

$$t_k = V_k^{-1}/V_K^{-1} = \left(\frac{1}{n_{1k}} + \frac{1}{n_{2k}} \right)^{-1} / \left(\frac{1}{n_{1K}} + \frac{1}{n_{2K}} \right)^{-1}$$

である。

2.4 α -spending function 法の適用上の問題点

Lan and DeMets は、 α -spending function 法により棄却限界値を算出する場合には、その時点以降の解析時点や総解析回数情報は必要ないと述べている²。しかし、 α -spending function 法は必ずしも解析時点の選択方法に依存せずに第一種の過誤を調整できるわけではない。Lan and DeMets³ や Proschan ら⁴ が指摘しているように、それまでの中間解析結果に依存して次の解析時点が選ばれる場合には、 α -spending function 法で棄却限界値を算出しても第一種の過誤確率が増加する可能性がある。これは、解析時点も一つの確率変数と考えられ、中間解析結果に依存して解析時点が決められる場合には第一種の過誤確率に影響するからである。本節では、中間解析結果に依存して解析時点が選ばれた場合の α -spending function 法と第一種の過誤確率についてこれまで得られている結果について述べる。

2.4.1 Lan and DeMets のルール

Lan and DeMets は α -spending function 法を用いて棄却限界値を算出した場合に、検定統計量の値が有意ではないが、棄却限界値の 8 割の値を越えたときにそ

の後の解析回数を計画の2倍にしたときの第一種の過誤確率と検出力について検討している³.

information fraction t に対し, 検定統計量が構成する確率過程 $W(t)$ は, ドリフトパラメータ μ のブラウン運動に従うとし,

$$W(t) = B(t) + \mu t$$

と表す. 帰無仮説を $\mu = 0$ とし, 2.2.5節で示した O'Brien-Fleming 型の $\alpha_1^*(t)$ を用いるとすると, $Z(t) = B(t)/\sqrt{t}$ に対する各解析時点での有意水準両側 5% の棄却限界値 $c(t)$ は表 4 のようになる.

表 4: 棄却限界値 $c(t_i)$ (Lan and DeMets³)

両側 $\alpha = 0.05$, $\alpha(t) = \alpha_1^*(t)$		t_i					
	.25	.375	.5	.625	.75	.825	1
a.	4.33		2.97		2.36		2.02
b.	4.33	3.49	2.98	2.65	2.40	2.22	2.07
c.	4.33		2.97	2.64	2.40	2.22	2.07
d.	4.33		2.97		2.36	2.21	2.07

Lan and DeMets はこれら a, b, c, d の解析計画を以下のルールで選択した場合について検討している. つまり, 初めに等間隔で 4 回の解析を行うと計画しておき, 中間解析の結果が有意に近くなった場合にその後の解析間隔を 1/2 にするという解析計画である.

- b. if $.8c(.25) \leq Z(.25) < c(.25)$
- c. if $Z(.25) < .8c(.25) \ \& \ .8c(.5) \leq Z(.5) < c(.5)$
- d. if $Z(.25) < .8c(.25) \ \& \ Z(.5) < .8c(.5) \ \& \ .8c(.75) \leq Z(.75) < c(.75)$
- a. その他

このような解析計画の下、O'Brien-Fleming型 $\alpha^*(1)$ および Pocock型 $\alpha^*(2)$ を用いて棄却限界値を求めたときの第一種の過誤確率をシミュレーションにより求めた結果が表5である。この結果から第一種の過誤確率はわずかではあるが有意水準を上回ることがわかる。

表 5: 棄却限界値を越える確率 (Lan and DeMets³)

α	$\alpha_1^*(t)$ の場合	$\alpha_2^*(t)$ の場合
.025	.025	.026
.050	.052	.051

2.4.2 Proschan らのルール

Proschan らは2回の中間解析を含め試験全体で3回の解析を行う場合に、1回目の中間解析結果から帰無仮説の下で2回目に有意になる確率が最も大きい解析時点を選んで解析を行った場合の第一種の過誤確率について検討している⁴。

帰無仮説の下で検定統計量が標準ブラウン運動 $B(t)$ に従うとし、1回目および2回目の解析時点を t_1, t_2 とする。棄却限界値の列 $\{b(t_i), i = 1, 2, 3\}$ に対し、

$$\begin{aligned} \varphi(b, t_2) = & \Pr\{B(t_2) > b(t_2) \mid B(t_1) = b\} \\ & + \Pr\{B(t_2) \leq b(t_2), B(1) > b(1) \mid B(t_1) = b\} \end{aligned}$$

と定義する。 $\varphi(b, t_2)$ が最大となる解析時点を $t_2 = t^*(b)$ とすると、このときの第一種の過誤確率は、

$$\alpha(t_1) + \int_{-\infty}^{b(t_1)} \varphi(b, t^*(b)) \frac{\phi(b/\sqrt{t_1})}{\sqrt{t_1}} db$$

と表される。ここで、 $\phi(x)$ は標準正規分布の密度関数である。

棄却限界値の列 $\{b(t_i), i = 1, 2, 3\}$ を 2.2.5 節で示した O'Brien-Fleming 型, Pocock 型, 直線型の α -spending function, $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ で求め, この第一種の過誤確率をシミュレーションで評価した結果が表 6 である。この場合もわずかながら第一種の過誤は有意水準よりも大きくなるが, 有意水準から大きく乖離することはないという結果であった。

表 6: 第一種の過誤確率 (Proschan ら⁴⁾)

t_1	α_1^*	α_2^*	α_3^*
$\alpha = .025$			
.1	.0261	.0272	.0268
.25	.0256	.0269	.0267
.5	.0260	.0260	.0261
$\alpha = .05$			
.1	.0528	.0537	.0529
.25	.0514	.0531	.0529
.5	.0520	.0517	.0518

2.5 α -spending function 法の最適性

これまで述べてきたように α -spending function 法の柔軟性は, α -spending function や解析時点の選択の任意性にある。しかし, 実際の試験計画時には第一種の過誤確率だけでなく, 検出力や早期終了の可能性などを考慮して解析時点をどのように選択すべきかということを議論することも重要である。 α -spending function 法を適用する上での最適性に関しては, どのようなタイプの α -spending function を用いるべきか, また, どのような解析回数や解析間隔で解析を行えばよいかと

ということが問題となる。Kim and DeMets²² や Li and Geller²³ は、検定統計量の大きさが同じ値であれば、より多くの情報量に基づく値、つまり試験がより進化した時点での検定統計量の値に比較上の重みがおかれるべきであるという考えを示している。この考え方に基づけば、各解析時に消費される α の量が information fraction の増加に対して単調増加であることが望ましいということになる。一方、解析回数の最適性に関しては McPherson が Pocock の方法で中間解析を行った場合について検出力や期待被験者数の観点から検討している²⁴。また、解析時点の最適性については Brittain and Bailey が、解析回数を 2 回及び 3 回としたときに何通りか解析時点を固定し、検出力を一定に保った場合と被験者数を一定とした場合の 2 通りの条件で期待被験者数の最小値とそのときの棄却限界値を検討している²⁵。

本節では、これらの α -spending function や解析時点の選択に関してこれまで得られている結果や提案について述べる。

2.5.1 α -spending function としての最適性

Slud and Wei²⁶ や Li and Geller²³ は、以下のような理由により中間解析に用いる棄却限界値の列は単調減少するものが望ましいと提案している。

- 解析回数が増加するにつれ、(その時点まで試験が継続されたという条件付きで) 試験が中止される確率が増加する。
- ある時点で有意になる値がその後の解析時点では有意にならないということがない。

この条件は棄却限界値の列を $b(t_1), b(t_2), \dots, b(t_i), \dots$ とすると,

$$b(t_1) \geq b(t_2) \geq \dots \geq b(t_i) \geq \dots$$

となることであるが, この条件を α -spending function 法において表現すると以下のようなになる.

$$\alpha^*(t_i) - \alpha^*(t_{i-1}) \geq \alpha^*(t_{i-1}) - \alpha^*(t_{i-2}) \text{ for every } i \quad (2.4.1)$$

しかし, 任意の t_{i-2}, t_{i-1}, t_i に対し, この条件が成り立つ α -spending function は存在しない. なぜならば, α -spending function が連続であるとする, $t_i - t_{i-1}$ を十分小さくとることにより, $\alpha^*(t_i) - \alpha^*(t_{i-1})$ を限りなく小さくできるからである.

また, α -spending function が不連続であれば, 不連続な時点では $\alpha^*(t)$ の値が定まらず, 棄却限界値を算出することができない²³.

従って, Li and Geller²³はこのSlud and Weiの条件を緩め, α -spending function として凸関数を用いることを提案している. つまり,

$$\frac{\alpha^*(t_i) - \alpha^*(t_{i-1})}{t_i - t_{i-1}} \geq \frac{\alpha^*(t_{i-1}) - \alpha^*(t_{i-2})}{t_{i-1} - t_{i-2}} \text{ for every } i \quad (2.4.2)$$

この(2.4.2)が成り立ち, 解析間隔 $t_i - t_{i-1}$ が i の増加に対して常に単調増加であれば(2.4.1)が成り立つ. しかしながら, 臨床試験の場合には解析間隔が物理的な時間に対して単調増加になるよう解析を行ったとしても information fraction の増分が必ずしも単調増加するとは限らず, 解析時点にはかなりの制約が必要となる.

2.5.2 Li and Geller の提案する α -spending function

さらに Li and Geller は α -spending function として次のような区分線形凸関数を用いることを提案している²³.

区間 $[0, 1]$ を2つの区間 $[0, t_0]$, $[t_0, 1]$ に分割し, 最初の区間で r_1 回, 後半の区間で r_2 回の解析を行うと仮定する. このとき,

$$t_1 = \frac{t_0}{r_1}, t_2 = \frac{2t_0}{r_1}, \dots, t_{r_1} = t_0$$

$$t_{r_1+1} = t_0 + \frac{1-t_0}{r_2}, t_{r_1+2} = t_0 + \frac{2(1-t_0)}{r_2}, \dots, t_{r_1+r_2} = 1$$

とし, 区分線形凸関数,

$$\alpha^*(t) = \begin{cases} at & 0 \leq t \leq t_0 \\ bt + c & t_0 < t \leq 1 \end{cases}$$

を α -spending function として定義する. ここで, a, b, c は次の関係式から求める.

$$a < b$$

$$at_0 = bt_0 + c$$

$$b + c = \alpha$$

$$b(1-t_0)/r_2 \geq \frac{at_0}{r_1}$$

これだけの条件からは, a, b, c は一意に定まらないが, Li and Geller は初期の解析で多くの α を消費したい場合には a を大きめにとり, そうでない場合には a を小さめにとって b, c を決めると提案している. このように α -spending function を設定することにより, 後半の解析に残しておきたい α の量をあらかじめ明確にしておくことができ, また, 必ずしも事前に定めた解析時点 $t_1, t_2, \dots, t_{r_1+r_2}$ に従わ

なくとも解析間隔が単調増加であれば(2.4.1)の関係を満たす。しかしながら、この方法が具体的にどの程度の利点があるかという定量的な評価はなされていない。

2.5.3 α -spending function 法と解析時点

中間解析に限らず、検定の多重性を考慮して第一種の過誤を調整すると検出力が低下する、つまり、得られる情報量が同じであれば、多重性の調整を行わない方が検出力が大きい²⁴。従って、検出力が同じであれば、中間解析を行ったときに必要な最大被験者数、つまり試験が途中で終了せずに最後まで継続されたときに必要となる被験者数は、中間解析を行わなかった場合に比べ大きくなる。その程度は、対立仮説や多重性の調整方法などに依存するが、検出力に影響する因子で制御可能なものについては可能な限りその影響を小さくすべきであると考えられる。特に臨床試験の場合には、倫理的な面から必要最小限の被験者数で試験を行うことが原則であり、また、実際に限られた被験者数で試験を実施しなければならないという制約がある。

一方、中間解析を行うことで試験を早期に終了できる可能性があり、必ずしも試験を最後まで実施しなくてもよい。従って、検出力を維持するために必要な最大被験者数が増大したとしてもそのような試験を繰り返した場合の平均的な被験者数は少なくなる可能性もある。これが sequential analysis の一つの利点である。従って、解析回数や解析時点を考慮する場合、単に検出力の低下を指標とするのではなく、期待被験者数も含めて検討することが重要である。

McPherson は Pocock 法を用いた場合の解析回数と期待被験者数について検討している²⁴。また、解析時点については Brittain and Bailey²⁵ が解析回数を 2 回も

しくは3回とし、それぞれの解析時点ごとに期待被験者数が最小となるような棄却限界値を求め、そのときの期待被験者数を Pocock 法や O'Brien-Fleming 法と比較している。本節では、この McPherson や Brittain and Bailey らが検討した解析回数や解析時点の最適性に関して述べる。

2.5.4 Pocock 法と解析回数

McPherson²⁴ は Pocock 法における解析回数と期待被験者数について検討している。Pocock 法は 2.2.1 節で述べたように、解析回数を事前に定め、統計的な情報量の増加に対して等間隔に解析したときに試験全体での有意水準が α となるように棄却限界値を与える方法であり、その棄却限界値は z 値に対して一定となる。

X_i, Y_i ($i = 1, 2, \dots$) をそれぞれ互いに独立な確率変数とし、

$$X_1, X_2, \dots \sim N(\mu_1, \sigma^2), \quad Y_1, Y_2, \dots \sim N(\mu_2, \sigma^2)$$

とする。ここで、 $N(\mu, \sigma^2)$ は平均 μ 、分散 σ^2 の正規分布を表し、 σ^2 は既知であるとする。また、帰無仮説、対立仮説をそれぞれ、

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 = \delta$$

とする。Pocock 法では、 n 例ずつの標本が得られるごとに最大 K 回の仮説検定を行い、いずれかの解析時に検定統計量の値が棄却限界値を越えれば、その時点で帰無仮説を棄却して試験を終える。もし、 K 回すべての検定で有意差が認められなければ、帰無仮説を棄却できずに試験を終えることになる。 \bar{X}_k, \bar{Y}_k ($k = 1, 2, \dots, K$) を $k-1$ 回目から k 回目の解析時点までに得られた n 例のデータの標本平均とする

と、 Pocock 法の棄却限界値 $z(K, \alpha)$ は,

$$Z_k = (\sqrt{kn})d_k/(\sqrt{2}\sigma)$$

に対し,

$$\alpha = \Pr \{Z_k > z(K, \alpha) \text{ for some } k = 1, \dots, K \mid H_0\}$$

となる値である。ここで d_k は,

$$d_k = \sum_{j=1}^k (\bar{X}_j - \bar{Y}_j)/k$$

である。(2.2.1 節および Appendix 2.6.1 節 参照)

表 7 に Pocock 法での解析回数と各解析時の名義的有意水準を示した。

McPherson はこの Pocock 法に対し、解析回数を 1 ~ 40 回に設定したときに検出力を一定に保つという条件の下で期待被験者数の観点から解析回数の最適性について論じている。彼は、 $\delta = 0.20$ のときに有意水準両側 5%、検出力 90%となる

表 7: Pocock 法での解析回数と名義的有意水準 (McPherson²⁴)

有意水準両側 5%の場合								
解析回数 :	40	20	10	8	5	4	2	1
有意水準 :	.0056	.0074	.0105	.0121	.0155	.0178	.0300	.0500

ために必要な最大被験者数とその最大被験者数で試験を実施した場合、実際の δ の値によって期待被験者数がどのように変化するかを検討し、表 8 の結果を示している。つまり、中間解析を行わない場合に、対立仮説 $H_1 : \mu_1 - \mu_2 = 0.20$ に対し、有意水準両側 5%、検出力 90%となるために必要な被験者数は 262 例であり、中間解析を行わないため δ の値に関わらず、常に被験者数は 262 例である。一方、解析

回数が5回の場合、同じ条件で必要な最大被験者数は317例であるが、仮定が正しく、 $\delta = 0.20$ であれば期待被験者数は174例で最大被験者数よりもかなり小さい値となる。

表 8: 解析回数と期待被験者数 (McPherson²⁴)

有意水準両側 5%, 検出力 90%								
解析回数 :	40	20	10	8	5	4	2	1
最大被験者数 :	361	349	333	323	317	309	285	262
$ \delta :$								
0.00 :	350	338	323	315	309	302	281	262
0.05 :	340	329	314	306	302	295	275	262
0.10 :	303	293	280	275	272	268	257	262
0.15 :	244	238	230	228	227	225	228	262
0.20 :	177	174	169	166	174	176	193	262
0.25 :	124	124	125	125	132	136	166	262
0.30 :	89	90	93	93	104	110	150	262
0.35 :	67	69	74	76	87	94	144	262
0.40 :	54	56	61	64	76	85	143	262

最大被験者数は解析回数に対して単調増加であり、期待被験者数が最小となる解析回数は δ の値に依存して決まることがわかる。また、 $|\delta|$ の値が小さいほど、解析回数が少ない方が期待被験者数は小さくなるが、逆に $|\delta|$ の値が大きければ、解析回数を増やした方が期待被験者数は小さくなる。ただし、最大被験者数の算出のために仮定した $\delta = 0.20$ の値が真値であれば、期待被験者数が最も小さくなる解析回数は8回であり、それ以上解析回数を増やしても多重性の調整が厳しくなり、検出力を確保するために必要な最大被験者数が増加してしまうことから期待被験者数もあまり低下しないことがわかる。

事前分布を考慮した場合の解析回数

McPherson は解析回数を選択するためにはこの δ の値の推定値がどの程度信頼できるかによるとして、 δ に対して次の 4 つの事前分布を仮定し、事前分布で重みづけた期待被験者数を基に解析回数を定めることも提案している。

1. $[-0.4, 0.4]$ 上の一様分布 : $U(-0.4, 0.4)$
2. 平均 0, 標準偏差 0.2 の正規分布 : $N(0, 0.2^2)$
3. 平均 0, 標準偏差 0.1 の正規分布 : $N(0, 0.1^2)$
4. 平均 0.1, 標準偏差 0.1 の正規分布 : $N(0.1, 0.1^2)$

分布 1 は、 δ に対して全く情報がない場合の仮定である。分布 2,3 は δ が 0 を中心に分布しているという事前情報があり、分布 3 の方がより強い事前情報がある場合である。分布 4 は $\delta = 0.1$ という強い事前情報がある場合である。McPherson は期待被験者数をこれらの事前分布で重み付けした次の 2 つの値が最小となる解析回数についても検討している。

$$\int_{-\infty}^{\infty} \text{ASN}_{\delta} \text{Pr}(\delta) d\delta \quad (1)$$

$$\int_{-\infty}^{\infty} |\delta| \text{ASN}_{\delta} \text{Pr}(\delta) d\delta \quad (2)$$

ここで ASN_{δ} は $\mu_1 - \mu_2 = \delta$ のときの期待被験者数であり、 $\text{Pr}(\delta)$ は仮定した事前分布の密度関数である。

(2) 式は治療間差が大きい場合の被験者数を少なくするという観点から期待被験者数を治療間差 $|\delta|$ とその事前分布で重みをつけた値である。つまり、(2) 式では治療間差が大きければ $|\delta|$ の値が大きくなるため、それだけ ASN_{δ} に重みがかかり、全体の値を小さくするためには $|\delta|$ の値が大きい場合の ASN_{δ} をなるべく小さくする必要がある。

表6の結果に対してこの2つの関数で解析回数を評価した結果、(1)を用いた場合には事前分布 $3. N(0, 0.1^2)$ を除き中間解析の回数が2回から10回の間で期待被験者数が最小となり、また、事前分布が $3. N(0, 0.1^2)$ のときは中間解析を行わない場合に最も期待被験者数が小さくなるという結果が得られている。

(2)の場合はいずれも解析回数を増やせばそれだけ値は低下する傾向があったが、解析回数が4回から8回の間ではほぼ一定の値になる傾向が見られ、それ以上解析回数を増やしてもあまり顕著な低下は認められなかった。この結果から McPherson は、 δ に関して事前情報が乏しい場合は中間解析回数を5～10回に設定すべきだが、事前情報があるときでもせいぜい4～8回で十分であると指摘している。

2.5.5 中間解析時点と被験者数

Brittain and Bailey は解析回数を2回 (two-stage デザイン) および3回 (three-stage デザイン) のとした場合に、検出力を一定に保ち期待被験者数が最小となる解析時点とそのときの棄却限界値について検討している²⁵。

2つの治療法の比較試験を考える。各治療法に N 例ずつの被験者を割り付けるとし、興味のあるパラメータを δ とする。

$$\text{帰無仮説} : \delta = 0$$

$$\text{対立仮説} : \delta > 0$$

とし、各解析時点の検定統計量が帰無仮説の下で標準正規分布 $N(0, 1)$ に従い、対立仮説の下では平均 $\delta = \xi/\sqrt{N}$ ($\xi > 0$)、分散1の正規分布 $N(\delta, 1)$ に従うと仮定する。また、 N_f を中間解析を行わない場合に有意水準 α 、検出力 $1 - \beta$ となる被験者数とする。

解析回数が2回するとき、中間解析時の information fraction を t_1 とすると有意水

準が両側 α となる棄却限界値 $b(t_1)$, $b(1)$ は次の式を満たす。

$$\alpha = [1 - \Phi(b(t_1))] + [1 - \Phi(b(1))] - L(\sqrt{t_1}, b(t_1), b(1))$$

ここで、 $\Phi(x)$ は標準正規分布の分布関数である。また、 L は解析時点 $t = t_1$, 1 の 2 点でともに棄却限界値を越える確率で、

$$\begin{aligned} & L(\sqrt{t_1}, b(t_1), b(1)) \\ &= \frac{1}{2\pi} \int_{\arccos \sqrt{t_1}}^{\pi} \exp[-.5(b(t_1)^2 + b(1)^2 - 2b(t_1)b(1) \cos w) \operatorname{csec}^2 w] dw \end{aligned}$$

である。このとき検出力が $1 - \beta$ となるための $\delta = \delta^*$ は次の式から求められる。

$$\begin{aligned} 1 - \beta &= [1 - \Phi(b(t_1) - \sqrt{t_1}\delta^*)] + [1 - \Phi(b(1) - \delta^*)] \\ &\quad - L(\sqrt{t_1}, b(t_1) - \sqrt{t_1}\delta^*, b(1) - \delta^*) \end{aligned}$$

同様にして、解析回数が 3 回のとき中間解析時の information fraction を t_1 , t_2 とすると、

$$\begin{aligned} \alpha &= [1 - \Phi(b(t_1))] + [1 - \Phi(b(t_2))] + [1 - \Phi(b(1))] \\ &\quad - L(\sqrt{t_1/t_2}, b(t_1), b(t_2)) - L(\sqrt{t_1}, b(t_1), b(1)) - L(\sqrt{t_2}, b(t_2), b(1)) \\ &\quad + LL(\sqrt{t_1/t_2}, \sqrt{t_2}, b(t_1), b(t_2), b(1)) \end{aligned}$$

$$\begin{aligned} 1 - \beta &= [1 - \Phi(b(t_1) - \sqrt{t_1}\delta^*)] + [1 - \Phi(b(t_2) - \sqrt{t_2}\delta^*)] \\ &\quad + [1 - \Phi(b(1) - \delta^*)] - L(\sqrt{t_1/t_2}, b(t_1) - \sqrt{t_1}\delta^*, b(t_2) - \sqrt{t_2}\delta^*) \\ &\quad - L(\sqrt{t_1}, b(t_1) - \sqrt{t_1}\delta^*, b(1) - \delta^*) \\ &\quad - L(\sqrt{t_2}, b(t_2) - \sqrt{t_2}\delta^*, b(1) - \delta^*) \\ &\quad + LL(\sqrt{t_1/t_2}, \sqrt{t_2}, b(t_1) - \sqrt{t_1}\delta^*, b(t_2) - \sqrt{t_2}\delta^*, b(1) - \delta^*) \end{aligned}$$

となる。ここで LL は、

$$\begin{aligned}
 & LL(\sqrt{t_1}, \sqrt{t_2}, H, K, M) \\
 &= \int_{\Phi(K)}^1 \left[1 - \Phi \left(\frac{H - \sqrt{t_1/t_2} \Phi^{-1}(y)}{\sqrt{1 - (t_1/t_2)}} \right) \right] \\
 & \quad \cdot \left[1 - \Phi \left(\frac{M - \sqrt{t_2} \Phi^{-1}(y)}{\sqrt{1 - t_2}} \right) \right] dy
 \end{aligned}$$

である。

中間解析を行わなかった場合に検出力が $1 - \beta$ となるための必要被験者数が N_f であることから、中間解析を行った場合と同じ検出力を得るために必要な最大被験者数は、

$$N_f \cdot \left(\frac{\delta^*}{\delta} \right)^2$$

から求めることができる。さらに、この値に試験終了時点での information fraction の期待値を乗じたものが期待被験者数となる。

この関係式から解析時点を与えたときに期待被験者数が最も小さくなる棄却限界値とそのときの期待被験者数を求めることができる。ただし、解析時点は連続量であり、すべての解析時点の組み合わせについて最小値を求めることは不可能である。このため、離散的にいくつかの解析時点の組み合わせについて期待被験者数の最小値を求め、それらの間を何らかの方法で補間することにより期待被験者数が最小となる解析時点について議論することになる。

Brittain らは中間解析を行わなかった場合に有意水準両側 5%、検出力 $1 - \beta = 0.70$ のときに $N_f = 1000$ となる対立仮説を設定し、最大被験者数を 1010 例 ~ 1200 例に増やして中間解析を 1 回行い、同じ有意水準、検出力という条件で表 7 に示した期待被験者数の最小値を求めている。つまり、表 9 は解析回数が 2 回として最大

被験者数を固定し、それぞれの最大被験者数に対し $1 - \beta \geq 0.70$ という条件から期待被験者数の最小値を求めたものである。この結果から、期待被験者数が最小

表 9: 最大および期待被験者数 (Brittain ら²⁵)

$N_f = 1000/\text{群}$			
α	$1 - \beta$	最大	期待
.05	.70	1010	909
		1030	896
		1040	894
		1050	894
		1070	896
		1100	902
		1200	938

となる場合の最大被験者数は 1040~1050 例であることがわかる。さらに Brittain らは、そのときの解析時点が information fraction として約 0.6 であることを指摘している。

この 1050 例に対して中間解析を行わなかった場合の検出力は 0.721 であり、Brittain らはこの値とあらかじめ設定した検出力との比が $0.7/0.721 \simeq 0.97$ であることに着目している。彼らは、検出力が 0.8 や 0.9 の場合、および解析回数が 3 回の場合も期待被験者数が最小になる最大被験者数に対して同様の検出力の比 0.97 が認められるとし、この値に関して理論的な裏付けはないものの、最大被験者数を設定するための一つの指標として利用することを提案している。実際、この関係を用いて $N_f = 1000$ 、両側 $\alpha = 0.05$ の条件で検出力が 0.8, 0.9 のときの期待被験者数の最小値を計算すると表 10 のようになる。

また、これらの期待症例数の最小値が得られる解析時点については、表 11 で与

表 10: 最大被験者数と期待被験者数²⁵

両側 $\alpha = 0.05$, $N_f = 1000$ /群

検出力	最大	期待	
		Two-stage	Three-stage
0.70	1050	893.8	863.2
0.80	1070	847.9	806.5
0.90	1100	775.9	718.2

表 11: 期待被験者数を最小にする解析時点

両側 $\alpha = 0.05$

検出力	Two-stage	Three-stage
0.70	0.60	(0.50,0.75)
0.80	0.55	(0.43,0.70)
0.90	0.50	(0.38,0.65)

えられる近傍であることも指摘している。

この方法で算出された棄却限界値は期待被験者数を最小にするように最適化されていることから、 Pocock 法や O'Brien-Fleming 法を適用したときよりも期待被験者数に関しては小さくなる。しかし、 Brittain らも指摘しているように Pocock 法や O'Brien-Fleming 法と比較してもそれほど大きな差はなく、 O'Brien-Fleming 法に関しては Brittain らの方法よりも最大被験者数が小さいという利点がある。また、解析時点を固定して棄却限界値を最適化しているため、あらかじめ解析時点を固定しなければならず、 α -spending function 法のように解析回数や解析時点についての柔軟性はない。

表 12: Pocock 法, O'Brien-Fleming 法との被験者数の比較 (Three-stage)

両側 $\alpha = 0.05$		$1 - \beta$		
		0.70	0.80	0.90
Pocock	最大	29.13	36.62	48.36
	期待	22.04	25.70	30.31
O'Brien-Fleming	最大	25.15	31.94	42.71
	期待	22.09	26.88	33.57
Brittain-Bailey	最大	25.96	33.49	46.77
	期待	21.31	25.33	30.17

2.6 選択問題としての方法

複数の治療法, 例えば, 同じ疾患に対する複数の薬剤や同一の薬剤の異なる用法用量などを同時に比較して, その中からある条件を満たす治療法を選択したいという状況がある. このような状況は一般に様々な分野で生じる問題であり, 統計学的には選択問題 (Selection problem) として古くから研究されている分野である^{27, 28}. 選択問題では, 選択目標とそれに対する選択ルールを定め, そのルールの下で試験群を選択した場合に選択目標を達成する確率 (選択確率, Selection probability) があらかじめ定めた値以上の大きさなるように試験をデザインすることになる.

選択問題として試験をデザインする場合に 'indifference-zone アプローチ' が最もよく用いられる. 選択確率は対象とする集団の確率分布を規定するパラメータに依存するため, 選択問題ではパラメータの変動範囲内で選択確率がある一定値以上になるように試験をデザインすることになる. Indifference-zone アプローチではこのパラメータ配置に対して選択確率を考慮しない領域 (indifference-zone) を設け, その領域以外の範囲でパラメータが変動したときの選択確率の最小値を議論

する方法である。本節では、初めにこの indifference-zone アプローチについて定義し、この方法の臨床試験への応用例として、複数の治療法の中から最も成功割合の高い治療法を選択するための Simon ら⁶の randomize phase II trials と Thall ら⁷が提案した複数の試験治療の中から対照治療よりも優る治療法を選択するための two-stage デザインについて述べる。

2.6.1 Indifference-zone approach

t 個の治療群 T_1, \dots, T_t が興味あるパラメータ $\pi_i, i = 1, \dots, t$ で特徴づけられているとする。便宜上、 π_i の値が大きいほど好ましい治療法とし、次の関係が成り立っているとする。

$$\pi_1 \leq \dots \leq \pi_t$$

これらの治療群に対し、

$$\{T_i \mid \pi_i \geq \pi\}$$

を選択すべき治療群と定義する。この場合、Indifference-zone アプローチでは、定数 Δ ($0 < \Delta < \infty$), P^* ($0 < P^* < 1$) を事前に定め、 $\pi_j \geq \pi$ なる最小の j に対し、

$$\pi_1 \leq \dots \leq \pi_{j-1} \leq \pi - \Delta$$

ならば、 $\{T_i \mid \pi_i \geq \pi\}$ に含まれる T_i のみをすべて選択する確率 P_{CS} が常に、

$$P_{CS} \geq P^*$$

となるような選択ルールと必要情報量を定めることが目標となる²⁷。

これは、 $\pi - \Delta < \pi_i < \pi$ であれば、この治療法 T_i に関しては選択されてもされなくてもよいという立場に立つということであり、 $(\pi - \Delta, \pi)$ の範囲は実験者とし

て無関心な領域 (indifference-zone) と考える²⁸。従って、 Δ は試験の計画時に明確に定義することが重要であるが、どのように設定するかは試験の目的によって異なると考えられる。「臨床試験における統計的原則」¹では非劣性試験（試験治療が対照治療に対してある一定範囲以上有効性で劣らないことを示す試験）の非劣性域を設定する際には試験の目的と臨床的意義を考慮して、妥当と判断される範囲を事前に規定すべきであると述べられている。従って、indifference-zone アプローチを用いて臨床試験をデザインする場合もこの Δ を臨床的に許容できる、あるいは意味のない差の下限として与えることが重要であり、医学的見地を踏まえて試験ごとに設定される値であると考えられる。

2.6.2 Randomized Phase II 試験

抗癌剤に関しては各薬剤の単独投与時の効果だけでは不十分なことが多く、複数の抗癌剤の併用投与時の有効性についてもしばしば検討が行われる。このような場合、有効性が見込まれる併用療法を臨床試験で同時に検討し、検出力重視で最適な治療法を選択することが望まれる。このため、仮説検定よりも選択問題としての試験デザインが望ましいことも多い。Simon ら⁶は癌治療研究における無作為化臨床試験にこの indifference-zone アプローチを用いることを提案している。

反応が有効もしくは無効の二値データとして得られるとし、 t 個の治療群の中から最も有効率の大きい治療群を選択したいという状況を考える。それぞれの治療群に n 人ずつの被験者を割り付け、選択ルールとして最も有効数の多かった治療群を選択することにする。この際、最も有効数の多かった治療群が複数存在したときにはいずれか一つを等確率で選択する。それぞれの治療群の有効率を $\pi_i, i = 1, \dots, t$,

indifference-zone の幅を Δ とし, π_i が最大となる i がただ一つ存在するとしたとき, P_{CS} の下限値は, 次の条件が満たされたときの値である.

$$\pi_{[1]} = \cdots = \pi_{[t-1]} = \pi_{[t]} - \Delta$$

このようなパラメータ配置を ‘least favorable configuration’ (LFC) と呼び²⁸, このときの P_{CS} の値は,

$$P_{CS} = \sum_{j=0}^n \left[b(j; \pi_{[t-1]} + \Delta, n) \sum_{l=0}^t g(j, l) / (l + 1) \right]$$

$$g(j, l) = \binom{t-1}{l} [b(j; \pi_{[t-1]}, n)]^l [B(j-1; \pi_{[t-1]}, n)]^{t-1-l}$$

で与えられる⁶. ここで, $b(i; \pi, n)$, $B(i; \pi, n)$ はそれぞれ二項確率 π , 試行数 n の二項分布の確率関数および分布関数である. Simon らは P^* , Δ を定め, $\pi_{[t-1]}$ についてはその推定値を代入することにより,

$$P_{CS} \geq P^*$$

となるために必要な被験者数を求めることを提案している.

表 13 に $t = 2, 3, 4$, $\pi_{[t-1]} = 0.10, \dots, 0.80$ のときに必要な被験者数を示す.

治療群ごとに割り付けられた被験者数が異なる場合には有効数ではなく, 有効割合を用いることにより同様の方法で試験をデザインすることが可能である. このような試験デザイン法を Randomized phase II 試験と呼ぶ.

2.6.3 対照群がある場合の試験デザイン

臨床試験では, 複数の試験治療の中から最も有効で対照治療と比較しても優れた治療法を選択したいという状況がある. 例えば, 単一の治療法ではそれほど大き

表 13: 必要被験者数 (/群)(Simon ら⁶)

$\pi_{[t-1]}$	t		
	2	3	4
0.10	21	31	37
0.20	29	44	52
0.30	35	52	62
0.40	37	55	67
0.50	36	54	65
0.60	32	49	59
0.70	26	39	47
0.80	16	24	29

な有効性が見込めない癌の治療法を開発する場合、複数の治療法をいくつか組み合わせ、その組み合わせの中からそれまでの標準治療法よりも優れるものを選択するという方法が良く用いられる。しかし、試験を行う治療法の組み合わせがすべて標準治療に優るということはほとんどなく、すべての試験治療に対して対照治療法との検証的な比較試験を実施することは倫理的にも適切ではない。このような状況に対し、Thallらは試験治療法の中から最もよいものを選択するスクリーニング試験とその試験で選ばれた試験治療と対照治療法との比較試験を一つの試験で行う two-stage デザインを提案している⁷。彼らが提案した試験デザインは、初めに試験治療のみで試験を行い、その中で最も高かった治療法の有効割合がある値よりも大きければ、その治療法と対照となる標準治療法との検証試験を行うというものである。これを統計学的に定式化すると以下のようなになる。

試験治療法を T_1, \dots, T_t , 対照となる標準治療法を T_c で表し、それぞれの有効率を表すパラメータを π_1, \dots, π_t , および π_c とする。また, δ_1, δ_2 ($0 < \delta_1 < \delta_2$) をそ

れぞれ, $\pi_c + \delta_2$ が臨床的に意義のある有効率であり, $\pi_c + \delta_1$ が臨床的にはあまり意義のない有効率の向上であるとする. 目的は, 試験治療法の中から

$$\{T_i : \pi_i = \max(\pi_1, \dots, \pi_t), \pi_i \geq \pi_c + \delta_2\}$$

となる T_i を選ぶことである. これを次のような試験手順で行う.

- **first-stage**

- t 個の試験治療群にそれぞれ n_1 例ずつ被験者を割り付けて試験を行い, 最も大きかった治療法の有効割合が λ 以上であればその試験治療を選択する.

- **second-stage**

- 選択された試験治療 T_i と対照治療 T_c の比較試験を行い, 帰無仮説 $H_0 : \pi_i = \pi_c$, 対立仮説 $H_1 : \pi_i > \pi_c$ の片側検定を行う.

λ は標準治療と比較して優る見込みの小さい試験治療法をむやみに検証試験に移行しないようにするための限界値である. この λ により試験が first-stage で終了する可能性がある. このような two-stage の試験を次の条件が満たされるようにデザインする.

まず, π_c の値は未知であるため, その推定値 $\hat{\pi}_c$ を用い, first-stage において, $\{T_i : \pi_i \geq \hat{\pi}_c + \delta_2\}$ である T_i が少なくとも一つ存在し, $\{T_i : \hat{\pi}_c + \delta_1 < \pi_i < \hat{\pi}_c + \delta_2\}$ である T_i が一つも存在しないとした場合に $\{T_i : \pi_i \geq \hat{\pi}_c + \delta_2\}$ である T_i が選ばれる確率を $1 - \beta_1$ 以上となるように n_1 を定める. また, second-stage の仮説検定

は帰無仮説 $H_0 : \pi_i = \hat{\pi}_c$, 対立仮説 $H_1 : \pi_i = \hat{\pi}_c + \delta_2$ のときに有意水準 α , 検出力 $1 - \beta_2$ となるように行う。つまり, indifference-zone を,

$$\hat{\pi}_c + \delta_1 < \pi < \hat{\pi}_c + \delta_2$$

と設定したときに, 最終的に選択すべき T_i が選択される確率

$$1 - \beta = (1 - \beta_1) \cdot (1 - \beta_2)$$

が事前に定めた値になるように n_1, n_2, λ を定め, 試験をデザインすることになる。ただし, これだけの条件だけでは n_1, n_2, λ は一意に定まらないため, Thallらは,

$$\max_i \{\pi_i\} = \pi_c$$

のときの期待被験者数 $E(N)$ が最小となるように n_1, n_2, λ を定めることを提案している。Thallらは次の定理を示し, $1 - \beta$ の計算には

$$(\pi_1, \dots, \pi_t) = (\hat{\pi}_c + \delta_1, \dots, \hat{\pi}_c + \delta_1, \hat{\pi}_c + \delta_2)$$

のときのみを考えればよいと指摘している。

定理 2.2 (Thallら⁷⁾)

$$\{T_i : \pi_i \geq \hat{\pi}_c + \delta_2\}$$

を満たす T_i が少なくとも一つ存在し,

$$\{T_i : \hat{\pi}_c + \delta_1 < \pi_i < \hat{\pi}_c + \delta_2\}$$

を満たす T_i が一つも存在しないとした場合, $1 - \beta$ が最小となるのは,

$$(\pi_1, \dots, \pi_t) = (\hat{\pi}_c + \delta_1, \dots, \hat{\pi}_c + \delta_1, \hat{\pi}_c + \delta_2)$$

のときである。

証明 2.7.5 節 参照。

従って,

$$(\pi_1, \dots, \pi_t) = (\hat{\pi}_c + \delta_1, \dots, \hat{\pi}_c + \delta_1, \hat{\pi}_c + \delta_2)$$

が LFC であり, このときに治療群が正しく選択できる確率を考えればよい。

便宜上, $\pi_t = \max_i \{\pi_i\}$ とする。 X_i を T_i を割り付けられた群の有効数とし,

$$\hat{\pi}_i = X_i/n_1, \quad i = 1, \dots, t$$

とすると,

$$\begin{aligned} 1 - \beta_1 &= \sum_{x=x_0}^{n_1} \sum_{j=0}^{t-1} (j+1)^{-1} \binom{t-1}{j} \binom{n_1}{x} (\hat{\pi}_c + \delta_2)^x (1 - \hat{\pi}_c - \delta_2)^{n_1-x} \\ &\quad \times \left[\binom{n_1}{x} (\hat{\pi}_c + \delta_1)^x (1 - \hat{\pi}_c - \delta_1)^{n_1-x} \right]^j \\ &\quad \times B(n_1, \hat{\pi}_c + \delta_1; x-1)^{t-1-j} \end{aligned}$$

である。ここで, x_0 は λn_1 に等しいか, それよりも大きい最小の整数であり,

$$B(n_1, \hat{\pi}_c + \delta_1; x) = \sum_{y=0}^x \binom{n_1}{y} (\hat{\pi}_c + \delta_1)^y (1 - \hat{\pi}_c - \delta_1)^{n_1-y}$$

である。

second-stage での検出力 $1 - \beta_2$ は, 帰無仮説, 対立仮説がそれぞれ,

$$H_0 : \pi_t = \hat{\pi}_c$$

$$H_1 : \pi_t = \hat{\pi}_c + \delta_2$$

のときの値とする.

試験が second-stage まで継続される確率を ξ で表すと, 期待被験者数 $E(N)$ は次のようになる.

$$E(N) = tn_1 + 2\xi n_2$$

帰無仮説 $\pi_1 = \dots = \pi_t = \hat{\pi}_c$ の下での ξ の値を ξ_0 とすると,

$$\xi = \xi_0 = 1 - [B(n_1, \hat{\pi}_c; n_0 - 1)]^t$$

である. ここで, n_0 は λn_1 以上の最小の整数である. また,

$$\pi_1 = \dots = \pi_{t-1} = \pi_c + \delta_1, \pi_t = \pi_c + \delta_2$$

の下 (LFC) での ξ の値 ξ^* は,

$$\xi^* = 1 - [B(n_1, \hat{\pi}_c + \delta_1; n_0 - 1)]^{t-1} B(n_1, \hat{\pi}_c + \delta_2; n_0 - 1)$$

である. 試験デザインにおいては検出力を確保した上で, 期待被験者数が小さくなるようにしたいが,

$$E_0(N) = tn_1 + \xi_0 n_2$$

を最小にしようとする, first-stage での被験者数が大きくなってしまふ. また,

$$E^*(N) = tn_1 + \xi^* n_2$$

を最小にしようとする λ の値を負にとって second-stage を必須にしなければならなくなるが, これは $E_0(N)$ に対して定義した帰無仮説の下では全く意味がない. 従って, Thallらは次のような重み付き期待被験者数をデザインの最適化の指標とした.

$$E_\omega(N) = \omega E_0(N) + (1 - \omega) E^*(N)$$

従って、この場合の試験デザインとは $t, \hat{\pi}_c, \alpha, \beta, \delta_1, \delta_2, \omega$ を与えたときに、この $E_\omega(N)$ が最小となるように n_1, n_2, λ を定めることである。Thallらは、 $t = 2, 3, 4, \hat{\pi}_c = 0.2, 0.4, 0.6, \alpha = 0.05, \beta = 0.2, 0.25, 0.3, \omega = 1/2$ のときの n_1, n_2, λ を与えているが、 $\beta = 0.2$ のときの結果が表 14 である。

表 14: $\delta_1 = 0.05, \delta_2 = 0.20, \omega = 1/2$ のときの試験デザイン (Thallら⁷)

$\alpha = 0.05, \beta = 0.20$							
t	$\hat{\pi}_c$	n_1	n_2	λ	$E_\omega(N)$	$1 - \beta_1$	$1 - \beta_2$
2	0.2	40	99	.280-.300	190.0	.8869	.9021
	0.4	47	119	.490-.510	224.1	.8800	.9091
	0.6	39	99	.695-.715	187.3	.8875	.9014
3	0.2	45	111	.290-.310	256.6	.8580	.9324
	0.4	51	129	.495-.505	302.5	.8612	.9290
	0.6	43	107	.700-.720	251.7	.8661	.9237
4	0.2	48	120	.295-.310	325.9	.8441	.9478
	0.4	56	142	.505-.515	384.5	.8437	.9482
	0.6	47	113	.705-.720	319.2	.8545	.9362

β_1 が λ に対して階段関数になるので、 $E_\omega(N)$ を最小にする λ の値には幅がある。この表から、 λ と $\hat{\pi}_c$ には、

$$\lambda \simeq \hat{\pi}_c + 0.10$$

という関係があることがわかる。また、 $1 - \beta_1$ に比べ、 $1 - \beta_2$ の方が大きいという傾向があるが、その分 first-stage での被験者数が少なくなっており、通常のパイロット試験の規模に近い被験者数であることがわかる。

Thallらは $1 - \beta$ だけでなく、検証的試験である second-stage の検出力 $1 - \beta_2$ も固定した場合についても検討している。この場合、 $\beta_1 \geq \beta/\beta_2$ となるように λ, n_1 を定めていくことになる。 $1 - \beta = 0.80, 1 - \beta_2 = 0.80, 0.85, 0.90, \hat{\pi}_c = 0.20$ のとき

の結果を表 15 に示す。表 15 から、 $1 - \beta$ のみを固定した場合に比べ、 $1 - \beta_1$ の値

表 15: $\delta_1 = 0.05$, $\delta_2 = 0.20$, $\omega = 1/2$ のときの試験デザイン (Thall ら⁷⁾)

$\alpha = 0.05, \beta = 0.20, \hat{\pi}_c = 0.20$						
t	n_1	n_2	λ	$E_\omega(N)$	$1 - \beta_1$	$1 - \beta_2$
2	54	84	.260-.275	207.6	.9423	.80
	43	91	.260-.275	197.0	.9126	.85
	38	98	.265-.285	192.5	.8921	.90
3	68	84	.265-.275	304.2	.9420	.80
	58	91	.280-.290	277.4	.9151	.85
	51	98	.275-.290	267.9	.8953	.90
4	79	84	.280-.290	410.0	.9417	.80
	67	71	.285-.295	368.9	.9148	.85
	59	98	.290-.305	344.8	.8891	.90

が大きくなり、first-stage での被験者数が多くなる。それに伴い second-stage の被験者数が少なくなることがわかる。Thall らはこれらの結果から、この Two-stage デザインを採用する利点として、帰無仮説が正しい場合には高い確率で first-stage を終了した段階で試験を終えることができ、不必要に多くの被験者を試験に組み入れてしまう可能性を小さくできること及び無作為化した試験結果により比較試験に移行する治療群を選択するため、無作為化されないパイロット試験で治療群を選択する場合に比べ、結果の信頼性が高いことをあげている。

2.7 Appendix

2.7.1 定理 2.1 の証明

定理 2.1 (Lan and Zucker⁸⁾)

$\mu = 0$ ならば、 $B(\bar{t})$ は $N \rightarrow \infty$ のときブラウン運動に従う。

証明

まず, $B(\tilde{t})$ が $N \rightarrow \infty$ のときに平均 0, 分散 \tilde{t} の正規分布 $N(0, \tilde{t})$ に従うことを示す.

$$E[B(\tilde{t})] = 0 \quad (1)$$

は $B(\tilde{t})$ の定義から自明.

また,

$$\text{Cov}(B(\tilde{t}_1), B(\tilde{t}_2)) = \min \left\{ \frac{[N\tilde{t}_1]}{N}, \frac{[N\tilde{t}_2]}{N} \right\} \quad (2)$$

$$\rightarrow \min\{\tilde{t}_1, \tilde{t}_2\}, \quad (N \rightarrow \infty) \quad (3)$$

である. ここで, $[x]$ は x を越えない最大の整数を表す.

中心極限定理より

$$\frac{S(t)}{\sqrt{[N\tilde{t}]}} \sim N(0, 1), \quad (N \rightarrow \infty)$$

が成り立ち,

$$\frac{[N\tilde{t}]}{N} \rightarrow \tilde{t}, \quad (N \rightarrow \infty)$$

であることから,

$$B(\tilde{t}) = \frac{S(t)}{\sqrt{N}} = \frac{S(t)}{\sqrt{[N\tilde{t}]}} \cdot \frac{\sqrt{[N\tilde{t}]}}{\sqrt{N}} \sim N(0, \tilde{t}), \quad (N \rightarrow \infty) \quad (4)$$

となる.

さらに, $0 = \tilde{t}_0 < \tilde{t}_1 < \tilde{t}_2 < \dots < \tilde{t}_n < 1$ に対し,

$$B(\tilde{t}_1) - B(\tilde{t}_0) = \frac{1}{\sqrt{N}} X_{n(t_1)}$$

$$B(\tilde{t}_2) - B(\tilde{t}_1) = \frac{1}{\sqrt{N}} (X_{n(t_1)+1} + \dots + X_{n(t_2)})$$

⋮

$$B(\tilde{t}_n) - B(\tilde{t}_{n-1}) = \frac{1}{\sqrt{N}} (X_{n(t_{n-1})+1} + \dots + X_{n(t_n)})$$

はその定義から、それぞれ独立である。従って、 $B(\bar{t})$ は独立増分構造を満たす。

また、任意の s に対し、

$$B(\bar{t} + s) - B(\bar{t}) = \frac{1}{\sqrt{N}}(X_{n(t)+1} + \cdots + X_{n(t+s)})$$

であることから、 $B(\bar{t} + s) - B(\bar{t})$ は t に依存せず $N \rightarrow \infty$ で正規分布に従う。よって $B(\bar{t})$ は定常増分構造も満たす。

以上から、 $B(\bar{t})$ がブラウン運動に従うことがわかる。

同様にして、 $\mu \neq 0$ のときは、 $B(t) + \sqrt{N}\mu t$ (ドリフト $\sqrt{N}\mu$ をもつブラウン運動) で近似できることも示される⁸。

2.7.2 Armitage らによる第一種の過誤確率の評価法

$$X_1, X_2, \dots$$

をそれぞれ独立に標準正規分布 $N(0, 1)$ に従う確率変数列とする。

$$S_n = \sum_{i=1}^n X_i$$

と定義し、

$$|S_n| \geq y_n, \quad y_n = k\sqrt{n}$$

を最初に満たした時点で標本抽出を止めるとする。ただし、 k は定数とする。 S_n

の密度関数は次のように表される。

$$f_n(s_n) = \begin{cases} \int_{-y_{n-1}}^{y_{n-1}} f_{n-1}(u) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(s_n - u)^2\right\} du, & -y_n \leq s_n \leq y_n \\ 0, & \text{その他} \end{cases}$$

従って、 f_n は標準正規分布の密度関数である f_1 から帰納的に定義することが可能である。

$$m = \min\{n : |S_n| \geq y_n\}$$

とすると、 m の分布関数 $P_n = \Pr\{m \leq n\}$ は、

$$P_n = 1 - \int_{-y_n}^{y_n} f_n(u) du$$

であることから、

$$\begin{aligned} P_n - P_{n-1} &= 2 \int_{-y_{n-1}}^{y_{n-1}} f_{n-1}(u) \{1 - \Phi(y_n - u)\} du \\ &= 2 \int_{y_n}^{\infty} \int_{-y_{n-1}}^{y_{n-1}} f_{n-1}(u) \frac{1}{\sqrt{2\pi}} \left\{ \frac{1}{2}(v - u)^2 \right\} dv du \end{aligned}$$

と表すことが可能である。従って、 f_n を求めれば、 P_n を計算することが可能となる。

Armitage ら¹⁷ は f_n の近似計算に、区間 $[-y_{n-1}, y_{n-1}]$ を区間幅 h の微少区間に分割し、

$$\lambda_{n-1}h \leq y_{n-1} < (\lambda_{n-1} + 1)h$$

なる λ_{n-1} に対し、

$$-\lambda_{n-1}h \leq u \leq \lambda_{n-1}h$$

では、Newton-Cotes 法の二次の公式を用い、端点

$$\lambda_{n-1}h \leq |u| \leq y_{n-1}$$

には、シンプソンの3点公式を適用している。

中間解析時の棄却限界値を求めるためにはあらかじめ定めた有意水準 α に対し、次の方程式を y_n について解けばよい。

$$\alpha = P_n$$

2.7.3 O'Brien-Fleming 型の α -spending function の導出

よく知られているブラウン運動の性質として次の関係式が成り立つ^{29, 30}。

定理 2.3 標準ブラウン運動 $B(t)$ に対し、 T_a を $B(t)$ が初めて a に達するまでの時間を示す確率変数とする。このとき、

$$\begin{aligned} \Pr\{T_a \leq t\} &= 2 \Pr\{B(t) \geq a\} \\ &= \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/(2t)} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{a/\sqrt{t}}^\infty e^{-y^2/2} dy \end{aligned}$$

が成り立つ (ROSS²⁹)。

証明

$a > 0$ とすると、

$$\begin{aligned} \Pr\{B(t) \geq a\} &= \Pr\{B(t) \geq a \mid T_a \leq t\} \Pr\{T_a \leq t\} \\ &\quad + \Pr\{B(t) \geq a \mid T_a > t\} \Pr\{T_a > t\} \end{aligned}$$

ここで、 T_a の定義から $\Pr\{B(t) \geq a \mid T_a > t\} = 0$ であり、また対称性から、

$$\Pr\{B(t) \geq a \mid T_a \leq t\} = \frac{1}{2}$$

が成り立つ。従って、

$$\Pr\{B(t) \geq a\} = \frac{1}{2} \Pr\{T_a \leq t\}$$

となる。対称性から $a < 0$ の場合も同様である。

従って、次のことが成り立つ。

系 2.6.2

標準ブラウン運動 $B(t)$ が $[0, 1]$ 上で $a = z_{\alpha/2}$ ($z_{\alpha/2}$ は正規分布の上側 $\alpha/2$ %点) を越える確率 $\Pr\{T_{z_{\alpha/2}} \leq 1\}$ は、

$$\Pr\{T_{z_{\alpha/2}} \leq 1\} = 2\Pr\{B(t) \geq z_{\alpha/2}\} = \alpha$$

となる。また、 τ を $B(t)$ が初めて $z_{\alpha/2}$ を越える時点とすると、その分布、 $\alpha_1^*(t) = \Pr\{\tau \leq t\}$ は、

$$\alpha_1^*(t) = \begin{cases} 0 & (t = 0) \\ 2[1 - \Phi(z_{\frac{1}{2}\alpha}/\sqrt{t})] & (0 < t \leq 1) \end{cases}$$

となる。

従って、ブラウン運動 $B(t)$ に対して一定値の棄却限界値を与える α -spending function が $\alpha_1^*(t)$ であり、 B -値に対して棄却限界値が一定である O'Brien-Fleming 法に近い棄却限界値を与える。

2.7.4 2.3.1 節の π_k の算出方法 (Jennison and Turnbull)

棄却限界値の列 $\{c_1, \dots, c_k\}$ が与えられているとし、2.3.1 節で定義した π_k を以下のように求める。

$k \geq 2$ に対し、次のように定義する。

$$F_k(z, r) = \Pr\{Z_k \leq z, R_k \leq r, T_i < c_i \text{ for all } i = 1, \dots, k-1\}$$

$$f_k(z, r) = \partial^2 F_k / (\partial z \partial r)$$

$k = 1$ のとき,

$$f_1(z, r) = q_1(z)h_1(r)$$

とする。ただし、 q_1 は平均 $n_1(\mu - \mu_0)/\sigma$ 、分散 n_1 の正規分布の密度関数とし、 h_1 は $\chi_{n_1-1}^2$ の密度関数である。 $k \geq 1$ に対し、次のように f_{k+1} を構成できる。

$$f_{k+1}(z, r) = \int \int_{C_k} f_k(u, v) q_{k+1}(z|u, v) h_{k+1}(r|u, v, z) dudv$$

$$C_k = \left\{ (u, v) : v > 0, u < c_k \sqrt{n_k v / (n_k - 1)} \right\}$$

ここで、 $q_{k+1}(z|u, v)$ は $Z_k = u$ 、 $R_k = v$ を与えたときの Z_{k+1} の条件付き密度であり、 $h_{k+1}(r|u, v, z)$ は $Z_k = u$ 、 $R_k = v$ 、 $Z_{k+1} = z$ を与えたときの R_{k+1} の条件付き密度である。これら q_{k+1} 、 h_{k+1} は 2.3.1 節の方法で求めることができる。

以上から、 π_k は、

$$\pi_k = \int \int_{D_k} f_k(u, v) dudv$$

$$D_k = \left\{ (u, v) : v > 0, u \geq c_k \sqrt{n_k v / (n_k - 1)} \right\}, k = 1, 2, \dots$$

により正確な値が算出可能である。

2.7.5 定理 2.2 の証明 (Thall ら)

A は許容的な試験治療の番号の集合、つまり、

$$A = \{i : \pi_i \geq \hat{\pi}_c + \delta_2\}$$

とする。従って、

$$A \cup A^c = \{1, \dots, t\}$$

である。

$\pi=(\pi_1, \dots, \pi_t)$ に対し, 正しく治療法を選択する確率を P_{CS} とすると,

$$\begin{aligned} P_{CS} &= \Pr \{2\text{nd stage で帰無仮説を棄却} \mid 1\text{st stage で治療群を正しく選択}\} \\ &\quad \times \Pr(\text{first stage で治療群を正しく選択}) \\ &= (1 - \beta_2(\pi))(1 - \beta_1(\pi)) \end{aligned}$$

と表せる. ただし, β_2 は $\hat{\pi}_c$ と first-stage で選択された治療法の有効率にのみ依存する.

$|A| = m$, $|A^c| = t - m$ とし, $A^c \neq \emptyset$, つまり $t - m \geq 1$ とする. また, $t \in A$ とすると, LFC は次の条件を満たす.

$$\pi_1 \leq \dots \leq \pi_{t-m} \leq \hat{\pi}_c + \delta_1, \quad \hat{\pi}_c + \delta_2 \leq \pi_{t-m+1} \leq \dots \leq \pi_t$$

治療群 T_i の有効数を X_i とすると, $1 - \beta_1(\pi)$ は π の下で,

- $\max\{X_i : i \in A\} \geq \max\{\lambda, \max\{X_j : j \in A^c\}\}$
- 有効数が最大の治療群が複数あった場合には A に含まれる T_i が選ばれる.

という事象に対する確率である.

また, $\max\{X_i : i \in A\}$, $\max\{X_j : j \in A^c\}$ ともそれぞれ $\{\pi_i : i \in A\}$, $\{\pi_j : j \in A^c\}$ に対し確率的に増加する. 従って, $1 - \beta_1(\pi)$ は,

$$\pi_1 = \dots = \pi_{t-m} = \hat{\pi}_c + \delta_1, \quad \pi_{t-m+1} = \dots = \pi_t = \hat{\pi}_c + \delta_2$$

のとき最小となる.

次にこの条件の下で $m = 1$ のときに P_{CS} が最小になることを示す.

$m > 1$ とする. U_1, \dots, U_t を互いに独立に $[0, \epsilon]$, ($0 < \epsilon < 1$) 上の一様分布に従う確率変数とする. このとき, $Y_i = X_i + U_i$ と定義すると $1 - \beta_1(\pi)$ は次のように表現できる.

$$1 - \beta_1(\pi) = \Pr \{ \max\{Y_i : i \in A\} > \max\{\lambda, \max\{Y_i : i \in A^c\}\} \}$$

これは, X_i の最大値にタイがあったときにランダムに一つの治療群を選ぶ工夫の一つである. A に含まれる治療群が一つ A^c に移ったとすると $1 - \beta_1(\pi)$ は明らかに減少する. 従って, 帰納的に P_{CS} は $m = 1$ のときに最小となる.

なお, $1 - \beta_2(\pi)$ が $\pi_\nu = \hat{\pi}_c + \delta_2$ で最小となることは自明である.

3 データに依存した中間解析計画への対応と解析時点の最適性

α -spending function 法を用いて中間解析を行う場合, どのような時点でもそれまでに得られている情報量から試験全体での有意水準を維持した棄却限界値を求めることができる. この性質は, データの蓄積率が時間に対して一定ではない臨床試験のデータ解析では非常に有用である. しかし, 解析時点や回数について任意であるとはいえ, 得られたデータに依存してその後の解析時点が決められる場合には試験全体での有意水準を維持できない可能性も指摘されており, 解析時点の定め方が完全に任意であるというわけではない. また, 実際にどの時点で中間解析を行うべきかという最適性の議論は, 最適性の指標と用いる α -spending function に依存する. 本節では, 解析時点がデータに依存して決められる場合に, 試験全体での有意水準を α に維持するための棄却限界値の算出方法について α -spending function 法の枠組みで提案する. また, 期待被験者数を指標とした解析時点の最

$m > 1$ とする. U_1, \dots, U_t を互いに独立に $[0, \epsilon]$, ($0 < \epsilon < 1$) 上の一様分布に従う確率変数とする. このとき, $Y_i = X_i + U_i$ と定義すると $1 - \beta_1(\pi)$ は次のように表現できる.

$$1 - \beta_1(\pi) = \Pr \{ \max\{Y_i : i \in A\} > \max\{\lambda, \max\{Y_i : i \in A^c\}\} \}$$

これは, X_i の最大値にタイがあったときにランダムに一つの治療群を選ぶ工夫の一つである. A に含まれる治療群が一つ A^c に移ったとすると $1 - \beta_1(\pi)$ は明らかに減少する. 従って, 帰納的に P_{CS} は $m = 1$ のときに最小となる.

なお, $1 - \beta_2(\pi)$ が $\pi_\nu = \hat{\pi}_c + \delta_2$ で最小となることは自明である.

3 データに依存した中間解析計画への対応と解析時点の最適性

α -spending function 法を用いて中間解析を行う場合, どのような時点でもそれまでに得られている情報量から試験全体での有意水準を維持した棄却限界値を求めることができる. この性質は, データの蓄積率が時間に対して一定ではない臨床試験のデータ解析では非常に有用である. しかし, 解析時点や回数について任意であるとはいえ, 得られたデータに依存してその後の解析時点が決められる場合には試験全体での有意水準を維持できない可能性も指摘されており, 解析時点の定め方が完全に任意であるというわけではない. また, 実際にどの時点で中間解析を行うべきかという最適性の議論は, 最適性の指標と用いる α -spending function に依存する. 本節では, 解析時点がデータに依存して決められる場合に, 試験全体での有意水準を α に維持するための棄却限界値の算出方法について α -spending function 法の枠組みで提案する. また, 期待被験者数を指標とした解析時点の最

適性について、代表的な α -spending function を用いて論じる。

3.1 中間解析結果に依存した解析計画に対する棄却限界値

3.1.1 中間解析結果と解析時点の独立性

Lan and DeMets³ や Proschan ら⁴ が指摘しているように α -spending function 法を用いて棄却限界値を算出しても解析時点の選択方法により第一種の過誤確率が増大する可能性がある。これは、解析時点を事前に固定する必要のない α -spending function 法の場合には解析時点の列 $\mathbf{T} = (t_1, t_2, \dots)$ も一つの確率変数と考えられるが、 α -spending function 法では第一種の過誤確率の評価に解析時点を確率変数として考慮していないためである。ここでは、 α -spending function 法が試験全体での有意水準が α に維持される必要条件について述べるとともに、その必要条件が満たされない状況における棄却限界値の算出方法を提案する。

定理 3.1 中間解析時点を表すベクトル $\mathbf{T} = (t_1, t_2, \dots)$ と検定統計量の列 $B(\mathbf{T}) = (B(t_1), B(t_2), \dots)$ が独立ならば、 α -spending function 法により定められた棄却限界値は有意水準を $\alpha = \alpha^*(1)$ に保つ。

証明

α -spending function $\alpha^*(t)$ を固定し、 $\mathbf{T} = (t_1, t_2, \dots)$ に対し、 $\{b(t_i), i = 1, 2, \dots\}$ を以下の条件を満たすように定めるとする。

$$\Pr \{B(t_j) \leq b(t_j), j = 1, \dots, i-1, B(t_i) > b(t_i)\} = \alpha^*(t_i) - \alpha^*(t_{i-1})$$

ただし、 $t_0 = 0$ とする。

\mathbf{T} の標本空間および分布関数をそれぞれ \mathcal{T} , $F(\mathbf{T})$ とすると, $B(\mathbf{T})$, \mathbf{T} の独立性から, 棄却限界値 $\{b(t_i), i = 1, 2, \dots\}$ を用いた場合の第一種の過誤確率は次のように表すことができる.

$$\begin{aligned}
 & \int_{\mathcal{T}} \Pr \{B(t_i) > b(t_i) \text{ for some } i \mid \mathbf{T} = (t_1, t_2, \dots)\} dF(\mathbf{T}) \\
 &= \int_{\mathcal{T}} \Pr \{B(t_i) > b(t_i) \text{ for some } i\} dF(\mathbf{T}) \\
 &= \alpha \int_{\mathcal{T}} dF(\mathbf{T}) \\
 &= \alpha
 \end{aligned}$$

従って, 検定統計量の列の分布とは独立に解析時点が決められるならば α -spending function 法により第一種の過誤は α に保たれる.

一方, Lan and DeMets³ や Proschan ら⁴ の例のように $B(\mathbf{T})$ に依存して \mathbf{T} が定められる場合には第一種の過誤確率は解析時点ごとの条件付き確率を積分して得られる.

例えば, Lan and DeMets(2.4.1 節) では, \mathbf{T} は次の 4 つの値をとる.

$$T_1 = (0.25, 0.50, 0.75, 1)$$

$$T_2 = (0.25, 0.375, 0.50, 0.625, 0.75, 0.875, 1)$$

$$T_3 = (0.25, 0.50, 0.625, 0.75, , 0.875, 1)$$

$$T_4 = (0.25, 0.50, 0.75, 0.825, 1)$$

従って、 $Z(t_i)$, $c(t_i)$ ($i = 1, 2, 3, 4$) をそれぞれ各解析時点での検定統計量および棄却限界値とすると第一種の過誤確率 α' は次のように表すことができる。

$$\alpha' = \sum_{k=1}^4 \Pr\{Z(t_i) > c(t_i) \text{ for some } t_i \in T_k \mid T = T_k\} \Pr\{T = T_k\}$$

つまり、データに依存して解析時点を変える場合には第一種の過誤確率は全確率の法則からすべての解析時点における第一種の過誤確率の和を計算する必要がある。従って、検定統計量の値と解析時点が独立でない場合には α -spending function 法で棄却限界値を定めても第一種の過誤確率は事前に定めた有意水準 α に一致するとは限らない。

3.2 データに依存した解析計画に対する棄却限界値の算出方法

本節では、中間解析結果に依存して1回だけ解析計画を変更する場合の棄却限界値の算出方法について検討する。中間解析結果に依存して解析時点が定められる場合には解析時点も確率変数として考え、選択される可能性のある解析時点での棄却限界値を α -spending function によりどのように与えるかが問題となる。ここでは、解析時点が選択される条件付き確率を用いて棄却限界値を算出する方法について示す。

解析時点 t での検定統計量を $X(t)$ 、棄却限界値を $b(t)$ で表す。 $X(t)$ は帰無仮説の下で標準ブラウン運動 $B(t)$ に従うとし、 t は information fraction とする。また、1回目 ($t = t_1$) の中間解析の結果で有意にならなかったが、検定統計量の値があらかじめ設定した範囲 S_1 に入れば試験終了までにもう一度中間解析を行い、そうでない場合は試験終了時点まで解析を行わないという解析計画を考える。例えば、Lan and DeMets(2.4.1 節) が検討した解析計画のように検定統計量の値が棄却限

界値の8割の値を越えた場合に解析時点を増やすとすると,

$$S_1 = \{x \mid 0.8 \cdot b(t_1) \leq x \leq b(t_1)\}$$

である. 一般に,

$$S_1 \subset \{x \mid x \leq b(t_1)\}$$

$$S_2 = \{x \mid x \leq b(t_1)\} \setminus S_1$$

$$\mathbf{T} = \begin{cases} (t_1, t_2, 1) & \text{if } B(t_1) \in S_1 \\ (t_1, 1) & \text{if } B(t_1) \in S_2 \end{cases}$$

とし, S_1 は固定された集合として考える. このとき, 次が成り立つ.

定理 3.2 $\alpha^*(t)$ を $\alpha^*(0) = 0$, $\alpha^*(1) = \alpha$ なる α -spending function とし, information fraction t に対し, 帰無仮説の下で検定統計量が $B(t)$ で表されるとする. このとき, $\alpha^*(t)$ を用い, $\mathbf{T} = (t_1, t_2, 1)$, $\mathbf{T} = (t_1, 1)$ のときの棄却限界値,

$$(b(t_1), \tilde{b}(t_2), \tilde{b}(1)), (b(t_1), b(1)))$$

を次のように定めれば有意水準は α に保たれる.

$$\begin{aligned} \alpha^*(t_1) &= \Pr\{B(t_1) > b(t_1)\} \\ \frac{\alpha^*(t_2) - \alpha^*(t_1)}{1 - \alpha^*(t_1)} &= \Pr\{B(t_2) > \tilde{b}(t_2) \mid B(t_1) \in S_1\} \\ \frac{\alpha^*(1) - \alpha^*(t_2)}{1 - \alpha^*(t_1)} &= \Pr\{B(t_2) \leq \tilde{b}(t_2), B(1) > \tilde{b}(1) \mid B(t_1) \in S_1\} \\ \frac{\alpha^*(1) - \alpha^*(t_1)}{1 - \alpha^*(t_1)} &= \Pr\{B(1) > b(1) \mid B(t_1) \in S_2\} \end{aligned}$$

証明

この解析計画に対する第一種の過誤確率 α' は次のように表される。

$$\begin{aligned}\alpha' &= \Pr\{B(t_1) > b(t_1)\} + \Pr\{B(t_1) \in S_1, B(t_2) > \tilde{b}(t_2)\} \\ &\quad + \Pr\{B(t_1) \in S_1, B(t_2) \leq \tilde{b}(t_2), B(1) > \tilde{b}(1)\} \\ &\quad + \Pr\{B(t_1) \in S_2, B(1) > b(1)\}\end{aligned}$$

今, $\alpha^*(t)$ を用いて次の式が成り立つように $c(t_1), \tilde{c}(t_2), \tilde{c}(1), c(1)$ を定めるとする。

$$\begin{aligned}\alpha^*(t_1) &= \Pr\{B(t_1) > c(t_1)\} \\ \alpha^*(1) - \alpha^*(t_1) &= \Pr\{B(t_1) \leq c(t_1), B(1) > c(1)\} \\ \alpha^*(t_2) - \alpha^*(t_1) &= \Pr\{B(t_1) \leq c(t_1), B(t_2) > \tilde{c}(t_2)\} \\ \alpha^*(1) - \alpha^*(t_2) &= \Pr\{B(t_1) \leq c(t_1), B(t_2) \leq \tilde{c}(t_2), B(1) > \tilde{c}(1)\}\end{aligned}$$

ここで,

$$E_{x,t} = \{B(t) > x(t)\}, \quad E_{x,t}^c = \{B(t) \leq x(t)\}, \quad E_t(S) = \{B(t) \in S\}$$

$$P_S = \Pr\{B(t_1) \in S \mid B(t_1) \leq b(t_1)\}$$

と表すことにする。定義から,

$$\begin{aligned}\alpha' &= \Pr\{E_{b,t_1}\} + \Pr\{E_{t_1}(S_1) \cap E_{\tilde{b},t_2}\} \\ &\quad + \Pr\{E_{t_1}(S_1) \cap E_{\tilde{b},t_2}^c \cap E_{\tilde{b},1}\} + \Pr\{E_{t_1}(S_2) \cap E_{b,1}\} \\ \alpha &= \left(\Pr\{E_{c,t_1}\} + \Pr\{E_{c,t_1}^c \cap E_{\tilde{c},t_2}\} + \Pr\{E_{c,t_1}^c \cap E_{\tilde{c},t_2}^c \cap E_{\tilde{c},1}\}\right) P_{S_1} \\ &\quad + \left(\Pr\{E_{c,t_1}\} + \Pr\{E_{c,t_1}^c \cap E_{c,1}\}\right) P_{S_2}\end{aligned}$$

である。従って,

$$\alpha' - \alpha = \Pr\{E_{b,t_1}\} - \Pr\{E_{c,t_1}\}$$

$$\begin{aligned}
& + \Pr \{E_{t_1}(S_1) \cap E_{\bar{b},t_2}\} - \Pr \{E_{c,t_1}^c \cap E_{\bar{c},t_2}\} P_{S_1} \\
& + \Pr \{E_{t_1}(S_1) \cap E_{\bar{b},t_2}^c \cap E_{\bar{b},1}\} - \Pr \{E_{c,t_1}^c \cap E_{\bar{c},t_2}^c \cap E_{\bar{c},1}\} P_{S_1} \\
& + \Pr \{E_{t_1}(S_2) \cap E_{b,1}\} - \Pr \{E_{c,t_1}^c \cap E_{c,1}\} P_{S_2} \\
= & \Pr \{E_{b,t_1}\} - \Pr \{E_{c,t_1}\} \\
& + \left(\frac{\Pr \{E_{t_1}(S_1) \cap E_{\bar{b},t_2}\}}{P_{S_1}} - \Pr \{E_{c,t_1}^c \cap E_{\bar{c},t_2}\} \right) P_{S_1} \\
& + \left(\frac{\Pr \{E_{t_1}(S_1) \cap E_{\bar{b},t_2}^c \cap E_{\bar{b},1}\}}{P_{S_1}} - \Pr \{E_{c,t_1}^c \cap E_{\bar{c},t_2}^c \cap E_{\bar{c},1}\} \right) P_{S_1} \\
& + \left(\frac{\Pr \{E_{t_1}(S_2) \cap E_{b,1}\}}{P_{S_2}} - \Pr \{E_{c,t_1}^c \cap E_{c,1}\} \right) P_{S_2}
\end{aligned}$$

よって,

$$\begin{aligned}
\Pr \{E_{b,t_1}\} & = \Pr \{E_{c,t_1}\} \\
\frac{\Pr \{E_{t_1}(S_1) \cap E_{\bar{b},t_2}\}}{P_{S_1}} & = \Pr \{E_{c,t_1}^c \cap E_{\bar{c},t_2}\} \\
\frac{\Pr \{E_{t_1}(S_1) \cap E_{\bar{b},t_2}^c \cap E_{\bar{b},1}\}}{P_{S_1}} & = \Pr \{E_{c,t_1}^c \cap E_{\bar{c},t_2}^c \cap E_{\bar{c},1}\} \\
\frac{\Pr \{E_{t_1}(S_2) \cap E_{b,1}\}}{P_{S_2}} & = \Pr \{E_{c,t_1}^c \cap E_{c,1}\}
\end{aligned}$$

が成り立てば,

$$\alpha' = \alpha$$

となることから, 定理の式が導かれる.

より一般に, 次が成り立つ.

定理 3.3 $\alpha^*(t)$ を $\alpha^*(0) = 0$, $\alpha^*(1) = \alpha$ なる α -spending function とする. information fraction t に対し, 帰無仮説の下で検定統計量が $B(t)$ で表されるとする.

総解析回数を K 回とし、各解析時に用いる棄却限界値の列を $b(t_i)$, ($i = 1, \dots, K$) とする。 $K - 1$ 回の中間解析時点の中から一時点 q ($1 \leq q \leq K - 1$) および次の条件を満たす S_1 を固定する。

$$S_1 \subset \{x \mid x \leq b(t_q)\}$$

$$S_2 = \{x \mid x \leq b(t_q)\} \setminus S_1$$

この S_1, S_2 に対し、次のような解析計画を用いるとする。

$$\mathbf{T} = \begin{cases} (t_1, \dots, t_q, t_{q'}, t_{q+1}, \dots, t_K = 1) & \text{if } B(t_q) \in S_1 \\ (t_1, \dots, t_q, t_{q+1}, \dots, t_K = 1) & \text{if } B(t_q) \in S_2 \end{cases}$$

このとき、 $\alpha^*(t)$ を用いて棄却限界値、

$$\begin{cases} (b(t_1), \dots, b(t_q), \tilde{b}(t_{q'}), \tilde{b}(t_{q+1}), \dots, \tilde{b}(1)) & \text{if } B(t_q) \in S_1 \\ (b(t_1), \dots, b(t_q), b(t_{q+1}), \dots, b(1)) & \text{if } B(t_q) \in S_2 \end{cases}$$

を次のように定めれば試験全体での有意水準を α に保つことができる。

- $i = 1, \dots, q$ のとき

$$\Pr\{B(t_j) \leq b(t_j), j = 1, \dots, i - 1, B(t_i) > b(t_i)\} = \alpha^*(t_i) - \alpha^*(t_{i-1})$$

- $B(t_q) \in S_1, i = q'$ のとき

$$\Pr\{B(t_j) \leq b(t_j), j = 1, \dots, q, B(t_{q'}) > \tilde{b}(t_{q'}) \mid B(t_q) \in S_1\} = \frac{\alpha^*(t_{q'}) - \alpha^*(t_q)}{1 - \alpha^*(t_q)}$$

- $B(t_q) \in S_1, i = q + 1, \dots, K$ のとき

$$\Pr\{B(t_j) \leq b(t_j), j = 1, \dots, q, B(t_m) \leq \tilde{b}(t_m),$$

$$m = q', q + 1, \dots, i - 1, B(t_i) > \tilde{b}(t_i) \mid B(t_q) \in S_1\} = \frac{\alpha^*(t_i) - \alpha^*(t_{i-1})}{1 - \alpha^*(t_q)}$$

- $B(t_q) \in S_2, i = q + 1, \dots, K$ のとき

$$\Pr\{B(t_j) \leq b(t_j), j = 1, \dots, i-1, B(t_i) > b(t_i) \mid B(t_q) \in S_2\} = \frac{\alpha^*(t_i) - \alpha^*(t_{i-1})}{1 - \alpha^*(t_q)}$$

と定めれば有意水準は α に保たれる。ただし, $t_0 = 0$ とする。

証明

$$\begin{aligned} \alpha' &= \sum_{i=1}^q \Pr\{B(t_j) \leq b(t_j), j = 1, \dots, i-1, B(t_i) > b(t_i)\} \\ &\quad + \Pr\{B(t_j) \leq b(t_j), j = 1, \dots, i-1, B(t_q) \in S_1, B(t_{q'}) > \bar{b}(t_{q'})\} \\ &\quad + \sum_{i=q+1}^K \Pr\{B(t_j) \leq b(t_j), j = 1, \dots, q-1, B(t_q) \in S_1, \\ &\quad\quad\quad B(t_m) \leq \bar{b}(t_m), m = q', q+1, \dots, i-1, B(t_i) > \bar{b}(t_i)\} \\ &\quad + \sum_{i=q+1}^K \Pr\{B(t_j) \leq b(t_j), j = 1, \dots, i-1, j \neq q, B(t_q) \in S_2, B(t_i) > b(t_i)\} \end{aligned}$$

α -spending function $\alpha^*(t)$ を用いて,

$$\{c(t_i), i = 1, \dots, K\}, \{\bar{c}(t_i), i = q', q+1, \dots, K\}$$

を次のように定める。

- $i = 1, \dots, q$ のとき,

$$\alpha^*(t_i) - \alpha^*(t_{i-1}) = \Pr\{B(t_j) \leq c(t_j), j = 1, \dots, i-1, B(t_i) > c(t_i)\}$$

- $i = q'$ のとき

$$\alpha^*(t_{q'}) - \alpha^*(t_q) = \Pr\{B(t_j) \leq c(t_j), j = 1, \dots, q, B(t_{q'}) > \bar{c}(t_{q'})\}$$

• $i = q + 1, \dots, K$ のとき,

$$\alpha^*(t_i) - \alpha^*(t_{i-1}) = \Pr \{B(t_j) \leq c(t_j), j = 1, \dots, q, B(t_{q'}) \leq \tilde{c}(t_{q'}), \\ B(t_m) \leq \tilde{c}(t_m), m = q + 1, \dots, i - 1, B(t_i) > \tilde{c}(t_i)\}$$

$$\alpha^*(t_i) - \alpha^*(t_{i-1}) = \Pr \{B(t_j) \leq c(t_j), j = 1, \dots, i - 1, B(t_i) > c(t_i)\}$$

ここで,

$$E_{x,t} = \{B(t) > x(t)\}, \quad E_{x,t}^c = \{B(t) \leq x(t)\}, \quad E_t(S) = \{B(t) \in S\}$$

$$P_{S,i} = \Pr \{B(t_i) \in S \mid B(t_i) \leq b(t_i)\}$$

と表すと,

$$\begin{aligned} \alpha' - \alpha &= \sum_{i=1}^q \Pr \left\{ \left(\bigcap_{j=1}^{i-1} E_{b,t_j}^c \right) \cap E_{b,t_i} \right\} - \sum_{i=1}^q \Pr \left\{ \left(\bigcap_{j=1}^{i-1} E_{c,t_j}^c \right) \cap E_{c,t_i} \right\} \\ &\quad + \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{b,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\tilde{b},t_{q'}} \right\} \\ &\quad - \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{c,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\tilde{c},t_{q'}} \right\} P_{S_1,q} \\ &\quad + \sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{b,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\tilde{b},t_{q'}}^c \cap \left(\bigcap_{m=q+1}^{i-1} E_{\tilde{b},t_m}^c \right) \cap E_{\tilde{b},t_i} \right\} \\ &\quad - \sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{c,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\tilde{c},t_{q'}}^c \cap \left(\bigcap_{m=q+1}^{i-1} E_{\tilde{c},t_m}^c \right) \cap E_{\tilde{c},t_i} \right\} P_{S_1,q} \\ &\quad + \sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{b,t_j}^c \right) \cap E_{t_q}(S_2) \cap \left(\bigcap_{m=q+1}^{i-1} E_{b,t_m}^c \right) \cap E_{b,t_i} \right\} \\ &\quad - \sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{c,t_j}^c \right) \cap E_{t_q}(S_2) \cap \left(\bigcap_{m=q+1}^{i-1} E_{c,t_m}^c \right) \cap E_{c,t_i} \right\} P_{S_2,q} \\ &= \left(\sum_{i=1}^q \Pr \left\{ \left(\bigcap_{j=1}^{i-1} E_{b,t_j}^c \right) \cap E_{b,t_i} \right\} - \sum_{i=1}^q \Pr \left\{ \left(\bigcap_{j=1}^{i-1} E_{c,t_j}^c \right) \cap E_{c,t_i} \right\} \right) \\ &\quad + \left(\Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{b,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\tilde{b},t_{q'}} \right\} / P_{S_1,q} \right) \end{aligned}$$

$$\begin{aligned}
& - \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{c,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\bar{c},t_{q'}} \right\} P_{S_1,q} \\
& + \left(\sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{b,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{b,t_{q'}}^c \cap \left(\bigcap_{m=q+1}^{i-1} E_{b,t_m}^c \right) \cap E_{b,t_i} \right\} / P_{S_1,q} \right. \\
& \quad \left. - \sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{c,t_j}^c \right) \cap E_{t_q}(S_1) \cap E_{\bar{c},t_{q'}}^c \cap \left(\bigcap_{m=q+1}^{i-1} E_{\bar{c},t_m}^c \right) \cap E_{\bar{c},t_i} \right\} \right) P_{S_1,q} \\
& + \left(\sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{b,t_j}^c \right) \cap E_{t_q}(S_2) \cap \left(\bigcap_{m=q+1}^{i-1} E_{b,t_m}^c \right) \cap E_{b,t_i} \right\} / P_{S_2,q} \right. \\
& \quad \left. - \sum_{i=q+1}^K \Pr \left\{ \left(\bigcap_{j=1}^{q-1} E_{c,t_j}^c \right) \cap E_{t_q}(S_2) \cap \left(\bigcap_{m=q+1}^{i-1} E_{c,t_m}^c \right) \cap E_{c,t_i} \right\} \right) P_{S_2,q}
\end{aligned}$$

となる。従って、定理の条件式が満たされれば、 $\alpha' = \alpha$ となる。

これらの結果から、中間解析結果に依存して1回だけ解析回数を増やす場合にも α -spending function を用いて試験全体での有意水準を α に維持する棄却限界値を定めることができることがわかる。

3.2.1 棄却限界値の算出例

定理 3.1.2 で示した方法で実際に棄却限界値を算出した例を示す。表 18,19,20 (p.107 ~ 109) は、

$$S_1 = \{x \mid 0.8 \cdot b(t_1) \leq x \leq b(t_1)\}$$

とした場合、つまり、検定統計量の値が棄却限界値の8割の値を越えたときに解析回数を一回増やすという解析計画に対する棄却限界値である。例えば、O'Brien-Fleming 型の α -spending function を用いて、

$$T = \begin{cases} (0.3, 0.5, 1) & \text{if } B(t_i) \in S_1 \\ (0.3, 1) & \text{if } B(t_i) \in S_2 \end{cases}$$

という解析計画を用いた場合の棄却限界値 $(b(.3), \bar{b}(.5), \bar{b}(1))$, $(b(.3), b(1))$ は,

$$(3.929, 4.554, 3.516), (3.929, 1.954)$$

である。この結果からわかるように、解析回数を増やした場合にはかなり保守的な棄却限界値となる。一方、解析回数を増やさなかった場合の最終解析時点での棄却限界値は、増やした場合よりもかなり小さい値である。他の α -spending function を用いた場合も同様の傾向があり、解析回数を増やすとかなり保守的な棄却限界値を用いなければならないことがわかる。

3.3 解析時点と期待被験者数

2.2.4 節で述べたように McPherson²⁴ は解析回数と期待被験者数について検討しているが、解析間隔に関しては等間隔であることを前提に議論したものである。また、解析回数が2回もしくは3回の場合は Brittain ら²⁵ の方法で棄却限界値を求めることにより期待被験者数を最小にできるが、あらかじめ解析時点は固定する必要がある。情報量の増加が時間に対して一定でない臨床試験の場合、 α -spending function 法のような解析回数や解析時点に依存しない方法を採用する方がより柔軟で実際的である。しかし、 α -spending function 法は有意水準の維持という点ではこの柔軟性を有する方法であるが、検出力や期待被験者数に関しては α -spending function の形状や解析回数、解析時点に依存して変化する。このため、本節では α -spending function 法を用いた場合の解析時点の最適性について期待被験者数の観点から論じる。ただし、期待被験者数は試験が最後まで継続された場合の総被験者数や用いる α -spending function にも依存するため、ここでは以下の条件の

下で検討することとする。

- 最大被験者数を固定した場合
- 検出力を一定とした場合

ここで最大被験者数とは、中間解析により試験が途中で終了されることなく最後まで継続された場合の総被験者数である。

また、 α -spending function には様々なもの考えられるが、ここでは 2.2.5 節で示した代表的な 3 つの α -spending function, O'Brien-Fleming 型 $\alpha_1^*(t)$, Pocock 型 $\alpha_2^*(t)$, 直線型 $\alpha_3^*(t)$ を用いて検討する。

- O'Brien-Fleming 型

$$\alpha_1^*(t) = 2 \left[1 - \Phi \left(z_{\alpha/2} / \sqrt{t} \right) \right]$$

- Pocock 型

$$\alpha_2^*(t) = \alpha \log \{ 1 + (e - 1)t \}$$

- 直線型

$$\alpha_3^*(t) = \alpha t$$

3.3.1 最大被験者数を固定した場合

一般に、中間解析時点ごとに仮説検定を繰り返すために生じる多重性を調整すると中間解析を行わなかった場合 (以下、固定デザインと呼ぶ) に比べ検出力は低

下する。このため、最大被験者数を固定した場合には期待被験者数と同時に検定の多重性の調整による検出力の低下も検討する必要がある。

本節では α -spending function 法を用いた場合の期待被験者数と検出力を最大被験者数を固定した条件の下で算出し、解析時点の最適性について検討する。

t を information fraction とし、時点 t での検定統計量 $X(t)$ が次のような標準ブラウン運動 $B(t)$ とドリフトパラメータ ξ の和で表されるとする。

$$X(t) = B(t) + \xi t \quad (0 \leq t \leq 1)$$

また、帰無仮説 H_0 、対立仮説 H_1 を次のように仮定する。

$$H_0 : \xi = 0$$

$$H_1 : \xi > 0$$

このとき、有意水準片側 α 、検出力 $1 - \beta$ となる ξ の値を ξ_β で表すことにする。もし中間解析を行わないのであれば、この ξ_β は、

$$\xi_\beta = z_\alpha + z_\beta$$

である⁸。ただし、 z_γ は標準正規分布の上側 $\gamma\%$ 点とする。従って、対立仮説を $\xi = \xi_\beta$ とし、固定デザインで $\xi = \xi_\beta$ となる被験者数を N_f とすれば、 α -spending function 法で中間解析を行った場合の検出力および試験の期待停止時間を算出することができる。これらの値から固定デザインでの値に対する相対的な検出力及び期待被験者数の低下量がわかる。

各解析時点 t_i での棄却限界値を $b(t_i)$ とし、 $\xi = \xi_\beta$ のときの第二種の過誤確率を

β^* とすると、検出力 $1 - \beta^*$ は次のように表される。

$$\begin{aligned} 1 - \beta^* &= \Pr\{X(t_i) > b(t_i) \text{ for some } i \mid \xi = \xi_\beta\} \\ &= \Pr\{B(t_i) + (z_\alpha + z_\beta)t_i > b(t_i) \text{ for some } i\} \end{aligned}$$

よって、 α , β と解析回数 K および解析時点 $T = (t_1, t_2, \dots, t_K)$ を与えることにより $1 - \beta^*$ は容易に計算できる。また、対立仮説の下で試験の期待停止時間 E_t も次のように求めることができる。

$$\begin{aligned} E_t &= \sum_{i=1}^K t_i \Pr\{X(t_j) \leq b(t_j), j = 1, \dots, i-1, X(t_i) > b(t_i) \mid \xi = \xi_\beta\} \\ &\quad + 1 \cdot \Pr\{X(t_j) \leq b(t_j), j = 1, \dots, K \mid \xi = \xi_\beta\} \end{aligned}$$

中間解析を行った場合の期待被験者数 $E(N)$ は、

$$E(N) = N_f \cdot E_t$$

で表されるが、ここでは、 N_f を固定して考えているため、 E_t が最小のとき期待被験者数 $E(N)$ も最小となる。

$\alpha = 0.025$, $1 - \beta = 0.80$, $K = 2, 3, 4$ のときの O'Brien-Fleming 型 $\alpha_1^*(t)$, Pocock 型 $\alpha_2^*(t)$, 直線型 $\alpha_3^*(t)$ を用いて棄却限界値を算出した場合の検出力と期待停止時間 E_t を表 21~24 (p.110 ~ 113) に示す。

表 21~24 の結果からいずれの $\alpha^*(t)$ に対しても解析回数が増えるにつれて検出力が低下し、期待被験者数も小さくなることがわかる。しかしながら、その性質は用いる α -spending function により異なる。O'Brien-Fleming 型の $\alpha_1^*(t)$ を用いた場合には検出力への影響が小さいが、Pocock 型の $\alpha_2^*(t)$ を用いた場合にはかな

り検出力が低下することがわかる。期待被験者数に関しては Pocock 型が最も小さくなるが、

$$(\text{期待被験者数の減少量})/(\text{検出力の低下量})$$

という比を考えると O'Brien-Fleming 型が最も大きい。

検討した範囲内で最も期待被験者数が小さくなる解析時点を表 16 に示した。

表 16: $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ を用いたときの期待被験者数の最小値と解析時点

$\alpha = 0.025, 1 - \beta = 0.80$			
		T	(β^*, E_t)
$K = 2$	$\alpha_1^*(t)$	(0.70,1)	(0.794, 0.861)
	$\alpha_2^*(t)$	(0.55,1)	(0.750, 0.784)
	$\alpha_3^*(t)$	(0.55,1)	(0.766, 0.798)
$K = 3$	$\alpha_1^*(t)$	(0.60,0.80,1)	(0.790, 0.819)
	$\alpha_2^*(t)$	(0.40,0.70,1)	(0.732, 0.736)
	$\alpha_3^*(t)$	(0.40,0.70,1)	(0.752, 0.748)
$K = 4$	$\alpha_1^*(t)$	(0.5,0.6,0.8,1)	(0.790, 0.804)
		(0.5,0.7,0.8,1)	(0.790, 0.804)
	$\alpha_2^*(t)$	(0.3,0.6,0.8,1)	(0.721, 0.718)
	$\alpha_3^*(t)$	(0.4,0.6,0.8,1)	(0.745, 0.727)

$\alpha_1^*(t)$ は試験の前半に中間解析を実施しても α をほとんど消費しないため、試験の後半に解析を行った方が期待被験者数を小さくできる。一方、 $\alpha_2^*(t)$ や $\alpha_3^*(t)$ は早い段階から α を消費できるため、解析回数に応じて等間隔に解析を行うことにより期待被験者数を小さくすることができる。

これらの結果は、被験者数に制約があり中間解析に伴う検出力の低下を補うことができない場合に利用可能である。この際、検出力の低下をどの程度まで許容できるかが問題となる。検出力の低下よりも期待被験者数を優先させる場合には、 $\alpha_2^*(t)$ のような早期から多くの α を消費するタイプの α -spending function を用

いるべきである。しかし、検出力の低下をなるべく小さくしたいという場合には $\alpha_1^*(t)$ を用いて検出力の低下を小さくし、解析回数を増やすことで期待被験者数を小さくするという方法も考えられる。例えば、期待被験者数を 20% 低下させたいとした場合、3 ~ 4% の検出力の低下を許容できるのであれば、 $\alpha_3^*(t)$ を用いて $0.50 \leq t \leq 0.60$ の範囲で中間解析を 1 回行えばよいが、もし、3 ~ 4% の検出力の低下が許容できないと判断されたならば、 $\alpha_1^*(t)$ を用いて $T = (0.5, 0.6, 0.8, 1)$ において 4 回の解析を行えば、期待被験者数をほぼ 20% 低下させることができ、検出力の低下も約 1% に抑えることができる。ただし、解析回数が多いほど期待被験者数が小さくなるというわけでもない。 $\alpha_1^*(t)$ を用いて

$$T = (0.50, 0.70, 1)$$

の 3 回の解析を行う場合と、

$$T = (0.20, 0.70, 0.80, 1)$$

の 4 回の解析を行う場合の検出力と試験の期待停止時間 E_t を比較するとそれぞれ $(0.794, 0.830)$, $(0.790, 0.835)$ であり、3 回の解析の方が検出力の低下も小さく期待被験者数も小さい。従って、検出力と期待被験者数という観点からは解析回数を 4 回とした利点はないことになる。

次節では、あらかじめ検出力を固定し、解析回数と解析時点を与えたときにその検出力を保つために必要な最大被験者数と期待被験者数を求め、どの時点で解析を行えば最も期待被験者数を小さくできるかについて検討する。

3.3.2 検出力一定のときの最大被験者数と期待被験者数

前節で述べたように中間解析を行うことで生じる検定の多重性の調整により検出力は低下する。従って、検出力が同じであれば固定デザインに比べ中間解析を行う試験に必要な最大被験者数は大きくなる。本節では、 $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ に対し、検出力を固定した場合の最大被験者数と期待被験者数を数値計算により求め、期待被験者数が最小となる時点について検討する。ただし、期待被験者数の最小値が検討の対象であることから、同じ検出力をもつ固定デザインの被験者数との相対値として議論する。

前節と同様の仮定の下、 α -spending function 法により中間解析を行うとする。また、解析時点を $\mathbf{T} = (t_1, \dots, t_K)$ とし、有意水準片側 α 、検出力 $1 - \beta$ となる ξ の値を $\tilde{\xi}_\beta = \xi_\beta(\alpha^*, \mathbf{T})$ とする。従って、各解析時点での棄却限界値の列を $\{b(t_i), i = 1, \dots, K\}$ とすると、

$$\alpha = \Pr\{B(t_i) > b(t_i) \text{ for some } i\} \quad (3.2.1)$$

$$1 - \beta = \Pr\{B(t_i) + \tilde{\xi}_\beta t_i > b(t_i) \text{ for some } i\} \quad (3.2.2)$$

を満たす。前節で述べたように、中間解析を行わなかったとした場合の $\tilde{\xi}_\beta$ は、

$$\tilde{\xi}_\beta = \xi_\beta = z_\alpha + z_\beta$$

であるから、固定デザインに必要な被験者数を 1 とした場合に、中間解析を行って検出力を $1 - \beta$ に保つために必要となる相対的な最大被験者数は

$$IF = IF(\alpha^*(t), \mathbf{T}) = \left(\frac{\tilde{\xi}_\beta}{z_\alpha + z_\beta} \right)^2$$

となる³¹。この IF を ‘inflation factor’ という^{32, 33}。

試験の停止時間を表す確率変数を T^* とすると、中間解析を行わなかった場合を 1 としたときの相対的な期待被験者数 E_{IF} は

$$\begin{aligned} E_{IF} &= IF(\alpha^*(t), T) \cdot E(T^* | H_1) \\ &= IF(\alpha^*(t), T) \left(\sum_{i=1}^K t_i \Pr \{ X(t_j) \leq b(t_j), j = 1, \dots, i-1, X(t_i) > b(t_i) \} \right. \\ &\quad \left. + \Pr \{ X(t_i) < b(t_i), i = 1, \dots, K \} \right) \end{aligned}$$

で表される。解析回数、解析時点を固定することにより、(3.2.1) および (3.2.2) から $\tilde{\xi}_\beta$ 値を求めることができる。 $\tilde{\xi}_\beta$ が与えられることにより、 $IF(\alpha^*(t), T)$ および $E(T^* | H_1)$ を求めることができるため、数値計算によりこの相対的な期待被験者数が $\alpha^*(t)$ や解析回数、解析時点でどのように変化するかを検討することが可能である。

3.3.3 数値計算による検討結果

O’Brien-Fleming 型 $\alpha_1^*(t)$, Pocock 型 $\alpha_2^*(t)$, 直線型 $\alpha_3^*(t)$, の α -spending function を用い、有意水準片側 2.5%, 検出力 80% という条件で解析回数を 2~4 回としたときの IF と E_{IF} を数値計算により検討した。解析時点は $[0, 1]$ 上の連続変数であるが、数値計算上、ここでは解析回数が $K = 2, 3$ のときは 0.05 間隔、 $K = 4$ のときは 0.10 間隔で離散的にとり、 IF と E_{IF} を求めることとした。表 17 に検討した範囲内で期待被験者数が最も小さくなる解析時点と IF , E_{IF} 値を示す。この表から、O’Brien-Fleming 型 $\alpha_1^*(t)$ の場合、期待被験者数を小さくするためには、試験の前半は解析を行わずに被験者数が最大被験者数の 1/2 以上集積された時点から

表 17: E_{IF} が最小となる解析時点

		T	(E_{IF}, IF)
$K = 2$	$\alpha_1^*(t)$	(0.70,1)	(0.872, 1.015)
	$\alpha_2^*(t)$	(0.50,1)	(0.855, 1.123)
	$\alpha_3^*(t)$	(0.55,1)	(0.849, 1.085)
$K = 3$	$\alpha_1^*(t)$	(0.60,0.75,1)	(0.835, 1.021)
	$\alpha_2^*(t)$	(0.40,0.70,1)	(0.818, 1.172)
	$\alpha_3^*(t)$	(0.40,0.70,1)	(0.808, 1.120)
$K = 4$	$\alpha_1^*(t)$	(0.5,0.6,0.8,1)	(0.820, 1.025)
	$\alpha_2^*(t)$	(0.3,0.5,0.7,1)	(0.801, 1.190)
	$\alpha_3^*(t)$	(0.3,0.5,0.7,1)	(0.792, 1.132)

解析回数に応じてほぼ等間隔で解析を行うとよいことがわかる。これは、前節でも述べたように $t \leq 1/2$ のときの $\alpha_1^*(t)$ の値は非常に小さいため、試験がその時点で終了する確率が極めて小さいく、初期の段階で解析を行っても最大被験者数や期待被験者数にはほとんど影響を与えないためである。一方、 Pocock 型 $\alpha_2^*(t)$ や直線型 $\alpha_3^*(t)$ の場合には試験の前半からある程度の α を消費できるため、試験全体で情報量の増加に対しほぼ等間隔で解析を行うことが望ましいことがわかる。

期待被験者数は、直線型、 Pocock 型、 O'Brien-Fleming 型の順で小さくなるが、最大被験者数は O'Brien-Fleming 型、直線型、 Pocock 型の順で小さいという結果であった。期待被験者数から見れば、直線型が最も良いという結果であるが、 O'Brien-Fleming 型との差は相対値で 0.02 ~ 0.03 であり、逆に最大被験者数は O'Brien-Fleming 型の方が直線型よりも 0.10 程度少ない。従って、 O'Brien-Fleming 型を用いて解析時点を適切に選ぶことにより最大被験者数をそれほど増大させずに直線型や Pocock 型に近い水準まで期待被験者数を低下させることが可能であるといえ

る。また、必ずしも E_{IF} を最小にする解析間隔は単調増加になっているわけではないことがわかる。 α -spending function の形状により解析間隔が単調増加でなくとも消費できる α の量は単調増加になることもあるため、 Li and Geller²³ が指摘するように解析間隔を単調増加にとる必要はない。

表 25~39 (p.114 ~ 128) に検討した範囲のすべての IF 値と E_{IF} 値について示す。これらの表から、いずれの α -spending function を用いても解析回数を増やすことにより検出力を 80% に保つために必要な最大被験者数は増加するが、逆に期待被験者数は低下させることができることがわかる。ただし、解析回数を増やしても解析時点の取り方によっては必ずしも期待被験者数を減らすことにはならない。 $\alpha_1^*(t)$ を用いて $T = (0.60, 0.75, 1)$ の 3 回の解析を行う場合と $T = (0.30, 0.40, 0.80, 1)$ の 4 回の解析を行う場合を比較するとそれぞれの (E_{IF}, IF) は $(0.835, 1.021)$, $(0.872, 1.022)$ であり、解析回数を 3 回とした方が期待被験者数、最大被験者数とも小さくなっていることがわかる。これは、検出力が同じであるという条件の下で比較しているため解析回数を増やせばそれだけ多重性の調整も厳しくなることに起因することである。

興味深い結果として、 $\alpha_2^*(t)$ と $\alpha_3^*(t)$ を比較した場合、 $\alpha_3^*(t)$ の方が E_{IF} , IF とも小さいということである。この場合、他のほとんどの時点でもこのような関係が成り立っているが、この関係が一般に成り立つというわけではない。ここでは示さなかったが、 $1 - \beta = 0.90$ とした場合には $\alpha_2^*(t)$ の E_{IF} 値の方が小さくなることが多い。試験の計画時にはこれらのことを考慮して α -spending function を設定することが重要である。

4 対照に対し許容的な治療法をすべて選択するための試験デザイン

Randomized phase II デザイン⁶ や Thall らのデザイン⁷ は最終的に有効率の最も大きな治療法を選択する方法であった。一方、スクリーニングの段階などでは必ずしも有効率が最も大きな治療法だけを選ぶのではなく、ある許容範囲以上の治療法はすべて選択したいという状況も存在する。一般的な方法論として、Gupta and Soble は indifference-zone アプローチを応用し、選ばれた群の中に許容的な群がすべて含まれる確率がある一定値以上になるような試験デザインについて検討している^{34, 35}。しかしながら、彼らの方法では許容的な群がすべて選ばれていればよく、選ばれた群の中に許容的でない群が含まれていても正しい選択とみなす。このため、必ずしも選ばれた群がすべて許容的であるというわけではない。これに対し、Tong は興味あるパラメータが正規分布の平均値の場合に複数の群の中から対照群より大きな平均値をもつ群とそうでない群に二分するための選択ルールを分散が既知の場合と未知の場合について与えている³⁶。興味あるパラメータが二項確率の場合にも同様の問題を考えることができるが、Tong のようにある対照と比較して、それより優るものと劣るものに二分するため選択ルールの最適性についてはほとんど検討されていない。

このため本節では、有効率を指標として複数の試験治療法の中から対照治療法に優る試験治療法のみを選択したいという状況に対して、indifference-zone アプローチを用いた試験デザインを提案する。ここでは、データがすべて収集された時点で一度だけ選択ルールを適用して治療群を選別する one-stage デザインと、試験の途

中で中間解析を行い，その時点で明らかに対照治療法に劣っていると判断される試験治療法があれば，その群への被験者の割付を打ち切ってしまう two-stage デザインを検討する．選択ルールは有効割合に基づいた単純なルールを用い，あらかじめ設定した選択確率を満たすために必要な被験者数が最小となるように試験をデザインすることを目的とする．ただし，two-stage デザインの場合は被験者数が確率変数となるので，期待被験者数が最小となるように中間解析時点と選択ルールを設定する．試験を two-stage で実施することにより，明らかに対照に劣った治療法を受けてしまう被験者の数を減らすことができるため，複数の試験治療群を同時においた臨床試験ではその利点は大きいと考えられる．

4.1 Indifference-zone

対照治療法を T_c ，試験治療法を T_1, \dots, T_t で表し，対応する治療法の有効率をそれぞれ $\pi_c, \pi_i, (i = 1, \dots, t)$ とする．indifference-zone の幅を Δ とし，選択すべき試験治療法の集合を

$$S = \{T_i \mid \pi_i \geq \pi_c\}$$

選択すべきでない試験治療法の集合を

$$R = \{T_i \mid \pi_i \leq \pi_c - \Delta\}$$

とする．また，

$$I = \{T_i \mid \pi_c - \Delta < \pi_i < \pi_c\}$$

に含まれる治療法は indifference-zone アプローチの考え方から，選択されても，されなくてもよいとする．目的は S に含まれる試験治療法 T_i をすべて選択し， R に

含まれる治療法はすべて棄却することである。従って、 P_{CS} は次のように定義される。

$$P_{CS} = \Pr \{ \text{select all } T_i \in \mathcal{S}, \text{ reject all } T_i \in \mathcal{R} \}$$

一般に選択すべき試験治療法の数は未知であるが、仮にその数が j ($1 \leq j \leq t$) であったとすると、indifference-zone アプローチにおいてこの P_{CS} が最小となるのは次の条件が成り立つときである。

$$\pi_{[1]} = \cdots = \pi_{[t-j]} = \pi_c - \Delta, \pi_{[t-j+1]} = \cdots = \pi_t = \pi_c$$

ただし、 $\pi_{[j]}$ は π_1, \dots, π_t の中で大きさが j 番目の値であるとする。従って、このパラメータ配置 (以下、 Π^*) における P_{CS} のみを考え、その値が事前に定めた選択確率以上となるように試験をデザインすることを考える。

4.2 One-stage デザイン

$(t+1)n$ 名の被験者を対照治療群と t 個の試験治療群に n 例ずつ割り付けるとする。 y_{ij} ($i = 1, 2, \dots, t, c, j = 1, \dots, n$) を治療法 T_i に割り付けられた j 番目の被験者の反応を表す変数とする。 y_{ij} は二値とし、割り付けられた治療法が有効であれば $y_{ij} = 1$ 、無効であれば $y_{ij} = 0$ とする。また、

$$y_i = \sum_{j=1}^n y_{ij}, \quad \bar{y}_i = y_i/n$$

と定義する。このとき、整数 δ に対し、次のような選択ルールを考える。

$$\mathcal{S}^* = \{T_i : \bar{y}_i > \bar{y}_c - \delta/n\}$$

つまり、試験の結果、試験治療群の中で有効割合が $\bar{y}_c - \delta/n$ よりも大きかった治療法をすべて選択するということである。このとき、治療法 T_i が選択される確率

$\Pr\{T_i \in \mathcal{S}^*\}$ は次のように表される.

$$\begin{aligned}\Pr\{T_i \in \mathcal{S}^*\} &= \sum_{x=0}^n \Pr\{\bar{y}_i > \bar{y}_c - \delta/n \mid y_c = x\} \Pr\{y_c = x\} \\ &= \sum_{x=0}^n \sum_{y_i=\max(0, x-\delta+1)}^n \binom{n}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n-y_i} \binom{n}{x} \pi_c^x (1 - \pi_c)^{n-x}\end{aligned}$$

便宜上, \mathcal{S} , \mathcal{R} に含まれる治療法がそれぞれ j , $t-j$ 個あるとし,

$$\mathcal{S} = \{T_1, \dots, T_j\}$$

$$\mathcal{R} = \{T_{j+1}, \dots, T_t\}$$

とする. \mathcal{S} に含まれる治療群をすべて選択し, \mathcal{R} に含まれる治療群はすべて棄却することが正しい選択であり, indifference-zone アプローチにおける P_{CS} の最小値 P_{CS}^* は次のように表される.

$$\begin{aligned}P_{CS}^* &= \sum_{x=0}^n \Pr\{\bar{y}_1, \dots, \bar{y}_j > x/n - \delta/n, \bar{y}_{j+1}, \dots, \bar{y}_t \leq x/n - \delta/n \mid y_c = x\} \Pr\{y_c = x\} \\ &= \sum_{x=0}^n A(\pi_c, x)^j B(\pi_c, x)^{t-j} \binom{n}{x} \pi_c^x (1 - \pi_c)^{n-x}\end{aligned}$$

ここで,

$$\begin{aligned}A(\pi, x) &= \sum_{y=\max(0, x-\delta+1)}^n \binom{n}{y} \pi^y (1 - \pi)^{n-y} \\ B(\pi, x) &= \sum_{y=0}^{x-\delta} \binom{n}{y} (\pi - \Delta)^y [1 - (\pi - \Delta)]^{n-y}\end{aligned}$$

である.

この P_{CS}^* が事前に定めた値 P^* よりも大きくなるように n , δ を定めることがここの試験デザインである. この P^* は 0.80 や 0.90 といった十分大きな値に設定されることになる. ただし, \mathcal{S} に含まれる治療法の数 j や π_c の値は一般に未知で

ある。このため、 π_c の値はそれまでに行われた他の試験結果などから得られた推定値 $\hat{\pi}_c$ を用いることにし、すべての j に対し、 $P_{CS}^* \geq P^*$ となるように n, δ を定めることにする。

ここで、indifference-zoneの幅が Δ であることから、 $\delta/n = \Delta$ となるように与えればいよいにも思えるが、ここで設定した条件の下で治療法を正しく選択できる確率 P_{CS} は $\delta/n = \Delta$ のときに必ずしも最大とはならない。このため、この δ を変化させ、 P_{CS} が最大となる δ を特定する必要がある。

理論的には、上に示した P_{CS}^* の式を n, δ の関数と見なし、各 j に対し、 $P_{CS}^* \geq P^*$ となる最小の n とそのときの δ の値を求めるという問題であるが、その解法は非常に困難である。このため、ここでは P_{CS} の計算が比較的容易であることを利用し、数値計算により n が最小となる δ を探索的に求めることにする。実際の数値計算による試験デザイン例については4.4節で述べる。

なお、この方法では $j = 0$ 、つまり、選択すべき試験治療群が全く存在しない場合にすべての試験治療群が棄却される確率も P^* 以上である。従って、仮説検定における第一種の過誤に相当する確率は $1 - P^*$ 以下であるといえる。

4.3 Two-stage デザイン

多群比較の場合には必然的にかなり多くの被験者数が必要となる。試験の計画時には、対照治療法に劣る試験治療法の数は未知であるが、可能な限り対照治療法よりも劣った治療法を受けてしまう被験者数を少なくすることは臨床試験を実施する上で非常に重要である。

劣った治療法を受ける被験者数を減らす一つの方法として、試験の途中で中間

解析を行い、対照治療よりも劣ると判断された治療法をなるべく早く試験から削除するという方法が考えられる。本節では、このような目的を含んだ試験デザイン法として、試験の途中で中間解析を一度だけ行う two-stage デザインを提案する。two-stage デザインでは中間解析の結果から対照治療よりも明らかに劣っていると考えられる治療法があった場合には、その試験治療群を試験から削除し、残った試験治療群のみでその後の試験を継続する。その後、すべてのデータが収集された時点で試験が継続された試験治療群の中から選択すべき試験治療法を特定することになる。このため、two-stage デザインでは中間解析時と最終時点それぞれに対し選択ルールが必要となり、被験者数も確率変数となる。従って、選択ルールの設定は期待被験者数が最小となるように行う。

4.3.1 Two-stage デザインと選択確率

n_1 を最初の試験段階 (first stage) で各治療群に割り付ける被験者数とし、 n_2 を中間解析後 (second stage) に割り付ける被験者数とする。従って、 $n = n_1 + n_2$ が各治療群に割り付けられる最大の被験者数であり、 $i = 1, 2, \dots, t, c$ に対し、

$$y_{i_1} = \sum_{j=1}^{n_1} x_{ij}, \quad y_{i_2} = \sum_{j'=n_1+1}^n x_{ij'}, \quad y_i = y_{i_1} + y_{i_2}$$

$$\bar{y}_{i_1} = \frac{y_{i_1}}{n_1}, \quad \bar{y}_{i_2} = \frac{y_{i_2}}{n_2}, \quad \bar{y}_i = \frac{y_i}{n}$$

と定義する。あらかじめ正の整数値 δ_1, δ_2 を定め、中間解析時と試験終了時において次のような二つの選択ルールを与える。

$$\text{中間解析時} : \mathcal{S}_1^* = \{T_i : \bar{y}_{i_1} > \bar{y}_{c_1} - \delta_1/n_1\}$$

$$\text{試験終了時} : \mathcal{S}_2^* = \{T_i : T_i \in \mathcal{S}_1^*, \bar{y}_i > \bar{y}_c - \delta_2/n\}$$

中間解析時に選択ルール S_1^* を満たさない試験治療群はその時点で試験から削除され、残った試験治療群のみで second stage を継続する。最終的に second stage まで進んだ試験治療群の中で S_2^* を満たした試験治療法のみが選択されるとする。

このとき、 n_1 および n_2 が固定されているとすると、試験治療法 T_i が選択される確率、 $P(T_i \in S_2^*)$ 、は次のように表される。

$$\begin{aligned} P(T_i \in S_2^*) &= \sum_{x_2=0}^{n_2} \sum_{x_1=0}^{n_1} \Pr\{T_i \in S_2^* \mid y_{c_1} = x_1, y_{c_2} = x_2\} \Pr\{y_{c_1} = x_1, y_{c_2} = x_2\} \\ &= \sum_{x_2=0}^{n_2} \sum_{x_1=0}^{n_1} C(\pi_i, x_1, x_2) \binom{n_1}{x_1} \pi_c^{x_1} (1 - \pi_c)^{n_1 - x_1} \binom{n_2}{x_2} \pi_c^{x_2} (1 - \pi_c)^{n_2 - x_2} \end{aligned}$$

ここで、

$$\begin{aligned} C(\pi, x_1, x_2) &= \Pr\{\bar{y}_{i_1} > \bar{y}_{c_1} - \delta_1/n_1, \bar{y}_i > \bar{y}_c - \delta_2/n, \mid y_{c_1} = x_1, y_{c_2} = x_2\} \\ &= \Pr\{y_{i_1} > x_1 - \delta_1, y_i > x_1 + x_2 - \delta_2\} \\ &= \sum_{y_{i_1}=d_1}^{n_1} \sum_{y_{i_2}=d_2}^{n_2} \binom{n_2}{y_{i_2}} \pi^{y_{i_2}} (1 - \pi)^{n_2 - y_{i_2}} \binom{n_1}{y_{i_1}} \pi^{y_{i_1}} (1 - \pi)^{n_1 - y_{i_1}} \end{aligned}$$

$$d_1 = \max(0, x_1 - \delta_1 + 1), \quad d_2 = \max(0, x_1 + x_2 - \delta_2 - y_{i_1} + 1)$$

である。仮に選択すべき試験治療法の数 j であったとすると、 P_{CS}^* の最小値は次のように表される。

$$P_{CS}^* = \sum_{x_2=0}^{n_2} \sum_{x_1=0}^{n_1} [C(\pi_c, x_1, x_2)]^j [1 - C(\pi_c - \Delta, x_1, x_2)]^{t-j} \Pr\{y_{c_1} = x_1\} \Pr\{y_{c_2} = x_2\}$$

One-stage デザインに比べより多くの変数が含まれるが、同様の考え方により (δ_1, δ_2) と (n_1, n_2) を設定することができる。ただし、one-stage デザインとは異なり、被験者数も確率変数となるため、次節で示す期待被験者数が最小となるように (δ_1, δ_2) および (n_1, n_2) を定めることとする。

4.3.2 期待被験者数

期待被験者数も選択すべき試験治療の数 j に依存するため、はじめに j を固定して考える。このときの期待被験者数 $E_j(N)$ は次のように表される。

$$E_j(N) = n(t+1) - n_2 \sum_{h=0}^t h \Pr\{|S_1^c| = h\}$$

ただし、 $|S_1^c|$ は中間解析後に削除された治療群の数であり、 $\Pr\{|S_1^c| = h\}$ は以下のように表される。

$$\begin{aligned} \Pr\{|S_1^c| = h\} &= \sum_{x_1=0}^{n_1} \sum_{h_1=0}^h \binom{j}{h_1} D(\pi_c, x_1)^{j-h_1} [1 - D(\pi_c, x_1)]^{h_1} \\ &\quad \times \binom{t-j}{h-h_1} D(\pi_c - \Delta, x_1)^{t-j-(h-h_1)} \\ &\quad \times [1 - D(\pi_c - \Delta, x_1)]^{h-h_1} \Pr\{y_{c_1} = x_1\} \end{aligned}$$

ここで、

$$D(\pi, x) = \sum_{y=\max(0, x-\delta+1)}^{n_1} \binom{n_1}{y} \pi^y (1-\pi)^{n_1-y}$$

である。一般に j の値は未知であるため、ここでは $j = 0, 1, \dots, t$ が等確率で生じると仮定して、その期待値 $E(N)$ で期待被験者数を評価することにする。

$$E(N) = \sum_{j=0}^t E_j(N)/(t+1)$$

つまり、この $E(N)$ が最小となるように $(n_1, n_2), (\delta_1, \delta_2)$ を定める。

当然のことではあるが、何らかの情報により j に対し他の分布が仮定できる場合にはその分布の下で $E(N)$ が最小となるように $(n_1, n_2), (\delta_1, \delta_2)$ を定めればよい。

4.4 数値計算による試験デザイン例と被験者数の比較

前節で提案した試験デザイン法について,

$$t = 2, 3, 4, \pi_c = 0.20(0.10)0.90$$

$$\Delta = 0.10, 0.15, 0.20, P_{CS} = 0.80$$

とした場合の被験者数および δ (もしくは (δ_1, δ_2)) を数値計算により求めた結果を表 40~45 (p.129 ~ 131) に示した. これらの表から, この試験デザインについていくつかの特徴を述べることができる.

まず, One-stage デザイン (表 40~42) では, 必要被験者数を最小にする δ の値が,

$$\delta/n \simeq \Delta/2$$

となることが分かる. これは, δ の値が大きすぎると許容的でない治療法を選択してしまう確率が増えるが, 逆に小さすぎると許容的な治療法を選択できない確率が増加するためであり, これらの確率の増減が indifference-zone の幅 Δ の中間で相殺されるためと考えられる. また, P_{CS} の値が最も小さくなる j の値は, t が偶数であれば $j = t/2$ のときであり, 奇数の時は $j = t/2 - 1$ もしくは $j = t/2 + 1$ のときである. これは δ の値が $\Delta/2$ に近いことから, 許容的あるいは非許容的のどちらか一方に試験治療法の数に偏っていた方が分離しやすいためである.

$\Delta = 0.20$ とした場合の two-stage デザインにおける選択ルールと期待被験者数を表 43 ~ 45 (p.132 ~ 134) に示した. この結果は $n = n_1 + n_2$ を t および π_c の値が対応する one-stage デザインで得られた被験者数と同じ値に固定して n_1, δ_1, δ_2 を得たものである. Two-stage デザインでは one-stage デザインで得られた被験者

数よりも最大被験者数を多くすることにより、期待被験者数をより小さくできる可能性もある。しかし、臨床試験を行う際には被験者数を数多く確保できることは少なく、ここでは one-stage デザインに必要な被験者数で two-stage デザインを採用した場合にどの程度期待被験者数が少なくなるかを検討するため、最大被験者数を one-stage デザインで必要な被験者数に固定して検討した。

表からいずれの場合も δ_1, n_1 の方が、 δ_2, n_2 よりも大きく、 $\delta_1/n_1 \simeq \Delta$ であることが分かる。期待被験者数はパラメータ値により違いがあるものの最大被験者数に比べ数%から 10%程度小さくなる。ここでは最大被験者数を one-stage デザインで得られた必要被験者数 (N_1) としていることから、first-stage 後に許容的でない試験治療群を削除できる確率はそれほど大きくなく、期待被験者数もそれほど大きく低下はしない。しかし、より劣った試験治療群が存在したならばその群を削除できる可能性はより大きくなり、複数の試験治療群を同時に比較する臨床試験の場合には非常に好ましい性質である。

4.4.1 試験デザイン例

4.2, 4.3 節で提案した試験デザインが適用可能な例として細菌感染症治療薬（抗菌薬）の開発における用量設定試験があげられる。抗菌薬の臨床評価ガイドライン³⁷によると後期第二相の段階においてそれまでの臨床、非臨床試験の結果を踏まえて実際の臨床用量を確認するため、2~3 用量の試験治療群に対照治療群をおいた「臨床用量確認試験」を実施することになっている。抗菌薬の有効性の評価は、疾患の原因となっている細菌が死滅したか、存続しているかという二値で行うため、この用量確認試験に本節で提案した one-stage および two-stage での試験

デザインを適用することが可能である。

例えば、高用量、低用量の2つの用量群と対照治療群を設定し、対照治療の有効率が90%と推定されているとする。Indifference-zoneの幅 Δ を0.2とした場合にOne-stageデザイン(表40)の必要被験者数は一群38例であり、選択ルールは、

$$S^* = \{T_i : \bar{y}_i > \bar{y}_c - 0.105\}$$

である。また、two-stageデザインを採用した場合の被験者数は、first-stageが20例、second-stageが18例で、選択ルールは、

$$S_1^* = \{T_i : \bar{y}_i > \bar{y}_c - 0.200\}$$

$$S_2^* = \{T_i : \bar{y}_i > \bar{y}_c - 0.105\}$$

となる(表43)。one-stageデザインでの必要被験者数は114例であり、two-stageデザインでの期待被験者数は103例であることから、two-stageデザインでは被験者数を平均的に11例少なくすることが可能である。

これまで抗菌薬の臨床用量確認試験では最終的な試験治療群の選択ルールが明確でなく、被験者数も実施可能な例数ということで一群10例から50例程度まで試験によって様々であった。この試験デザイン法は少なくとも選択ルールの設定根拠と必要被験者数を統計学的に評価可能であるという点で、これまでの曖昧な試験計画と選択ルールを改善する一つの方法論となると考えられる。

4.5 Weighted assignment デザイン

これまで示してきた試験デザイン法では、対照治療群の有効割合が試験治療法の選択ルールの基準になるため、対照治療法の有効率の推定精度をあげることに

より選択確率が増加する可能性も考えられる。このため、ここでは対照治療群への割付被験者数を試験治療群よりも多くした場合の試験デザインについて検討する。

4.5.1 One-stage デザイン

n_c および n をそれぞれ対照治療群と試験治療群に割り付ける被験者数とする。選択ルールは 4.2 節の one-stage デザインと同様とすると、 P_{CS} の最小値 P_{CS}^* は以下のように表される。

$$P_{CS}^*(\pi_c, \delta) = \sum_{x=0}^{n_c} \tilde{A}(\pi_c, x)^j \tilde{B}(\pi_c, x)^{t-j} \binom{n_c}{x} \pi_c^x (1 - \pi_c)^{n-x}$$

ここで、

$$\begin{aligned} \tilde{A}(\pi, x) &= \sum_{y=\max(0, \lceil nx/n_c - \delta \rceil + 1)}^n \binom{n}{y} \pi^y (1 - \pi)^{n-y} \\ \tilde{B}(\pi, x) &= \sum_{y=0}^{\lfloor nx/n_c - \delta \rfloor} \binom{n}{y} (\pi - \Delta)^y [1 - (\pi - \Delta)]^{n-y} \end{aligned}$$

従って、4.2 節と同様の方法で δ , n_c および n を定めることができる。ただし、割付比 $\gamma = n_c/n$ は固定して考えることにする。割付比を変えて最適な δ や n_c , n を設定することも理論的に可能であるが、計算が煩雑になるだけでなく、実際の割付自体が難しくなってしまうため、ここでは割付比は固定して考える。

4.5.2 Two-stage デザイン

Two-stage design も考え方は同じであるが、割付比 $n_{c_1}/n_1 = n_{c_2}/n_2 = \gamma$ は試験計画時に定められ、その比は first stage および second stage で同一とする。 n_{c_1} および n_{c_2} をそれぞれ対照治療群に割り付けられる first stage, second stage にお

ける被験者数とし, $n_{c_1} + n_{c_2} = n_c$ とする. この他の記号の定義は前節と同様とする. このとき, P_{CS} の最小値 P_{CS}^* は次のように表される.

$$\begin{aligned}
P_{CS}^*(\pi_c, \delta_1, \delta_2) &= \sum_{x_2=0}^{n_{c_2}} \sum_{x_1=0}^{n_{c_1}} [\tilde{C}(\pi_c, x_1, x_2)]^j [1 - \tilde{C}(\pi_c - \Delta, x_1, x_2)]^{t-j} \\
&\quad \times \Pr\{y_{c_1} = x_1\} \Pr\{y_{c_2} = x_2\} \\
&= \sum_{x_2=0}^{n_{c_2}} \sum_{x_1=0}^{n_{c_1}} [\tilde{C}(\pi_c, x_1, x_2)]^j [1 - \tilde{C}(\pi_c - \Delta, x_1, x_2)]^{t-j} \\
&\quad \times \binom{n_{c_1}}{x_1} \pi_c^{x_1} (1 - \pi_c)^{n_{c_1} - x_1} \binom{n_{c_2}}{x_2} \pi_c^{x_2} (1 - \pi_c)^{n_{c_2} - x_2}
\end{aligned}$$

ここで,

$$\begin{aligned}
\tilde{C}(\pi, x_1, x_2) &= \sum_{y_1=d_1}^{n_1} \sum_{y_2=d_2}^{n_2} \binom{n_2}{y_2} \pi^{y_2} (1 - \pi)^{n_2 - y_2} \binom{n_1}{y_1} \pi^{y_1} (1 - \pi)^{n_1 - y_1} \\
d_1 &= \max(0, \lceil n_1 x_1 / n_{c_1} - \delta_1 \rceil + 1) \\
d_2 &= \max(0, \lceil n(x_1 + x_2) / n_c - \delta_2 - y_{i_1} \rceil + 1)
\end{aligned}$$

前節同様, 試験デザインの最適性は期待被験者数 $E(N)$ で評価することとし, $E(N)$ が最小となるように (n_{c_1}, n_{c_2}) , (n_1, n_2) , および (δ_1, δ_2) を設定することにする.

4.5.3 Weighted assignment デザインでの選択ルールと被験者数

表 46~48 (p.135 ~ 137) は $\gamma = 1.5$ としたときの weighted one-stage デザインの結果であり, N_{w1} は必要被験者数である. なお, γn_i が整数でない場合には四捨五入することとした. これらの表から割付比 $\gamma = 1.5$ で試験をデザインした場合, π_c の値によっては割付比 $\gamma = 1$ の one-stage デザインに比べ, 必要被験者数を数%程度低下させることができることがわかる. 特に π_c の値が小さいときには割付比が

1 の two-stage デザインの被験者数の期待値よりも小さくなる。ただし、 π_c の値が 1 に近いほどその比は小さくなり、 t が小さいと $\pi_c = 0.90$ では逆に必要被験者数は多くなってしまうこともある。 δ/n の値は割付比 1 の one-stage デザインと同様 $\Delta/2$ の値に近く、また、試験治療群に割付けられる被験者数は割付比 1 のときの約 80%~90%である。

表 49 (p.138) は $\gamma = 1.5$ としたときの weighted two-stage デザインであり、4.3 節の two-stage デザインと同様、最大被験者数を対応する weighted one-stage デザインで得られた値に固定して期待被験者数が最も小さくなるように算出した結果である。weighted one-stage デザインよりもさらに期待被験者数を 3% ~ 10%程小さくなるのが分かる。

これらの結果から、対照群に被験者を多く割り付けることにより π_c の値によっては必要被験者数を数%程度減らすことが可能であることがわかる。ただし、対照群に被験者を多く割り付けるということはそれだけ試験治療群に割り付けられる被験者数が少なくなるため、試験治療の有効率の推定には不利になる。従って、試験治療群の被験者数が許容できる範囲内で割り付け比を定める必要があるといえる。

5 まとめ

本稿では、臨床試験における中間解析手法である α -spending function 法と二値データに対する indifference-zone アプローチを応用した試験デザインについて検討した。以下ではそれぞれの手法を分けて議論することにする。

1 の two-stage デザインの被験者数の期待値よりも小さくなる。ただし、 π_c の値が 1 に近いほどその比は小さくなり、 t が小さいと $\pi_c = 0.90$ では逆に必要被験者数は多くなってしまうこともある。 δ/n の値は割付比 1 の one-stage デザインと同様 $\Delta/2$ の値に近く、また、試験治療群に割付けられる被験者数は割付比 1 のときの約 80%~90%である。

表 49 (p.138) は $\gamma = 1.5$ としたときの weighted two-stage デザインであり、4.3 節の two-stage デザインと同様、最大被験者数を対応する weighted one-stage デザインで得られた値に固定して期待被験者数が最も小さくなるように算出した結果である。weighted one-stage デザインよりもさらに期待被験者数を 3% ~ 10%程小さくなるのが分かる。

これらの結果から、対照群に被験者を多く割り付けることにより π_c の値によっては必要被験者数を数%程度減らすことが可能であることがわかる。ただし、対照群に被験者を多く割り付けるということはそれだけ試験治療群に割り付けられる被験者数が少なくなるため、試験治療の有効率の推定には不利になる。従って、試験治療群の被験者数が許容できる範囲内で割り付け比を定める必要があるといえる。

5 まとめ

本稿では、臨床試験における中間解析手法である α -spending function 法と二値データに対する indifference-zone アプローチを応用した試験デザインについて検討した。以下ではそれぞれの手法を分けて議論することにする。

5.1 仮説検定を基にした中間解析手法

本稿では α -spending function 法に関する適用上の問題点と解析時点の最適性について議論した。 α -spending function 法は Lan and DeMets² により提案された手法であり、統計学的には中間解析手法と確率過程論を統合した理論構築がなされたという点で意義のある手法である。実際の応用面においては、それまで事前に解析回数を定め、情報量の増加に対し等間隔でしか解析ができなかった制約を解放し、任意の解析時点で棄却限界値を算出できるという点で非常に適用しやすい方法である。しかしながら、柔軟な方法であるが故に用いるべき α -spending function や解析時点、解析間隔の最適性についての議論は非常に難しいといえる²³。どのような α -spending function を定義しても、検出力や被験者数といった指標に対して常に最適となるものを定義することは困難である。このため、むしろ許容可能な範囲と許容できない範囲を検討することにより、最適ではないにしても良好な結果を得ることができるような中間解析計画が重要であると考えられる。その一つの例が中間解析を行う時点の問題である。 α -spending function 法では任意の解析時点で棄却限界値を算出できることが一つの大きな利点ではあるが、中間解析を行う一つの目的は可能な限り試験結果を早期に判断して次のステップへ進むということである。試験の早期停止が解析時点に依存する以上、その目的を最大限に達成できるように解析時点を選択することは自然な要求である。しかしながら、解析時点が任意に選択できる α -spending function 法に関してはこれまでどのように解析時点を選ぶべきかという提案があまりなされていなかった。 Li and Geller²³ は α を徐々に多く消費していく方が望ましいという観点から

解析間隔の最適性について論じており、Kim and DeMets²² はいくつかの解析時点に対し α -spending function 法を用いた場合の期待試験停止時間を算出しているが、具体的に α -spending function を与えたときにどの時点で解析すべきかという提案まではなされなかった。本稿ではこの問題を検討するため、O'Brien-Fleming 型、Pocock 型、直線型の 3 つの α -spending function に対し、解析時点と解析回数を変えたときにそれぞれの検出力や期待被験者数がどのように変化するかを数値計算により検討した。その結果、用いる α -spending function によって期待被験者数や検出力の低下が異なるだけでなく、同じ α -spending function を用いても解析時点によってこれらがかなり異なることが示された。また、検出力一定という条件下で期待被験者数が最小となる解析時点についても検討した。その結果、O'Brien-Fleming 型を用いた場合には information fraction が $1/2$ を越えた時点から解析回数に応じて information fraction に対しほぼ等間隔に解析するとよく、また、Pocock 型や直線型を用いた場合には information fraction に対し全体的に等間隔に解析を行えばよいとの結果が得られた。つまり、これらの α -spending function を用いる限り、あまり極端な解析計画は期待被験者数の観点からは好ましくないといえる。この結果は試験計画時に解析回数やデータモニタリングの時期を検討する上で非常に重要な情報になると考えられる。臨床試験の場合、情報量の増加は必ずしも時間に対して一定ではないため、定期的にデータモニタリングを行えばいいというわけではないが、これらの結果を踏まえ、各データモニタリング時にそれまでのデータ集積量を考慮しながら次回のデータモニタリング時点を検討することは可能であると思われる。また、試験開始後のデータモニタリング時においても次回以降どの時点で解析を行うべきかを検討するための指針となると考

えられる。ただし、中間解析結果に依存して解析時点が決められるとすると本稿 2.4 節で取り上げた α -spending function 法の乱用となる。しかし、解析時点と用いる α -spending function との関係から次の解析時点を選択するのであれば第一種の過誤への影響はないため、このような解析時点の選択方法は積極的に用いることができる。

本稿 3.1 節で提案した棄却限界値の算出方法は、解析時点を一度だけ中間解析結果に依存して増やす場合の提案であるが、複数回の解析時点の変更については複雑な条件付き確率を評価する必要がある、非常に難しい問題である。Lan and DeMets³ や Proschan ら⁴ が検討した状況も解析時点は一時点の中間解析結果にのみ依存する場合であり、解析時点が複数回の中間解析結果に依存したときに第一種の過誤確率がどの程度増大するかは未解決である。

本稿では取り上げなかったが、 α -spending function 法を適用する上でもう一つの大きな問題点は information fraction の推定である。2.1.3 節で定義したように information fraction は検定統計量に対する各解析時点での Fisher 情報量と最終解析時点での Fisher 情報量との比であり、中間解析を行う場合には、

$$(\text{解析時点}) = (\text{information fraction})$$

として議論される。臨床試験では、多くの場合、中間解析時点において最終解析時点での Fisher 情報量は未知であることから、何らかの推定値を用いる必要がある。もしこの推定値が実際よりも過小評価された場合には試験の途中で information fraction の値が 1 を越えるため、その後の棄却限界値は定義できなくなる。この問題に対する一般的な対処方法は最終時点での Fisher 情報量を過大評価しておくことで

あるが、便宜的に過大評価するというだけで理論的な裏付けはない³⁸。information fraction をどのように推定すべきか、あるいは information fraction に依存しない中間解析方法が存在するかどうかについては今後の研究課題の一つである。また、試験結果が中間解析時に有意になったにもかかわらずその後も試験が継続された場合 (overrunning) や有意でないにもかかわらず試験を終了した場合 (underrunning) に対し、その後の解析方法や解析結果の解釈についても統計学的に解決しているわけではない。Whitehead³⁹ が overrunning や underrunning が生じた試験でのパラメータの推定方法について提案しているが、それ以外の議論は今のところほとんどない。

本稿では、推定の問題については触れなかったが、中間解析を含む試験デザインにおけるパラメータ推定についても多くの議論がある。Jennison and Turnbull⁴⁰ は中間解析時の棄却限界値を用いたパラメータの区間推定法 (repeated confidence interval) を提案しており、また、Whitehead⁴¹ は有意確率を示す P 値関数と呼ばれる量を定義して、その値を用いてパラメータの点推定や区間推定を行う方法を示している。中間解析を行う試験では、固定デザインでの推定方法を用いることは不適切であり、このような中間解析を行うことを考慮した推定方法を採用しなければならない。

5.2 Indifference-zone アプローチ

本稿では、indifference-zone アプローチを用いて試験をデザインした場合の中間解析の利点についても検討した。初めにも述べたように、中間解析を含む臨床試験デザインは検証的な比較試験で多く試みられており、特に長期間にわたる大規

模臨床試験では必須のものとなっている。一方で、開発の初期の段階におけるスクリーニングを目的とした試験のように、複数の治療法を同時に比較するような試験での中間解析手法についてはあまり提案されていない。このような試験では検証的な比較試験に進む前段階として複数の治療法の中に有望なものがいくつか存在したならば、それらをすべて選択してさらに追加の試験を行ったり、あるいは別の評価指標からその後のステップに進む治療法を選択していくという方法が採られることが多い。このため、検出力重視の試験デザインが望ましい。従って、仮説検定のような検証的な手法を用いる方法以外に、有効性のような一つの指標に対してある許容範囲を設け、その許容範囲を越えて劣らないものはすべて選択するという試験デザインも考えられる。例えば、二値データの発現割合に関し、対照群と複数の試験群を比較する一般的な方法として、古くは、Paulson⁴²が逆 sin 変換を用いた正規近似により複数の試験群と対照群の中から最も良い(有効割合が大きい)ものを選ぶ方法を提案しており、Gupta and Sobel³⁴は indifference-zone アプローチの観点から対照よりも優る試験群をすべて選択する方法を示している。このような選択問題の臨床試験への応用は本稿 2.6 節で示した方法以外にも Simon⁴³や Thall⁴⁴などが主に検証試験前の試験治療法のスクリーニング目的の試験デザインとして提案している。ただし、これらの方法は有効割合が最も大きい治療法を選ぶことや許容的な治療法を含む集合を選ぶことを目的とした方法であり、必ずしも許容範囲以上の治療法だけがすべて選ばれることを目的としているわけではない。

一方、ある許容範囲を定義してその範囲に含まれる群と含まれない群を完全に二分することは統計学的にも非常に厳しい制約であり、多くの被験者数を必要と

する。本稿では、このような問題に対して期待被験者数を最小にすることを選択ルールの指標とし、存在する許容的な治療法の数に関わらず正しい選択を行う確率が一定値以上になる one-stage および two-stage での試験デザインとそれぞれの必要被験者数および期待被験者数を検討した。ここで、two-stage の試験デザインとは試験の途中で中間解析を一度行い、明らかに対照治療に劣っていると判断された治療群が存在した場合にはその治療群を削除し、残った治療群のみで試験を継続するという方法である。複数の試験治療群が設定されているということはそれだけ被験者数も多く必要なり、また、必ずしもすべての試験治療が十分な効果を持つとは限らない。このため、このようなデザインを採用することで劣った治療法を不必要に多くの被験者に施すことを避けることができ、試験全体での期待被験者数を低下させることができる利点がある。本稿ではこのような試験デザインを indifference-zone アプローチを用いて行い、期待被験者数が最小となる選択ルールと各 stage に必要な被験者数を数値計算により与えた。この結果、two-stage デザインでは中間解析を行わない one-stage デザインよりも期待被験者数を約 10% 程度少なくできることが示された。

さらに、試験治療群よりも対照治療群へより多くの被験者を割り付けることで、若干ではあるが、期待被験者数を小さくすることができることも示した。試験治療群よりも対照治療群に多くの被験者を割り付けることにより試験治療の効果の推定には不利なデザインになるため、実際にこのような試験デザインを採用すべきかどうかは議論のあるところである。逆に対照治療群よりも試験治療群の方に多くの被験者を割り付けるという方法も考えられるため、それぞれの利点と欠点についてさらなる検討が必要である。

また、Dunnett の多重比較法⁴⁵ のような対照群と複数の試験群との比較を前提とした仮説検定手法と今回提案した選択問題としての one-stage および two-stage デザインとの比較ということも今後検討する必要があると考えられる。試験を計画する際にどちらの方法を採用するとよいかという指針を与えることにより、試験デザインの選択の幅をより広げることが可能であると考えられる。

現在、臨床試験では仮説検定に基づく方法が圧倒的に多く採用されており、選択問題としての試験デザインはあまり普及していない。二群比較の場合はどちらも同じ論理に帰着するため本質的な違いはないが、複数の試験治療群を比較する場合はお互い結果の解釈の仕方が異なるため、評価指標や試験の目的、統計的方法の性質を考慮して選択すべき問題であると考えられる。しかしながら、現在ではこれらの試験デザインの比較に関して理論的な根拠というものは考慮されていない。数理統計学の分野では、統計的実験を一つの集合と考えて半順序を定義し、実験間の比較を行うという理論構築もなされている⁴⁶。特に(完全)逐次的な二項実験に対しては、完備類を与える研究結果も得られている⁴⁷。これまで臨床試験デザインの検討にこれらの理論的なデザイン比較の研究結果を取り入れた試みは少ないが、よりよい臨床試験デザインを追求する上でこのような数理統計学における理論結果を応用することも今後の検討課題の一つであると考えられる。

表 18: 中間解析に依存した解析計画に対する棄却限界値

$\alpha_1^*(t)$					
t_1	t_2	$b(t_1)$	$b(1)$	$\tilde{b}(t_2)$	$\tilde{b}(1)$
0.2	0.30	4.878	1.959	5.762	3.606
	0.40	4.878	1.959	5.347	3.609
	0.50	4.878	1.959	4.937	3.617
	0.60	4.878	1.959	4.583	3.632
	0.70	4.878	1.959	4.282	3.653
	0.80	4.878	1.959	4.024	3.679
	0.90	4.878	1.959	3.801	3.707
0.3	0.40	3.929	1.954	4.730	3.507
	0.50	3.929	1.954	4.554	3.516
	0.60	3.929	1.954	4.315	3.532
	0.70	3.929	1.954	4.082	3.554
	0.80	3.929	1.954	3.869	3.582
	0.90	3.929	1.954	3.676	3.610
0.4	0.50	3.357	1.939	4.067	3.395
	0.60	3.357	1.939	3.995	3.413
	0.70	3.357	1.939	3.849	3.437
	0.80	3.357	1.939	3.688	3.466
	0.90	3.357	1.939	3.531	3.495
0.5	0.60	2.963	1.916	3.603	3.278
	0.70	2.963	1.916	3.584	3.306
	0.80	2.963	1.916	3.490	3.337
	0.90	2.963	1.916	3.375	3.368
0.6	0.70	2.669	1.889	3.258	3.157
	0.80	2.669	1.889	3.268	3.194
	0.90	2.669	1.889	3.207	3.229
0.7	0.80	2.438	1.862	2.988	3.026
	0.90	2.438	1.862	3.016	3.071
0.8	0.90	2.250	1.834	2.770	2.875

表 19: 中間解析に依存した解析計画に対する棄却限界値

$\alpha_2^*(t)$					
t_1	t_2	$b(t_1)$	$b(1)$	$\tilde{b}(t_2)$	$\tilde{b}(1)$
0.2	0.30	2.438	2.038	3.372	2.906
	0.40	2.438	2.038	3.326	2.957
	0.50	2.438	2.038	3.233	3.002
	0.60	2.438	2.038	3.139	3.041
	0.70	2.438	2.038	3.053	3.074
	0.80	2.438	2.038	2.977	3.096
	0.90	2.438	2.038	2.908	3.094
0.3	0.40	2.312	2.062	3.197	3.005
	0.50	2.312	2.062	3.227	3.056
	0.60	2.312	2.062	3.184	3.101
	0.70	2.312	2.062	3.124	3.139
	0.80	2.312	2.062	3.063	3.168
	0.90	2.312	2.062	3.003	3.176
0.4	0.50	2.224	2.078	3.060	3.059
	0.60	2.224	2.078	3.132	3.110
	0.70	2.224	2.078	3.123	3.155
	0.80	2.224	2.078	3.089	3.192
	0.90	2.224	2.078	3.046	3.209
0.5	0.60	2.157	2.086	2.949	3.078
	0.70	2.157	2.086	3.047	3.130
	0.80	2.157	2.086	3.061	3.176
	0.90	2.157	2.086	3.046	3.204
0.6	0.70	2.103	2.084	2.858	3.064
	0.80	2.103	2.084	2.971	3.119
	0.90	2.103	2.084	3.003	3.161
0.7	0.80	2.059	2.070	2.781	3.016
	0.90	2.059	2.070	2.905	3.073
0.8	0.90	2.021	2.036	2.714	2.922

表 20: 中間解析に依存した解析計画に対する棄却限界値

$\alpha_3^*(t)$					
t_1	t_2	$b(t_1)$	$b(1)$	$\bar{b}(t_2)$	$\bar{b}(1)$
0.2	0.30	2.576	1.999	3.507	2.896
	0.40	2.576	1.999	3.444	2.934
	0.50	2.576	1.999	3.331	2.969
	0.60	2.576	1.999	3.218	3.002
	0.70	2.576	1.999	3.113	3.031
	0.80	2.576	1.999	3.019	3.054
	0.90	2.576	1.999	2.933	3.061
0.3	0.40	2.432	2.010	3.302	2.988
	0.50	2.432	2.010	3.314	3.028
	0.60	2.432	2.010	3.253	3.065
	0.70	2.432	2.010	3.175	3.099
	0.80	2.432	2.010	3.095	3.127
	0.90	2.432	2.010	3.017	3.139
0.4	0.50	2.326	2.016	3.138	3.035
	0.60	2.326	2.016	3.191	3.077
	0.70	2.326	2.016	3.163	3.116
	0.80	2.326	2.016	3.111	3.150
	0.90	2.326	2.016	3.051	3.170
0.5	0.60	2.241	2.017	3.002	3.047
	0.70	2.241	2.017	3.080	3.093
	0.80	2.241	2.017	3.076	3.134
	0.90	2.241	2.017	3.042	3.162
0.6	0.70	2.170	2.012	2.888	3.029
	0.80	2.170	2.012	2.980	3.078
	0.90	2.170	2.012	2.993	3.118
0.7	0.80	2.108	1.998	2.789	2.977
	0.90	2.108	1.998	2.891	3.031
0.8	0.90	2.054	1.970	2.703	2.884

表 21: $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ を用いたときの検出力

$\alpha = 0.025, 1 - \beta = 0.80$

$K = 2$	$\alpha_1^*(t)$	$\alpha_2^*(t)$	$\alpha_3^*(t)$
t_1	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$
0.20	(0.800, 1.000)	(0.772, 0.906)	(0.783, 0.926)
0.25	(0.800, 0.999)	(0.767, 0.875)	(0.780, 0.898)
0.30	(0.800, 0.994)	(0.763, 0.847)	(0.776, 0.871)
0.35	(0.800, 0.984)	(0.759, 0.823)	(0.774, 0.847)
0.40	(0.800, 0.966)	(0.756, 0.805)	(0.771, 0.826)
0.45	(0.799, 0.943)	(0.753, 0.792)	(0.769, 0.811)
0.50	(0.799, 0.918)	(0.751, 0.785)	(0.767, 0.801)
0.55	(0.798, 0.895)	(0.750, 0.784)	(0.766, 0.798)
0.60	(0.797, 0.876)	(0.750, 0.789)	(0.766, 0.800)
0.65	(0.795, 0.865)	(0.751, 0.800)	(0.765, 0.808)
0.70	(0.794, 0.861)	(0.753, 0.816)	(0.766, 0.822)
0.75	(0.793, 0.866)	(0.756, 0.837)	(0.768, 0.841)
0.80	(0.792, 0.880)	(0.761, 0.863)	(0.770, 0.865)
0.85	(0.791, 0.901)	(0.767, 0.892)	(0.774, 0.894)
0.90	(0.791, 0.929)	(0.775, 0.925)	(0.779, 0.926)

表 22: $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ を用いたときの検出力

$\alpha = 0.025, 1 - \beta = 0.80$

$K = 3$		$\alpha_1^*(t)$	$\alpha_2^*(t)$	$\alpha_3^*(t)$
t_1	t_2	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$
0.20	0.30	(0.800, 0.994)	(0.759, 0.855)	(0.774, 0.877)
	0.40	(0.800, 0.966)	(0.749, 0.805)	(0.767, 0.827)
	0.50	(0.799, 0.918)	(0.741, 0.772)	(0.761, 0.791)
	0.60	(0.797, 0.876)	(0.736, 0.760)	(0.757, 0.776)
	0.70	(0.794, 0.861)	(0.735, 0.770)	(0.755, 0.784)
	0.80	(0.792, 0.880)	(0.739, 0.799)	(0.757, 0.814)
	0.90	(0.791, 0.929)	(0.749, 0.846)	(0.764, 0.863)
0.30	0.40	(0.800, 0.966)	(0.751, 0.804)	(0.768, 0.825)
	0.50	(0.799, 0.917)	(0.741, 0.766)	(0.761, 0.784)
	0.60	(0.797, 0.874)	(0.735, 0.745)	(0.756, 0.761)
	0.70	(0.794, 0.858)	(0.732, 0.745)	(0.753, 0.760)
	0.80	(0.792, 0.876)	(0.733, 0.763)	(0.753, 0.780)
	0.90	(0.791, 0.924)	(0.741, 0.798)	(0.759, 0.818)
0.40	0.50	(0.799, 0.914)	(0.744, 0.771)	(0.763, 0.789)
	0.60	(0.797, 0.866)	(0.736, 0.745)	(0.757, 0.759)
	0.70	(0.794, 0.845)	(0.732, 0.736)	(0.752, 0.748)
	0.80	(0.792, 0.857)	(0.731, 0.744)	(0.751, 0.757)
	0.90	(0.791, 0.901)	(0.737, 0.767)	(0.755, 0.784)
0.50	0.60	(0.796, 0.862)	(0.741, 0.762)	(0.759, 0.773)
	0.70	(0.794, 0.830)	(0.735, 0.746)	(0.754, 0.755)
	0.80	(0.791, 0.832)	(0.732, 0.745)	(0.751, 0.754)
	0.90	(0.790, 0.864)	(0.735, 0.758)	(0.753, 0.771)
0.60	0.70	(0.794, 0.833)	(0.741, 0.775)	(0.758, 0.781)
	0.80	(0.790, 0.819)	(0.737, 0.767)	(0.754, 0.773)
	0.90	(0.789, 0.836)	(0.738, 0.772)	(0.754, 0.780)
0.70	0.80	(0.790, 0.836)	(0.746, 0.808)	(0.759, 0.811)
	0.90	(0.788, 0.837)	(0.744, 0.807)	(0.757, 0.811)
0.80	0.90	(0.788, 0.870)	(0.756, 0.860)	(0.765, 0.861)

表 23: $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ を用いたときの検出力

$\alpha = 0.025$, $1 - \beta = 0.80$

$K = 4$			$\alpha_1^*(t)$	$\alpha_2^*(t)$	$\alpha_3^*(t)$
t_1	t_2	t_3	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$
0.20	0.30	0.40	(0.800, 0.966)	(0.747, 0.806)	(0.766, 0.827)
		0.50	(0.799, 0.917)	(0.738, 0.766)	(0.759, 0.784)
		0.60	(0.797, 0.874)	(0.731, 0.746)	(0.754, 0.761)
		0.70	(0.794, 0.858)	(0.728, 0.746)	(0.751, 0.761)
		0.80	(0.792, 0.876)	(0.730, 0.767)	(0.751, 0.782)
	0.40	0.50	(0.791, 0.924)	(0.738, 0.804)	(0.756, 0.822)
		0.60	(0.798, 0.914)	(0.738, 0.766)	(0.759, 0.784)
		0.70	(0.797, 0.866)	(0.730, 0.738)	(0.753, 0.753)
		0.80	(0.794, 0.845)	(0.725, 0.729)	(0.748, 0.742)
		0.90	(0.792, 0.857)	(0.724, 0.738)	(0.747, 0.753)
	0.50	0.60	(0.791, 0.900)	(0.730, 0.765)	(0.751, 0.782)
		0.70	(0.796, 0.862)	(0.731, 0.744)	(0.753, 0.758)
		0.80	(0.794, 0.830)	(0.724, 0.727)	(0.748, 0.739)
		0.90	(0.791, 0.832)	(0.721, 0.726)	(0.745, 0.739)
		0.90	(0.790, 0.864)	(0.725, 0.742)	(0.747, 0.757)
0.60	0.70	(0.793, 0.833)	(0.727, 0.742)	(0.750, 0.754)	
	0.80	(0.790, 0.819)	(0.722, 0.734)	(0.745, 0.746)	
	0.90	(0.789, 0.836)	(0.723, 0.740)	(0.745, 0.753)	
	0.90	(0.790, 0.835)	(0.727, 0.760)	(0.749, 0.772)	
	0.90	(0.788, 0.837)	(0.725, 0.759)	(0.746, 0.771)	
0.30	0.40	0.50	(0.788, 0.869)	(0.733, 0.795)	(0.752, 0.809)
		0.50	(0.798, 0.913)	(0.740, 0.764)	(0.760, 0.782)
		0.60	(0.797, 0.865)	(0.731, 0.737)	(0.754, 0.751)
		0.70	(0.794, 0.844)	(0.726, 0.728)	(0.749, 0.741)
		0.80	(0.792, 0.857)	(0.726, 0.737)	(0.748, 0.752)
	0.50	0.60	(0.791, 0.900)	(0.731, 0.764)	(0.752, 0.781)
		0.60	(0.796, 0.861)	(0.731, 0.736)	(0.753, 0.751)
		0.70	(0.794, 0.828)	(0.724, 0.719)	(0.748, 0.731)
		0.80	(0.791, 0.830)	(0.722, 0.719)	(0.745, 0.732)
		0.90	(0.790, 0.862)	(0.725, 0.736)	(0.747, 0.750)

表 24: $\alpha_1^*(t)$, $\alpha_2^*(t)$, $\alpha_3^*(t)$ を用いたときの検出力

$\alpha = 0.025, 1 - \beta = 0.80$

$K = 4$			$\alpha_1^*(t)$	$\alpha_2^*(t)$	$\alpha_3^*(t)$	
t_1	t_2	t_3	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$	$(1 - \beta^*, E_t)$	
0.30	0.60	0.70	(0.793, 0.830)	(0.726, 0.726)	(0.748, 0.738)	
		0.80	(0.790, 0.817)	(0.721, 0.718)	(0.744, 0.730)	
		0.90	(0.789, 0.834)	(0.721, 0.725)	(0.744, 0.738)	
	0.70	0.80	(0.790, 0.832)	(0.724, 0.734)	(0.746, 0.747)	
		0.90	(0.788, 0.834)	(0.722, 0.733)	(0.744, 0.746)	
		0.80 0.90	(0.788, 0.865)	(0.728, 0.759)	(0.748, 0.774)	
0.40	0.50	0.60	(0.796, 0.857)	(0.734, 0.743)	(0.755, 0.755)	
		0.70	(0.794, 0.825)	(0.728, 0.726)	(0.750, 0.737)	
		0.80	(0.791, 0.827)	(0.725, 0.726)	(0.747, 0.737)	
		0.90	(0.790, 0.859)	(0.728, 0.742)	(0.749, 0.755)	
	0.60	0.70	(0.793, 0.822)	(0.727, 0.726)	(0.749, 0.736)	
		0.80	(0.790, 0.809)	(0.722, 0.718)	(0.745, 0.727)	
		0.90	(0.789, 0.826)	(0.723, 0.725)	(0.745, 0.736)	
		0.70 0.80	(0.790, 0.819)	(0.724, 0.724)	(0.746, 0.735)	
	0.70	0.80	(0.788, 0.820)	(0.722, 0.724)	(0.743, 0.734)	
		0.90	(0.787, 0.847)	(0.725, 0.739)	(0.746, 0.752)	
		0.50	0.60	(0.793, 0.818)	(0.732, 0.743)	(0.752, 0.751)
			0.80	(0.790, 0.804)	(0.728, 0.735)	(0.748, 0.742)
0.90	(0.789, 0.822)		(0.728, 0.742)	(0.748, 0.750)		
0.60	0.70	0.80	(0.790, 0.804)	(0.727, 0.735)	(0.747, 0.741)	
		0.90	(0.787, 0.805)	(0.725, 0.734)	(0.745, 0.741)	
	0.80 0.90	(0.787, 0.821)	(0.727, 0.740)	(0.746, 0.748)		
	0.70	0.80	(0.789, 0.806)	(0.734, 0.764)	(0.751, 0.768)	
0.90		(0.787, 0.808)	(0.732, 0.763)	(0.749, 0.768)		
0.80 0.90		(0.786, 0.809)	(0.732, 0.762)	(0.749, 0.767)		
0.70	0.80 0.90	(0.786, 0.825)	(0.741, 0.804)	(0.754, 0.806)		

表 25: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 2$	$\alpha_1^*(t)$			
t_1	ξ_β	$E(\mathbf{T})$	E_{IF}	IF
0.25	2.8016	0.999	0.999	1.0000
0.30	2.8018	0.994	0.994	1.0001
0.35	2.8022	0.984	0.984	1.0004
0.40	2.8030	0.966	0.967	1.0010
0.45	2.8045	0.943	0.945	1.0021
0.50	2.8068	0.918	0.921	1.0037
0.55	2.8098	0.894	0.900	1.0059
0.60	2.8135	0.875	0.883	1.0085
0.65	2.8178	0.863	0.873	1.0116
0.70	2.8224	0.859	0.872	1.0149
0.75	2.8269	0.864	0.880	1.0182
0.80	2.8310	0.878	0.896	1.0211
0.85	2.8336	0.899	0.920	1.0230
0.90	2.8332	0.928	0.949	1.0227
0.95	2.8267	0.962	0.979	1.0180

表 26: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 2$		$\alpha_2^*(t)$		
t_1	ξ_β	$E(\mathbf{T})$	E_{IF}	IF
0.25	2.9154	0.864	0.936	1.0829
0.30	2.9305	0.832	0.910	1.0941
0.35	2.9433	0.805	0.888	1.1038
0.40	2.9540	0.783	0.871	1.1118
0.45	2.9623	0.769	0.860	1.1180
0.50	2.9683	0.762	0.855	1.1225
0.55	2.9716	0.762	0.857	1.1250
0.60	2.9721	0.768	0.865	1.1254
0.65	2.9695	0.782	0.878	1.1234
0.70	2.9633	0.801	0.896	1.1188
0.75	2.9530	0.826	0.917	1.1110
0.80	2.9380	0.854	0.940	1.0998
0.85	2.9174	0.887	0.962	1.0844
0.90	2.8897	0.923	0.982	1.0639
0.95	2.8529	0.960	0.962	1.0379

表 27: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 2$		$\alpha_3^*(t)$		
t_1	ξ_β	$E(\mathbf{T})$	E_{IF}	IF
0.25	2.8730	0.892	0.938	1.0516
0.30	2.8836	0.862	0.913	1.0594
0.35	2.8932	0.835	0.891	1.0665
0.40	2.9016	0.813	0.872	1.0727
0.45	2.9087	0.796	0.858	1.0780
0.50	2.9145	0.786	0.851	1.0823
0.55	2.9188	0.782	0.849	1.0854
0.60	2.9214	0.785	0.854	1.0873
0.65	2.9219	0.795	0.865	1.0877
0.70	2.9200	0.811	0.881	1.0863
0.75	2.9152	0.832	0.901	1.0828
0.80	2.9069	0.858	0.924	1.0766
0.85	2.8938	0.889	0.949	1.0669
0.90	2.8747	0.924	0.972	1.0528
0.95	2.8465	0.961	0.992	1.0323

表 28: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 3$		$\alpha_1^*(t)$			
t_1	t_2	ξ_β	$E(\mathbf{T})$	E_{IF}	IF
0.20	0.25	2.8016	0.999	0.999	1.0000
	0.30	2.8018	0.994	0.994	1.0001
	0.35	2.8022	0.984	0.984	1.0004
	0.40	2.8030	0.966	0.967	1.0010
	0.45	2.8045	0.943	0.945	1.0021
	0.50	2.8068	0.918	0.921	1.0037
	0.55	2.8098	0.894	0.899	1.0059
	0.60	2.8135	0.875	0.883	1.0085
	0.65	2.8178	0.863	0.873	1.0116
	0.70	2.8224	0.859	0.872	1.0149
	0.75	2.8269	0.864	0.880	1.0182
	0.80	2.8310	0.878	0.896	1.0211
	0.85	2.8336	0.899	0.920	1.0230
	0.90	2.8332	0.928	0.949	1.0227
	0.95	2.8267	0.961	0.979	1.0180
0.30	0.35	2.8022	0.983	0.984	1.0004
	0.40	2.8031	0.965	0.967	1.0011
	0.45	2.8046	0.942	0.944	1.0021
	0.50	2.8068	0.916	0.920	1.0038
	0.55	2.8098	0.892	0.898	1.0059
	0.60	2.8136	0.873	0.880	1.0086
	0.65	2.8178	0.860	0.870	1.0116
	0.70	2.8225	0.856	0.869	1.0150
	0.75	2.8270	0.860	0.876	1.0183
	0.80	2.8311	0.873	0.892	1.0211
0.85	2.8334	0.895	0.915	1.0230	
0.90	2.8333	0.922	0.943	1.0228	
0.95	2.8268	0.956	0.973	1.0181	

表 29: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 3$		$\alpha_1^*(t)$			
t_1	t_2	ξ_β	$E(T)$	E_{IF}	IF
0.4	0.45	2.8047	0.942	0.944	1.0022
	0.50	2.8070	0.913	0.917	1.0039
	0.55	2.8101	0.887	0.892	1.0061
	0.60	2.8139	0.864	0.872	1.0088
	0.65	2.8182	0.859	0.859	1.0119
	0.70	2.8229	0.842	0.855	1.0153
	0.75	2.8276	0.844	0.860	1.0187
	0.80	2.8317	0.854	0.873	1.0216
	0.85	2.8344	0.873	0.894	1.0236
	0.90	2.8343	0.898	0.919	1.0235
0.5	0.95	2.8279	0.930	0.947	1.0189
	0.55	2.8103	0.888	0.894	1.0062
	0.60	2.8144	0.861	0.869	1.0092
	0.65	2.8190	0.840	0.850	1.0125
	0.70	2.8240	0.827	0.840	1.0161
	0.75	2.8289	0.823	0.839	1.0196
	0.80	2.8334	0.828	0.847	1.0228
	0.85	2.8365	0.840	0.861	1.0251
0.6	0.90	2.8368	0.860	0.882	1.0253
	0.95	2.8310	0.886	0.905	1.0211
	0.65	2.8192	0.850	0.861	1.0126
	0.70	2.8248	0.830	0.844	1.0167
	0.75	2.8305	0.818	0.835	1.0207
	0.80	2.8357	0.815	0.835	1.0245
	0.85	2.8396	0.820	0.842	1.0273
0.7	0.90	2.8409	0.832	0.855	1.0282
	0.95	2.8363	0.850	0.871	1.0249
	0.75	2.8301	0.843	0.860	1.0204
	0.80	2.8367	0.832	0.853	1.0252
	0.85	2.8421	0.829	0.853	1.0291
0.8	0.90	2.8450	0.833	0.859	1.0312
	0.95	2.8425	0.843	0.868	1.0294
	0.85	2.8400	0.870	0.894	1.0276
	0.90	2.8458	0.867	0.894	1.0318
	0.95	2.8466	0.869	0.898	1.0324

表 30: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 3$		$\alpha_2^*(t)$			
t_1	t_2	ξ_β	$E(T)$	E_{IF}	IF
0.20	0.25	2.9210	0.87137	0.9472	1.0871
	0.30	2.9412	0.83931	0.9250	1.1021
	0.35	2.9591	0.80785	0.9013	1.1156
	0.40	2.9750	0.77969	0.8792	1.1276
	0.45	2.9888	0.75658	0.8611	1.1381
	0.50	3.0005	0.73960	0.8483	1.1470
	0.55	3.0098	0.72931	0.8417	1.1541
	0.60	3.0165	0.72587	0.8415	1.1593
	0.65	3.0205	0.72911	0.8475	1.1624
	0.70	3.0211	0.73867	0.8590	1.1629
	0.75	3.0180	0.75403	0.8750	1.1604
	0.80	3.0102	0.77462	0.8943	1.1545
	0.85	2.9968	0.79986	0.9152	1.1442
	0.90	2.9762	0.82918	0.9358	1.1285
	0.95	2.9457	0.86213	0.9531	1.1055
0.30	0.35	2.9517	0.80722	0.8961	1.1101
	0.40	2.9701	0.77944	0.8760	1.1239
	0.45	2.9862	0.75390	0.8565	1.1362
	0.50	3.0003	0.73275	0.8404	1.1469
	0.55	3.0121	0.71714	0.8290	1.1559
	0.60	3.0216	0.70763	0.8231	1.1632
	0.65	3.0284	0.70437	0.8230	1.1685
	0.70	3.0322	0.70720	0.8284	1.1714
	0.75	3.0324	0.71581	0.8386	1.1715
	0.80	3.0282	0.72978	0.8526	1.1683
	0.85	3.0186	0.74866	0.8691	1.1609
	0.90	3.0018	0.77207	0.8864	1.1480
	0.95	2.9751	0.79976	0.9019	1.1277

表 31: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 3$		$\alpha_2^*(t)$			
t_1	t_2	ξ_β	$E(\mathbf{T})$	E_{IF}	IF
0.4	0.45	2.9738	0.76331	0.8601	1.1267
	0.50	2.9905	0.74191	0.8454	1.1394
	0.55	3.0049	0.72344	0.8322	1.1504
	0.60	3.0169	0.70948	0.8227	1.1596
	0.65	3.0265	0.70072	0.8177	1.1670
	0.70	3.0332	0.69742	0.8175	1.1722
	0.75	3.0366	0.69949	0.8217	1.1748
	0.80	3.0359	0.70671	0.8298	1.1742
	0.85	3.0300	0.71880	0.8408	1.1697
	0.90	3.0174	0.73551	0.8532	1.1600
0.5	0.95	2.9948	0.75678	0.8648	1.1427
	0.55	2.9868	0.74694	0.8490	1.1366
	0.60	3.0017	0.73212	0.8405	1.1480
	0.65	3.0141	0.72032	0.8337	1.1574
	0.70	3.0236	0.71262	0.8301	1.1648
	0.75	3.0301	0.70943	0.8299	1.1698
	0.80	3.0329	0.71082	0.8330	1.1719
0.6	0.85	3.0309	0.71669	0.8388	1.1704
	0.90	3.0224	0.72693	0.8460	1.1639
	0.95	3.0045	0.74160	0.8529	1.1501
	0.65	2.9891	0.75877	0.8638	1.1384
	0.70	3.0020	0.74967	0.8608	1.1482
	0.75	3.0117	0.74336	0.8591	1.1556
	0.80	3.0179	0.74056	0.8593	1.1604
0.7	0.85	3.0197	0.74150	0.8615	1.1618
	0.90	3.0157	0.74627	0.8647	1.1587
	0.95	3.0028	0.75498	0.8673	1.1488
	0.75	2.9783	0.79541	0.8989	1.1302
	0.80	2.9884	0.79075	0.8997	1.1378
0.8	0.85	2.9942	0.78856	0.9007	1.1422
	0.90	2.9947	0.78931	0.9019	1.1426
	0.95	2.9873	0.79327	0.9019	1.1370
	0.85	2.9501	0.85164	0.9443	1.1088
0.9	0.90	2.9556	0.85017	0.9462	1.1129
	0.95	2.9540	0.85088	0.9460	1.1118

表 32: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 3$		$\alpha_3^*(t)$			
t_1	t_2	ξ_β	$E(T)$	E_{IF}	IF
0.20	0.25	2.8767	0.89785	0.9467	1.0544
	0.30	2.8906	0.86841	0.9245	1.0646
	0.35	2.9034	0.83859	0.9007	1.0740
	0.40	2.9150	0.81106	0.8781	1.0826
	0.45	2.9255	0.78771	0.8589	1.0904
	0.50	2.9346	0.76986	0.8447	1.0972
	0.55	2.9424	0.75832	0.8365	1.1031
	0.60	2.9486	0.75345	0.8346	1.1077
	0.65	2.9529	0.75529	0.8391	1.1110
	0.70	2.9550	0.76362	0.8496	1.1125
	0.75	2.9544	0.77801	0.8652	1.1121
	0.80	2.9503	0.79794	0.8849	1.1090
	0.85	2.9418	0.82283	0.9073	1.1026
	0.90	2.9271	0.85209	0.9302	1.0916
	0.95	2.9032	0.88522	0.9506	1.0738
0.30	0.35	2.8988	0.83785	0.8970	1.0706
	0.40	2.9122	0.81022	0.8755	1.0806
	0.45	2.9243	0.78438	0.8546	1.0896
	0.50	2.9352	0.76260	0.8371	1.0976
	0.55	2.9447	0.74620	0.8244	1.1048
	0.60	2.9526	0.73589	0.8174	1.1107
	0.65	2.9589	0.73197	0.8165	1.1154
	0.70	2.9630	0.73439	0.8214	1.1185
	0.75	2.9645	0.74287	0.8318	1.1197
	0.80	2.9628	0.75701	0.8466	1.1184
0.85	2.9567	0.77634	0.8647	1.1138	
0.90	2.4324	0.80042	0.8842	1.1047	
0.95	2.9233	0.82893	0.9025	1.0888	

表 33: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 3$		$\alpha_3^*(t)$			
t_1	t_2	ξ_β	$E(T)$	E_{IF}	IF
0.4	0.45	2.9166	0.79178	0.8581	1.0838
	0.50	2.9295	0.76907	0.8409	1.0934
	0.55	2.9408	0.74939	0.8257	1.1019
	0.60	2.9506	0.73446	0.8147	1.1092
	0.65	2.9588	0.72507	0.8087	1.1154
	0.70	2.9650	0.72151	0.8081	1.1200
	0.75	2.9687	0.72374	0.8127	1.1229
	0.80	2.9694	0.73151	0.8218	1.1234
	0.85	2.9659	0.74448	0.8344	1.1208
	0.90	2.9567	0.76233	0.8491	1.1138
0.5	0.95	2.9385	0.78493	0.8635	1.1001
	0.55	2.9293	0.76926	0.8410	1.0933
	0.60	2.9415	0.75268	0.8297	1.1024
	0.65	2.9518	0.73960	0.8211	1.1101
	0.70	2.9602	0.73120	0.8163	1.1164
	0.75	2.9662	0.72787	0.8159	1.1210
	0.80	2.9694	0.72964	0.8197	1.1234
	0.85	2.9688	0.73636	0.8269	1.1229
0.6	0.90	2.9626	0.74782	0.8363	1.1183
	0.95	2.9479	0.76400	0.8459	1.1072
	0.65	2.9358	0.77363	0.8495	1.0981
	0.70	2.9469	0.76301	0.8442	1.1064
	0.75	2.9556	0.75584	0.8412	1.1130
	0.80	2.9615	0.75285	0.8413	1.1174
	0.85	2.9639	0.75420	0.8441	1.1192
0.7	0.90	2.9611	0.75988	0.8489	1.1171
	0.95	2.9503	0.76993	0.8538	1.1090
	0.75	2.9336	0.80354	0.8810	1.0965
	0.80	2.9430	0.79793	0.8805	1.1035
	0.85	2.9487	0.79546	0.8812	1.1078
	0.90	2.9498	0.79655	0.8830	1.1086
0.8	0.95	2.9434	0.80139	0.8846	1.1038
	0.85	2.9186	0.85491	0.9278	1.0853
	0.90	2.9243	0.85312	0.9295	1.0895
	0.95	2.9232	0.85410	0.9299	1.0887

表 34: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 4$			$\alpha_1^*(t)$				
t_1	t_2	t_3	ξ_β	$E(\mathbf{T})$	E_{IF}	IF	
0.20	0.30	0.40	2.803	0.965	0.967	1.0011	
		0.50	2.807	0.916	0.920	1.0038	
		0.60	2.814	0.873	0.880	1.0086	
		0.70	2.822	0.856	0.867	1.0150	
		0.80	2.831	0.873	0.892	1.0211	
		0.90	2.833	0.922	0.943	1.0228	
	0.40	0.50	0.60	2.807	0.913	0.917	1.0039
			0.70	2.814	0.864	0.872	1.0088
			0.80	2.823	0.842	0.855	1.0153
			0.90	2.832	0.854	0.873	1.0216
			0.90	2.834	0.898	0.919	1.0235
	0.50	0.60	0.70	2.814	0.861	0.869	1.0092
			0.80	2.824	0.827	0.840	1.0161
			0.90	2.833	0.828	0.847	1.0228
			0.90	2.837	0.860	0.882	1.0253
0.60	0.70	0.80	2.825	0.830	0.844	1.0167	
		0.90	2.836	0.815	0.835	1.0245	
		0.90	2.841	0.832	0.855	1.0282	
0.70	0.70	0.80	2.837	0.832	0.853	1.0252	
		0.80	2.845	0.833	0.859	1.0312	
0.80	0.90	2.846	0.867	0.894	1.0318		

表 35: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 4$			$\alpha_1^*(t)$				
t_1	t_2	t_3	ξ_β	$E(T)$	E_{IF}	IF	
0.3	0.40	0.50	2.807	0.913	0.916	1.0039	
		0.60	2.814	0.864	0.871	1.0088	
		0.70	2.823	0.841	0.854	1.0153	
		0.80	2.832	0.854	0.872	1.0216	
		0.90	2.834	0.898	0.919	1.0235	
	0.50	0.60	2.814	0.859	0.867	1.0092	
		0.70	2.824	0.825	0.839	1.0161	
		0.80	2.833	0.826	0.845	1.0229	
		0.90	2.837	0.858	0.880	1.0253	
		0.60	0.70	2.825	0.827	0.841	1.0167
			0.80	2.836	0.813	0.833	1.0245
			0.90	2.841	0.829	0.853	1.0283
0.70	0.70	2.837	0.828	0.849	1.0252		
		0.80	2.845	0.829	0.855	1.0313	
	0.80	0.90	2.846	0.862	0.890	1.0319	
		0.40	0.50	0.60	2.815	0.855	0.863
0.70	2.824			0.822	0.835	1.0162	
0.80	2.834			0.823	0.842	1.0230	
0.90	2.837			0.855	0.877	1.0254	
0.60	0.70		2.825	0.819	0.833	1.0169	
	0.80		2.836	0.804	0.824	1.0247	
	0.90		2.841	0.821	0.844	1.0285	
0.70	0.70		2.837	0.814	0.835	1.0256	
			0.80	2.846	0.815	0.841	1.0316
	0.80		0.90	2.847	0.843	0.870	1.0323
0.50	0.60		0.70	2.826	0.814	0.828	1.0172
			0.80	2.836	0.799	0.820	1.0250
		0.90	2.842	0.817	0.840	1.0288	
	0.70	0.80	2.838	0.799	0.820	1.0262	
		0.90	2.847	0.799	0.825	1.0323	
0.80	0.90	2.848	0.815	0.843	1.0335		
	0.60	0.70	0.80	2.839	0.801	0.823	1.0267
			0.90	2.847	0.802	0.827	1.0329
0.80		0.90	2.851	0.820	0.849	1.0356	

表 36: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 4$			$\alpha_2^*(t)$					
t_1	t_2	t_3	ξ_β	$E(\mathbf{T})$	E_{IF}	IF		
0.20	0.30	0.40	2.980	0.780	0.882	1.1315		
		0.50	3.011	0.731	0.844	1.1547		
		0.60	3.032	0.706	0.827	1.1714		
		0.70	3.043	0.707	0.834	1.1799		
		0.80	3.039	0.732	0.861	1.1769		
		0.90	3.013	0.777	0.898	1.1566		
	0.40	0.50	0.50	3.011	0.730	0.844	1.1552	
			0.60	3.038	0.695	0.817	1.1758	
			0.70	3.055	0.683	0.813	1.1888	
		0.60	0.80	0.80	3.058	0.695	0.828	1.1912
				0.90	3.039	0.727	0.856	1.1769
				0.90	3.034	0.703	0.825	1.1727
			0.70	0.70	3.056	0.682	0.812	1.1901
				0.80	3.066	0.681	0.815	1.1978
				0.90	3.056	0.700	0.833	1.1898
0.80	0.90	0.70	3.047	0.702	0.830	1.1827		
		0.80	3.063	0.692	0.827	1.1957		
		0.90	3.062	0.699	0.835	1.1943		
0.70	0.80	0.70	3.047	0.725	0.858	1.1830		
		0.80	3.054	0.723	0.860	1.1885		
0.80	0.90	0.90	3.029	0.768	0.898	1.1690		

表 37: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05, 1 - \beta = 0.80$)

$K = 4$			$\alpha_2^*(t)$				
t_1	t_2	t_3	ξ_β	$E(T)$	E_{IF}	IF	
0.3	0.40	0.50	3.006	0.730	0.841	1.1511	
		0.60	3.033	0.696	0.815	1.1716	
		0.70	3.049	0.684	0.810	1.1846	
		0.80	3.052	0.696	0.826	1.1871	
		0.90	3.034	0.728	0.853	1.1729	
	0.50	0.60	3.033	0.695	0.815	1.1722	
		0.70	3.056	0.673	0.801	1.1896	
		0.80	3.066	0.673	0.805	1.1974	
		0.90	3.056	0.692	0.823	1.1896	
	0.60	0.70	3.052	0.682	0.809	1.1864	
		0.80	3.068	0.671	0.806	1.1995	
		0.90	3.067	0.679	0.814	1.1983	
	0.70	0.70	3.058	0.692	0.825	1.1916	
		0.80	3.065	0.690	0.827	1.1973	
		0.90	3.047	0.723	0.855	1.1831	
0.40	0.50	0.60	3.023	0.706	0.822	1.1644	
		0.70	3.046	0.685	0.809	1.1817	
		0.80	3.055	0.684	0.813	1.1894	
		0.90	3.045	0.703	0.831	1.1816	
		0.60	0.70	3.046	0.684	0.809	1.1824
	0.60	0.80	3.063	0.673	0.805	1.1955	
		0.90	3.062	0.681	0.814	1.1944	
		0.70	0.70	3.059	0.682	0.813	1.1921
	0.70	0.80	3.066	0.680	0.815	1.1979	
		0.80	0.90	3.055	0.699	0.831	1.1890
		0.60	0.70	3.031	0.708	0.829	1.1703
	0.50	0.60	0.80	3.047	0.698	0.826	1.1832
			0.90	3.046	0.706	0.834	1.1821
			0.70	0.80	3.049	0.697	0.826
	0.70	0.80	0.90	3.056	0.696	0.828	1.1901
0.80			0.90	3.052	0.703	0.835	1.1865
0.60			0.70	0.80	3.027	0.736	0.859
0.60	0.70	0.90	3.034	0.735	0.862	1.1727	
		0.70	0.80	0.90	3.006	0.785	0.904

表 38: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 4$			$\alpha_3^*(t)$			
t_1	t_2	t_3	ξ_β	$E(\mathbf{T})$	E_{IF}	IF
0.20	0.30	0.40	2.919	0.811	0.880	1.0854
		0.50	2.942	0.761	0.840	1.1026
		0.60	2.959	0.735	0.820	1.1158
		0.70	2.970	0.734	0.825	1.1238
		0.80	2.970	0.759	0.853	1.1238
		0.90	2.952	0.804	0.893	1.1101
		0.40	0.50	2.942	0.761	0.839
0.40	0.60	0.60	2.964	0.724	0.810	1.1191
		0.70	2.978	0.712	0.804	1.1301
		0.80	2.983	0.723	0.820	1.1337
		0.90	2.971	0.757	0.851	1.1242
		0.50	0.60	2.961	0.731	0.817
0.50	0.70	0.70	2.980	0.708	0.801	1.1315
		0.80	2.990	0.707	0.806	1.1389
		0.90	2.983	0.728	0.826	1.1339
		0.60	0.70	2.974	0.727	0.820
0.60	0.80	0.80	2.989	0.716	0.815	1.1383
		0.90	2.989	0.725	0.825	1.1383
		0.70	0.70	2.978	0.748	0.846
0.70	0.80	0.80	2.985	0.747	0.848	1.1356
		0.80	0.90	2.968	0.791	0.888

表 39: ξ_β と E_{IF} 値 (両側 $\alpha = 0.05$, $1 - \beta = 0.80$)

$K = 4$			$\alpha_3^*(t)$				
t_1	t_2	t_3	ξ_β	$E(T)$	E_{IF}	IF	
0.3	0.40	0.50	2.939	0.760	0.836	1.1008	
		0.60	2.961	0.723	0.808	1.1168	
		0.70	2.975	0.711	0.802	1.1278	
		0.80	2.980	0.723	0.818	1.1314	
		0.90	2.968	0.756	0.849	1.1220	
	0.50	0.60	2.961	0.723	0.808	1.1174	
		0.70	2.980	0.699	0.792	1.1317	
		0.80	2.990	0.699	0.796	1.1391	
		0.90	2.984	0.720	0.817	1.1343	
		0.60	0.70	2.978	0.708	0.800	1.1298
	0.60	0.80	2.993	0.697	0.795	1.1412	
		0.90	2.993	0.706	0.806	1.1414	
		0.70	0.70	2.986	0.718	0.815	1.1360
		0.80	2.994	0.716	0.818	1.1417	
		0.80	0.90	2.981	0.749	0.848	1.1321
0.40	0.50	0.60	2.956	0.730	0.812	1.1129	
		0.70	2.974	0.707	0.797	1.1272	
		0.80	2.984	0.707	0.802	1.1346	
		0.90	2.978	0.727	0.822	1.1297	
	0.60	0.70	2.975	0.706	0.797	1.1280	
		0.80	2.991	0.695	0.792	1.1395	
		0.90	2.991	0.704	0.803	1.1397	
	0.70	0.70	2.988	0.704	0.801	1.1373	
		0.80	2.995	0.703	0.803	1.1431	
0.80	0.90	2.987	0.723	0.833	1.1371		
0.50	0.60	0.70	2.966	0.726	0.813	1.1209	
		0.80	2.981	0.715	0.810	1.1323	
		0.90	2.981	0.724	0.820	1.1325	
	0.70	0.80	2.983	0.714	0.809	1.1334	
		0.90	2.990	0.713	0.812	1.1392	
	0.80	0.90	2.987	0.721	0.820	1.1370	
0.60	0.70	0.80	2.969	0.747	0.839	1.1231	
		0.90	2.977	0.746	0.842	1.1288	
0.70	0.80	0.90	2.960	0.791	0.883	1.1163	

表 40: One-stage デザイン: $P_{CS} \geq 0.80$ となる被験者数 n と δ

$t = 2$		P_{CS}					
Δ	π_c	n	δ	(δ/n)	$j = 0$	$j = 1$	$j = 2$
0.10	0.2	186	10	(.054)	0.8532	0.8010	0.8209
	0.3	259	14	(.054)	0.8315	0.8002	0.8369
	0.4	305	16	(.052)	0.8334	0.8003	0.8327
	0.5	325	17	(.052)	0.8290	0.8004	0.8359
	0.6	318	16	(.050)	0.8383	0.8001	0.8246
	0.7	285	15	(.053)	0.8183	0.8001	0.8431
	0.8	226	12	(.053)	0.8105	0.8006	0.8493
	0.9	139	7	(.050)	0.8208	0.8011	0.8341
	0.15	0.2	76	7	(.092)	0.8392	0.8001
0.3		110	9	(.082)	0.8465	0.8006	0.8257
0.4		133	10	(.075)	0.8609	0.8003	0.8070
0.5		143	11	(.077)	0.8444	0.8007	0.8215
0.6		141	11	(.078)	0.8325	0.8002	0.8304
0.7		128	10	(.078)	0.8266	0.8008	0.8350
0.8		103	8	(.078)	0.8226	0.8016	0.8368
0.9		65	5	(.077)	0.8144	0.8015	0.8377
0.20		0.3	59	7	(.119)	0.8385	0.8015
	0.4	72	8	(.111)	0.8379	0.8001	0.8319
	0.5	80	8	(.100)	0.8634	0.8016	0.8053
	0.6	80	8	(.100)	0.8548	0.8020	0.8117
	0.7	73	8	(.110)	0.8143	0.8027	0.8499
	0.8	59	6	(.102)	0.8326	0.8014	0.8253
	0.9	38	4	(.105)	0.8071	0.8024	0.8433

表 41: One-stage デザイン: $P_{CS} \geq 0.80$ となる被験者数 n と δ

$t = 3$		P_{CS}						
Δ	π_c	n	δ	(δ/n)	$j = 0$	$j = 1$	$j = 2$	$j = 3$
0.10	0.2	237	13	(.055)	0.8456	0.8054	0.8003	0.8325
	0.3	333	18	(.054)	0.8304	0.8011	0.8042	0.8428
	0.4	394	21	(.053)	0.8265	0.8003	0.8055	0.8448
	0.5	420	22	(.052)	0.8275	0.8008	0.8051	0.8423
	0.6	409	21	(.051)	0.8295	0.8002	0.8021	0.8360
	0.7	370	19	(.051)	0.8260	0.8008	0.8053	0.8409
	0.8	294	15	(.051)	0.8225	0.8004	0.8066	0.8426
	0.9	182	9	(.049)	0.8222	0.8009	0.8056	0.8371
	0.15	0.2	97	9	(.093)	0.8348	0.8038	0.8058
0.3		142	12	(.085)	0.8331	0.8030	0.8056	0.8463
0.4		171	14	(.082)	0.8289	0.8015	0.8057	0.8451
0.5		186	15	(.081)	0.8261	0.8020	0.8081	0.8476
0.6		181	14	(.077)	0.8327	0.8007	0.8002	0.8316
0.7		167	13	(.078)	0.8258	0.8019	0.8071	0.8426
0.8		133	10	(.075)	0.8291	0.8009	0.8016	0.8309
0.9		89	7	(.079)	0.8016	0.8023	0.8239	0.8708
0.20		0.3	75	9	(.120)	0.8295	0.8008	0.8050
	0.4	93	10	(.108)	0.8505	0.8085	0.8005	0.8267
	0.5	102	11	(.108)	0.8337	0.8032	0.8042	0.8388
	0.6	103	11	(.107)	0.8281	0.8029	0.8073	0.8437
	0.7	95	10	(.105)	0.8255	0.8033	0.8092	0.8452
	0.8	77	8	(.104)	0.8184	0.8003	0.8083	0.8449
	0.9	51	5	(.098)	0.8253	0.8056	0.8080	0.8337

表 42: One-stage デザイン: $P_{CS} \geq 0.80$ となる被験者数 n と δ

$t = 4$		P_{CS}							
π_c	Δ	n	δ	(δ/n)	$j = 0$	$j = 1$	$j = 2$	$j = 3$	$j = 4$
0.10	0.2	277	15	(.054)	0.8536	0.8156	0.8007	0.8041	0.8316
	0.3	390	20	(.051)	0.8554	0.8165	0.8005	0.8021	0.8245
	0.4	461	24	(.052)	0.8379	0.8088	0.8003	0.8091	0.8393
	0.5	492	25	(.051)	0.8419	0.8107	0.8001	0.8067	0.8335
	0.6	484	24	(.050)	0.8474	0.8133	0.8003	0.8040	0.8268
	0.7	434	22	(.051)	0.8319	0.8064	0.8000	0.8102	0.8402
	0.8	346	17	(.049)	0.8389	0.8102	0.8007	0.8072	0.8317
	0.9	218	11	(.050)	0.8140	0.8022	0.8052	0.8226	0.8574
	0.15	0.2	111	10	(.090)	0.8487	0.8145	0.8019	0.8084
0.3		166	14	(.084)	0.8344	0.8080	0.8018	0.8138	0.8517
0.4		200	16	(.080)	0.8399	0.8103	0.8012	0.8097	0.8405
0.5		216	17	(.079)	0.8357	0.8080	0.8003	0.8099	0.8408
0.6		215	16	(.074)	0.8522	0.8154	0.8001	0.8016	0.8217
0.7		195	15	(.077)	0.8293	0.8054	0.8002	0.8113	0.8421
0.8		158	12	(.076)	0.8253	0.8047	0.8013	0.8134	0.8440
0.9		104	8	(.077)	0.8041	0.8001	0.8087	0.8306	0.8693
0.20		0.3	87	10	(.115)	0.8504	0.8154	0.8021	0.8075
	0.4	109	12	(.110)	0.8411	0.8115	0.8023	0.8108	0.8433
	0.5	120	13	(.108)	0.8320	0.8073	0.8019	0.8137	0.8480
	0.6	121	13	(.107)	0.8243	0.8044	0.8025	0.8171	0.8534
	0.7	111	11	(.099)	0.8495	0.8153	0.8011	0.8031	0.8229
	0.8	91	9	(.099)	0.8380	0.8112	0.8019	0.8078	0.8309
	0.9	60	6	(.100)	0.8104	0.8008	0.8039	0.8201	0.8520

表 43: Two-stage デザイン: $P_{CS} \geq 0.80$ となる被験者数 n と δ

$t = 2$

Δ	π_c	n	n_1	n_2	δ_1 (δ_1/n_1)	δ_2 (δ_2/n)	$E(N)$	$E(N)/N_1$
0.20	0.3	59	47	12	9 (.191)	7 (.119)	169.54	.958
	0.4	72	53	19	13 (.245)	8 (.111)	209.40	.969
	0.5	80	52	28	11 (.212)	8 (.100)	225.61	.940
	0.6	80	42	38	10 (.238)	8 (.100)	223.93	.933
	0.7	73	44	29	9 (.205)	8 (.110)	202.96	.927
	0.8	59	31	28	8 (.258)	6 (.102)	163.30	.923
	0.9	38	21	17	4 (.190)	4 (.105)	102.78	.902

表 44: Two-stage design デザイン: $P_{CS} \geq 0.80$ となる被験者数 n と δ

$t = 3$

Δ	π_c	n	n_1	n_2	δ_1 (δ_1/n_1)	δ_2 (δ_2/n)	$E(N)$	$E(N)/N_1$
0.20	0.3	75	61	14	11 (.180)	9 (.120)	285.87	.953
	0.4	93	67	26	15 (.224)	10 (.108)	355.53	.956
	0.5	102	65	37	13 (.200)	11 (.108)	377.33	.925
	0.6	103	49	54	12 (.182)	11 (.107)	374.72	.910
	0.7	95	53	42	10 (.189)	10 (.105)	341.43	.899
	0.8	77	43	34	8 (.186)	8 (.104)	275.99	.896
	0.9	51	27	24	5 (.185)	5 (.098)	180.82	.886

表 45: Two-stage デザイン: $P_{CS} \geq 0.80$ となる被験者数 n と δ

$t = 4$

Δ	π_c	n	n_1	n_2	δ_1 (δ_1/n_1)	δ_2 (δ_2/n)	$E(N)$	$E(N)/N_1$
0.20	0.3	87	58	29	12 (.207)	10 (.115)	404.92	.931
	0.4	109	71	38	14 (.197)	12 (.110)	502.14	.921
	0.5	120	76	44	15 (.197)	13 (.108)	550.87	.918
	0.6	121	77	44	15 (.195)	13 (.107)	554.90	.917
	0.7	111	56	55	13 (.232)	11 (.099)	510.34	.920
	0.8	91	51	40	10 (.200)	9 (.099)	409.20	.899
	0.9	60	33	27	6 (.182)	6 (.100)	264.99	.883

表 46: Weighted assignment one-stage デザイン ($n_c/n = 1.5$): $P_{CS} \geq 0.80$ となる
被験者数 n と δ および one-stage デザインとの被験者数の比較

$t = 2$								
Δ	π_c	n_c	n	δ	(δ/n)	N_{w1}	N_1	N_{w1}/N_1
0.10	0.2	226	151	8	(.053)	528	558	.946
	0.3	318	212	11	(.052)	742	777	.955
	0.4	378	252	13	(.052)	882	915	.964
	0.5	405	270	14	(.052)	945	975	.969
	0.6	397	265	13	(.049)	927	954	.972
	0.7	357	238	12	(.050)	833	855	.974
	0.8	286	191	9	(.047)	668	678	.985
	0.9	181	121	5	(.041)	423	417	1.014
	0.15	0.2	90	60	5	(.083)	210	228
0.3		135	90	7	(.078)	315	330	.955
0.4		163	109	8	(.073)	381	399	.955
0.5		177	118	9	(.076)	413	429	.963
0.6		177	118	9	(.076)	413	423	.976
0.7		160	107	8	(.075)	374	384	.974
0.8		130	87	6	(.069)	304	309	.984
0.9		87	58	4	(.069)	203	195	1.041
0.20		0.3	70	47	5	(.106)	164	177
	0.4	88	59	6	(.102)	206	216	.954
	0.5	99	66	7	(.106)	231	240	.963
	0.6	99	66	7	(.106)	231	240	.963
	0.7	91	61	6	(.098)	213	219	.973
	0.8	75	50	5	(.100)	175	177	.989
	0.9	51	34	3	(.088)	119	114	1.044

表 47: Weighted assignment one-stage デザイン ($n_c/n = 1.5$): $P_{CS} \geq 0.80$ となる
被験者数 n と δ および one-stage デザインとの被験者数の比較

$t = 3$								
Δ	π_c	n_c	n	δ	(δ/n)	N_{w1}	N_1	N_{w1}/N_1
0.10	0.2	295	197	11	(.056)	886	948	.935
	0.3	420	280	15	(.054)	1260	1332	.946
	0.4	492	328	17	(.052)	1476	1576	.937
	0.5	528	352	18	(.051)	1584	1680	.943
	0.6	520	347	17	(.049)	1561	1636	.954
	0.7	472	315	15	(.048)	1417	1480	.957
	0.8	376	251	12	(.048)	1129	1176	.960
	0.9	235	157	7	(.045)	706	728	.970
	0.15	0.2	117	78	7	(.090)	351	388
0.3		178	119	10	(.084)	535	568	.942
0.4		211	141	11	(.078)	634	684	.927
0.5		232	155	12	(.077)	697	744	.937
0.6		234	156	12	(.077)	702	724	.970
0.7		214	143	10	(.070)	643	668	.963
0.8		172	115	8	(.070)	517	532	.972
0.9		115	77	5	(.065)	346	356	.972
0.20		0.3	91	61	7	(.115)	274	300
	0.4	115	77	8	(.104)	346	372	.930
	0.5	129	86	9	(.105)	387	408	.949
	0.6	132	88	9	(.102)	396	412	.961
	0.7	120	80	8	(.100)	360	380	.947
	0.8	100	67	6	(.090)	301	308	.977
	0.9	69	46	4	(.087)	207	204	1.015

表 48: Weighted assignment one-stage デザイン ($n_c/n = 1.5$): $P_{CS} \geq 0.80$ となる
被験者数 n と δ および one-stage デザインとの被験者数の比較

$t = 4$								
Δ	π_c	n_c	n	δ	(δ/n)	N_{w1}	N_1	N_{w1}/N_1
0.10	0.2	348	232	13	(.056)	1276	1385	.921
	0.3	490	327	17	(.052)	1798	1950	.922
	0.4	585	390	20	(.051)	2145	2305	.931
	0.5	627	418	21	(.050)	2299	2460	.935
	0.6	618	412	21	(.051)	2266	2420	.936
	0.7	556	371	18	(.049)	2040	2170	.940
	0.8	445	297	14	(.047)	1633	1730	.944
	0.9	282	188	9	(.048)	1034	1090	.949
	0.15	0.2	136	91	8	(.088)	500	555
0.3		205	137	11	(.080)	753	830	.907
0.4		252	168	13	(.077)	924	1000	.924
0.5		274	183	14	(.077)	1006	1080	.931
0.6		274	183	14	(.077)	1006	1075	.936
0.7		250	167	12	(.072)	918	975	.942
0.8		204	136	10	(.074)	748	790	.947
0.9		135	90	6	(.067)	495	520	.952
0.20		0.3	108	72	8	(.111)	396	435
	0.4	138	92	10	(.109)	506	545	.928
	0.5	151	101	10	(.099)	555	600	.925
	0.6	154	103	10	(.097)	566	605	.936
	0.7	142	95	9	(.095)	522	555	.941
	0.8	120	80	8	(.100)	440	455	.967
	0.9	81	54	5	(.093)	297	300	.990

表 49: Weighted assignment two-stage デザイン ($n_c/n = 1.5$): $P_{CS} \geq 0.80$ となる
被験者数 n と δ および one-stage デザインとの被験者数の比較

$\Delta = 0.20$										
t	π_c	n_c	n	n_{c_1}	n_1	$\delta_1(\delta_1/n_1)$	$\delta_2(\delta_2/n)$	$E_w(N)$	E_w/N_{w1}	
2	0.3	70	47	48	32	7(.219)	5(.106)	157.11	.958	
	0.4	88	59	60	40	9(.225)	6(.102)	197.92	.961	
	0.5	99	66	57	38	8(.211)	7(.106)	216.89	.939	
	0.6	99	66	63	42	9(.214)	7(.106)	219.56	.950	
	0.7	91	61	45	30	7(.233)	6(.098)	199.85	.938	
	0.8	75	50	45	30	6(.200)	5(.100)	164.17	.938	
	0.9	51	34	30	20	4(.200)	3(.088)	111.45	.937	
	3	0.3	91	61	65	46	9(.209)	7(.115)	261.38	.954
		0.4	115	77	83	55	12(.218)	8(.104)	332.43	.961
0.5		129	86	74	49	10(.204)	9(.105)	359.34	.929	
0.6		132	88	81	54	10(.185)	9(.102)	365.30	.922	
0.7		120	80	60	41	9(.225)	8(.100)	334.33	.929	
0.8		100	67	57	38	8(.211)	6(.090)	280.03	.930	
0.9		69	46	39	26	5(.192)	4(.087)	190.21	.919	
4		0.3	108	72	68	45	9(.200)	8(.111)	368.72	.931
		0.4	138	92	87	58	11(.190)	10(.109)	466.16	.921
	0.5	151	101	108	72	15(.208)	10(.099)	527.56	.951	
	0.6	154	103	96	64	12(.188)	10(.097)	520.17	.919	
	0.7	142	95	83	55	11(.200)	9(.095)	480.89	.921	
	0.8	120	80	69	46	8(.174)	8(.100)	395.98	.900	
	0.9	81	54	45	30	5(.167)	5(.093)	264.88	.892	

参考文献

1. Japanese Ministry of Health and Welfare. *Statistical principles for clinical trials*, 1998.
2. Lan, K. K. G. & DeMets, D. L. 'Discrete sequential boundaries for clinical trials', *Biometrika*, **70**, 659-663 (1983).
3. Lan, K. K. G. & DeMets, D. L. 'Changing frequency of interim analysis in sequential monitoring', *Biometrics*, **45**, 1017-1020 (1989).
4. Proschan, M. A. , Follmann, D. A. & Waclawiw, M. A. 'Effects of assumption violations on type I error rate in group sequential monitoring', *Biometrics*, **48**, 1131-1143 (1992).
5. Bechhofer, R. E. 'A single-sample multiple decision procedure for ranking means of normal populations with known variances', *Annals of Mathematical Statistics*, **25**, 16-39 (1954).
6. Simon, R., Wittes, R. E. & Ellenberg, S. S. 'Randomized Phase II clinical trials', *Cancer Treatment Reports*, **69**, 1375-1381 (1985).
7. Thall, P. F., Simon, R. & Ellenberg, S. S. 'A Two-stage design for choosing among several experimental treatments and a control in clinical trials', *Biometrics*, **45**, 537-547 (1989).

8. Lan, K. K. G. & Zucker, D. M. 'Sequential monitoring of clinical trials : The role of information and Brownian motion', *Statistics in Medicine*, **12**, 753-765 (1993).
9. Lan, K. K. G. & Wittes, J. 'The B-value : a tool for monitoring data', *Biometrics*, **44**, 579-585 (1988).
10. Lan, K. K. G., Reboussin, D. M. & DeMets, D. L. 'Information and information fractions for design and sequential monitoring of clinical trials', *Communication in Statistics, Theory and Methods*, **23**, 403-420 (1994).
11. Armitage, P. *Sequential Medical Trials*. Oxford, Blackwell 1975.
12. Wald, A. *Sequential Analysis*. Wiley, New York, 1947.
13. Pocock, S. J. 'Group sequential methods in the design and analysis of clinical trials', *Biometrika*, **64**, 191-199 (1977).
14. O'Brien, P. C. & Fleming, T. R. 'A multiple testing procedure for clinical trials', *Biometrics*, **35**, 549-556 (1979).
15. Peace, K. E. *Biopharmaceutical Sequential Statistical Application*. Marcel Dekker, 1992.
16. DeMets, D. L. & Lan, K. K. G. 'Interim analysis : The alpha spending function approach', *Statistics in Medicine*, **13**, 1341-1352 (1994).

17. Armitage, P., McPherson, C. K. & Rowe, B. C. 'Repeated significance tests on accumulating data', *Journal of the Royal Statistical Society , Series A*, **132**, 15-25 (1969).
18. Geller, N. L. 'Discussion of interim analysis: The alpha spending approach', *Statistics in Medicine*, **13**, 1353-1356 (1994).
19. Hwang, I. K. , Shih, W. J & De Cani, J. S. 'Group sequential designs using a family of type I error probability spending functions', *Statistics in Medicine*, **9**, 1439-1445 (1990).
20. Jennison, C & Turnbull, B. W. 'Exact calculations for sequential t , χ^2 and F tests' *Biometrika*, **78**, 133-141 (1991).
21. Bickel, P. J. & Doksum, K. A. *Mathematical statistics*, Holden-day, 1977.
22. Kim, K. & DeMets, D. L. 'Design and analysis of group sequential tests based on the type I error spending rate function', *Biometrika*, **74**, 149-154 (1987).
23. Li, Z. & Geller, N. L. 'On the choice of times for data analysis in group sequential clinical trials', *Biometrics*, **47**, 745-750 (1991).
24. McPherson, K. 'On choosing the number of interim analyses in clinical trials', *Statistics in Medicine*, **1**, 25-36 (1982).
25. Brittain, E. H. & Bailey, R. K. 'Optimization of multistage testing times and critical values in clinical trials', *Biometrics*, **49**, 763-772 (1993).

26. Slud, E. V., & Wei, L. J. 'Two-sample repeated significance tests based on the modified Wilcoxon statistics', *Journal of American Statistics Association*, **77**, 862-868 (1982).
27. Bechhofer, R.E., Santner, T. J. & Goldsman, D. M. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*, Wiley, New York, 1995.
28. Gibbson, J. D., Olkin, I. & Sobel, M. *Selecting and Ordering Populations: A New Statistical Methodology*, Wiley, New York, 1977.
29. Ross, S. *Stochastic process* Second Edition, Wiley, New York, 1996.
30. Breiman, L. *Probability* (Classic Edition), SIAM, 1992.
31. Kim, K. & DeMets, D. L. 'Sample size determination for group sequential clinical trials with immediate response' *Statistics in Medicine*, **11**, 1391-1399 (1992).
32. Wang, S. K. & Tsiatis, A. A. 'Approximately optimal one-parameter boundaries for group sequential trials' *Biometrics*, **43**, 193-199 (1987).
33. Scharfstein, D. O. & Tsiatis, A. A. 'The use of simulation and bootstrap in information-based group sequential studies' *Statistics in Medicine*, **17**, 75-87 (1998).

34. Gupta, S. S. & Sobel, M. 'On selecting a subset which contains all populations better than a standard', *Annals of Mathematical Statistics*, **29**, 235-244 (1958).
35. Gupta, S. S. & Sobel, M. 'Selecting a subset containing the best of several binomial populations', *Contributions to Probability and Statistics*. Stanford University Press, 224-248, 1960.
36. Tong, Y. L. 'On partitioning a set of normal populations by their locations with respect to a control', *Annals of Mathematical Statistics*, **40**, 1300-1324 (1969).
37. 抗菌薬臨床評価のガイドライン, 厚生省 1998.
38. Lan, K. K. G. & DeMets, D. L. 'Group sequential procedures : Calendar versus information time', *Statistics in Medicine*, **8**, 1191-1198 (1989).
39. Whitehead, J. 'Overrunning and underrunning in sequential clinical trials', *Controlled Clinical Trials*, **13**, 106-121 (1992).
40. Jennison, C. & Turnbull, B. W. 'Interim analysis: The repeated confidence interval approach', *Journal of Royal Statistics Society, Series B*, **51**, 305-361 (1989).
41. Whitehead, J. *The Design and Analysis of Sequential Clinical Trials* Revised second edition, Wiley, New York, 1997.

42. Paulson, E. 'On the comparison of several experimental categories with a control', *Annals of Mathematical Statistics*, **23**, 239-246 (1952).
43. Simon, R., Thall, P. F. & Ellenberg, S. S. 'New designs for the selection of treatments to be tested in randomized clinical trials', *Statistics in Medicine*, **13**, 417-429 (1994).
44. Thall, P. F., Simon, R. & Ellenberg, S. S. 'Two-stage selection and testing designs for comparative clinical trials', *Biometrika*, **75**, 303-310 (1988).
45. 永田靖, 吉田道弘. 統計的多重比較法の基礎, サイエンス社 1997.
46. Torgersen, E. *Comparison of Statistical Experiments*. Cambridge University Press, Cambridge (1991).
47. Kusama, T. & Koyama, N. 'Complete classes in comparison of sequential binomial experiments', *Statistics & Decisions*, **18**, 27-34 (2000) *To appear*.