

抗がん剤の2段階デザインにおける 奏効確率の推定法の性能比較

豊 泉 滋 之*, 浜 田 知久馬**, 吉 村 功**

Comparison of response rate estimator of anti-cancer drug in two-stage design

Shigeyuki Toyozumi*, Chikuma Hamada** and Isao Yoshimura**

抗がん剤の開発の中で, 第II相臨床試験の主目的は, 被験薬を患者に投与することで腫瘍縮小効果を評価し, 開発を継続するか中止するかを決定することである. 開発を継続する場合には, 腫瘍が規定の大きさまで縮小する患者(奏効患者)の母集団における比率, すなわち奏効確率の推定が重要である. 奏効確率の推定には, 最尤推定量(MLE)を用いるのが自然であるが, 2段階デザインにおいては, 一様最小分散不偏推定量(UMVUE)の方を用いるべきだという主張もあり, 両者の長短は明確でない. 本論文では, 優れた抗がん剤を開発する立場から, 開発を継続するときのみの奏効確率の推定という実務的視点で, 両者を理論的・数量的に比較した. 結果として, 奏効確率が問題になるような条件を想定した場合には, バイアスという意味でも, 平均二乗誤差という意味でも, MLEの方が優れているという結論に達した.

The role of phase II study for the test drug in cancer clinical trials is to evaluate whether it has a desirable level of anti-tumor activity. A two-stage design is often used in the phase II study for determining the succession of development. Jung and Kim (2004) explicitly showed the uniformly minimum variance unbiased estimator (UMVUE) of the response rate in the two-stage design and asserted its superiority to the maximum likelihood estimator (MLE). We mathematically and numerically examined the performance of UMVUE and MLE from the viewpoint that the estimator would be used only when the succeeding phase of development was provided for the test drug. It was revealed that MLE had smaller bias and smaller mean squared error than UMVUE with practically reasonable population response rates.

Key Words and Phrases: response rate estimator, UMVUE, MLE, two-stage design phase II study

1. はじめに

抗がん剤の開発の中で第II相臨床試験の主目的は, 患者に被験薬を投与した場合の腫瘍縮小効果を評価し, 開発を継続するか中止するかを決定することである. 第II相試験の対象となる被験薬の数は多く, 速やかに有効な被験薬を選別する必要があるため, 評価項目として長期間の追跡調査が必要な直接の延命効果ではなく, 腫瘍縮小の有無が用いられる. この腫瘍縮小を測る指標として, 治療により患者が奏効する確率が用いられる. この確率を本稿では「奏効確率」と定義する. また, 「奏効」の医学的な定義は, 「CTなどの画像診断で測定可能な腫瘍の面積が50%以上縮小した状態が4週間以上持続すること」(福田治彦ら(2004))とするのが

* ブリストル・マイヤーズ株式会社, 〒163-1328, 東京都新宿区西新宿 6-5-1 新宿アイランドタワー

** 東京理科大学工学部経営工学科, 〒162-8601, 東京都新宿区神楽坂 1-3

一般的であるが、がんの種類や病変により奏効の定義は異なる。本稿では全被験者の中で奏効した被験者の割合を標本奏効割合とよんで、母数の奏効確率と区別する。

腫瘍縮小効果のない被験薬による治療を受ける被験者を少なくしたいという倫理的な理由により、試験の途中で中間解析を行い効果のない場合は、試験を早期に終了する2段階デザインが広く用いられている。実際にはこの統計的な決定と被験薬の効果や毒性などの臨床・医学的な考察を含めて、最終的に開発を続けるか否かが決定される。このとき奏効確率の推定値は重要な腫瘍縮小効果の指標となる。特に奏効例数が開発の進行を決める境界例数付近の場合や、被験薬の効果を他の薬剤と比較した相対的な位置付けを行う場合には、奏効確率の信頼区間を含めた推定が重要になる。さらに、奏効確率の推定値は第Ⅲ相試験を含めた次相における症例数設計等の試験デザインの参考値となる。このように、奏効確率の推定値は様々な場面で利用されるため、適切に奏効確率を推定する必要がある。特に、第Ⅱ相試験のように開発の初期段階では情報が少なく、薬剤の効果や毒性が不明確な状況であるため、腫瘍縮小効果を定量的に評価する上で奏効確率を推定することは重要である。

直感的に考えると、2段階デザインでの奏効確率の推定には、各段階における最尤推定量 (maximum likelihood estimator; MLE), すなわち標本奏効割合を用いるのが自然である。しかし、一様最小分散不偏推定量 (uniformly minimum variance unbiased estimator; UMVUE) が存在するなら、それを用いた方がよいという考え方もある。実際、Jung and Kim (2004) は Lehmann (1983) に示されている一般的な UMVUE の構成法を2段階デザインに適用して明示的な推定量を構成し、MLE よりこれを使うべきだと主張している。これは尤度原理と不偏性原理の優劣の議論に帰着する問題であるが、本稿ではそのような一般論ではなく、抗がん剤開発の第Ⅱ相臨床試験における、2段階デザインの下での奏効確率の推定という実務的立場から、抗がん剤の有効性の評価にはどちらが合理的であるかを検討した。以下にその結果を示す。

2. 試験デザインと手順

2.1 二者択一のデザイン

第Ⅱ相試験の試験デザインには、腫瘍縮小効果に応じて被験薬の開発を進める合格か、不合格かの二者択一の決定方式が求められる。医師や患者が望む被験薬を開発するために、決定方式は2つの条件を満たさなければならない。第1の条件は「効果のない被験薬をできるだけ低い確率で合格と判定すること」であり、第2の条件は「効果のある被験薬をできるだけ高い確率で合格と判定すること」である。すなわち、奏効確率 π の被験薬の合格確率を $\Pr(\text{合格} | \pi)$ と表したとき、これ以下では開発の価値がない「閾値奏効確率」 π_0 と、これ以上であれば開発の価値がある「期待奏効確率」 π_1 を定めておいて、

$$\Pr(\text{合格} | \pi_0) \leq \alpha \quad (1)$$

$$\Pr(\text{合格} | \pi_1) \geq 1 - \beta \quad (2)$$

を試験デザインが満たすべき条件とすることである。

π_0 と π_1 は対象とするがんの種類や進行度を表すステージなどにより異なるため、他の同様な薬剤の試験や、海外で行われた試験で得られた情報を下に、試験ごとに決定される。従来の抗がん剤の奏効確率は低く一般に5~30%を閾値奏効確率とし、期待奏効確率は閾値奏効確率より15~20%大きい値を設定することが多い。 α は5~10%、 β は10~20%に設定することが多い。これら4つのパラメータ $\pi_0, \pi_1, \alpha, \beta$ の値を定め、条件(1), (2)式を満たすように試験デザインを作成する。

2.2 2段階デザイン

1段階デザインでは全被験者の奏効の有無を診断した上で、薬剤の合格・不合格を判定することになる。しかしながら、一般にがんの臨床試験で用いられる被験薬は奏効の有無に関わらず、毒性は常に存在するため、試験途中で被験薬に腫瘍縮小効果がないと確信できた場合には早急に試験を中止し、腫瘍縮小効果のない治療を受ける被験者を少なくしたいという倫理的な要求が存在する。また無効による早期中止を考慮することにより、製薬企業にとっては時間や費用などの資源を他の候補薬剤の開発に回すことが可能になる。これらの要求を満たすために、試験を2段階に分け、中間解析を行う2段階デザインが広く用いられる。中間解析の時点で極端に効果が大きいと分かった場合に試験を中止し、次相の試験を繰り上げて開始することも考えられるが、極端に強い効果がある被験薬は少なくこれまで有効による早期中止が考慮されることは少なかった。

第1段階、第2段階の症例数をそれぞれ n_1, n_2 として試験を行うことを考える。この試験結果として各段階の奏効例数 X_1, X_2 が得られる。これらは互いに独立な確率変数であり、それぞれ二項分布 $B(n_1, \pi), B(n_2, \pi)$ に従うとする。この仮定の下で条件(1), (2)式を満たす2段階デザインは数多く提案されている。

Fleming (1982) は二項分布の正規近似を利用して、条件を満たす2段階デザインを一意に決定する方法を提案した。このFleming法は中間解析において有効の早期中止と無効の早期中止を共に考慮した方法である。臨床試験の被験者数は、倫理的・制約的な側面から必要最低限にする必要があり、Simon (1989) はこの要求を考慮するために、条件(1), (2)式を満たす試験デザインとして、各段階の症例数 n_1, n_2 と、中間・最終解析の境界例数 r, R の組合せを列挙し、その中から最適な n_1, n_2, r , および R の組を選択する方法を提案した。最適なデザインを選択する規準として、症例数の期待値 $E[n]$ が最小である期待症例数最小化規準と、最大の症例数 $n_1 + n_2$ が最小であるミニマックス規準の2つがある。Simon法は無効による早期中止のみを考慮した試験デザインであり、抗がん剤の第II相試験において頻用されている。本稿ではSimon法による2段階デザインを主に取り上げる。

2.3 試験手順

第II相試験で最終的に被験薬の開発を継続すると判断されるためには、中間解析と最終解析で、共に開発の継続と判定される必要がある。本稿では中間解析での継続判定を「予選合格」、最終解析での継続判定を「合格」とよぶことにする。Simon法は、各段階の症例数 n_1, n_2 と、予選合格、合格の判定基準となる境界例数 r, R により一意に定まり、以下の手順で行なわれる。

STEP 1 n_1 例の被験者で第1段階の試験を行う。

STEP 2 n_1 例の試験が終了した時点で中間解析を行う。

$$X_1 \leq r \Rightarrow \text{早期中止}$$

$$X_1 > r \Rightarrow \text{予選合格, STEP 3 の第2段階へ進める}$$

STEP 3 新たに n_2 例の被験者を追加して第2段階の試験を行う。

STEP 4 n_2 例の試験が終了した時点で最終解析を行う。

$$X_1 + X_2 \leq R \Rightarrow \text{不合格}$$

$$X_1 + X_2 > R \Rightarrow \text{合格, 次相へ開発を進める}$$

3. 奏効確率の推定量

3.1 奏効確率の完備十分統計量

無効による早期中止の2段階デザインを定式化する. 第2段階へ進むか否かの予選合格の境界例数 r は $0 \leq r < n_1$ を満たす整数である. $X_1 \leq r$ のときは X_2 を観測せず, $X_1 > r$ のときのみ X_2 を観測する. このデザインで得られる X_1 と X_2 の実現値により奏効確率 π を推定する. ここで2つの実現値をまとめた累積奏効例数の統計量 $S = s(X_1, X_2)$ を $S = s(X_1, X_2) = X_1 + X_2 I_r(X_1)$ と定義する. ただし, $I_r(\bullet)$ は $\bullet \leq r$ のとき0, そうでなければ1という値を取る指示関数である. 試験の終了段階ごとの累積奏効例数 S は(3)式で示される.

$$S = \begin{cases} X_1, & X_1 \leq r \text{ のとき} \\ X_1 + X_2, & X_1 > r \text{ のとき} \end{cases} \quad (3)$$

さらに, 累積奏効例数 S はその定義により式(4)の性質が生じる.

$$\begin{aligned} S &\leq r, & X_1 &\leq r \text{ のとき} \\ S &> r, & X_1 &> r \text{ のとき} \end{aligned} \quad (4)$$

Jung and Kim (2004) は有効による早期中止を含めた一般的な多段階デザインの試験において, 累積奏効例数 S と試験終了段階 M の組合せ (S, M) が奏効確率の完備十分統計量になることを示したが, 本稿が対象とする無効による早期中止の2段階デザインでは, その特別な場合として, 上記のように指示関数を取り入れた累積奏効例数 S が完備十分統計量となっている.

3.2 最尤推定量 (MLE)

奏効確率の推定量としてMLEである標本奏効割合が最も一般的に用いられている. 標本奏効割合 $\hat{\pi}$ は(5)式のように累積奏効例数を累積症例数で割ったものである.

$$\hat{\pi} = \frac{S}{n_1 + I_r(S)n_2} = \begin{cases} \frac{X_1}{n_1}, & S \leq r \\ \frac{X_1 + X_2}{n_1 + n_2}, & S > r \end{cases} \quad (5)$$

MLEは早期中止の場合に第1段階の奏効割合となり, 予選合格して第2段階で試験が終了した場合には各段階の結果を単純に併合した, 直感的で自然な統計量となる. 医師や臨床家など統計に詳しくない実務家にも理解しやすい指標といえる. また, 最尤推定量であるため, 漸近有効性が期待される. 1段階デザインであればMLEは不偏推定量であるが, 2段階デザインではバイアスが生じることが知られている. 特に無効による早期中止デザインにおいては奏効確率を過小評価する (Chang (1989), Jung and Kim (2004)). 真の奏効確率が π であるときのMLEのバイアスは(6)式で示される.

$$\text{bias}(\hat{\pi}|\pi) = E[\hat{\pi}] - \pi = \frac{n_2}{n_1(n_1 + n_2)} \sum_{x_1=0}^r (x_1 - n_1\pi) {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n_1-x_1} \quad (6)$$

MLEの $100(1-\alpha)\%$ 信頼区間は一般的な二項分布から得られる信頼区間として求められる. Clopper and Pearson (1934) による正確な信頼区間 $(\hat{\pi}_L, \hat{\pi}_U)$ は(7), (8)式に示したF分布の%点を利用して推定される.

$$\hat{\pi}_L = \frac{\nu_2}{\nu_2 + \nu_1 F_{\alpha/2}(\nu_1, \nu_2)} \quad \nu_1 = 2(n_1 + I_r(S)n_2 - S + 1), \nu_2 = 2S \quad (7)$$

$$\hat{\pi}_U = \frac{\lambda_1 F_{\alpha/2}(\lambda_1, \lambda_2)}{\lambda_2 + \lambda_1 F_{\alpha/2}(\lambda_1, \lambda_2)} \quad \lambda_1 = 2(S + 1), \lambda_2 = 2(n_1 + I_r(S)n_2 - S) \quad (8)$$

3.3 一様最小分散不偏推定量 (UMVUE)

Jung and Kim (2004) により多段階デザインにおける奏効確率の UMVUE が明示的に示された。(9) 式に無効による早期中止 2 段階デザインの場合の UMVUE $\hat{\pi}$ を示す。第 1 段階で早期中止する場合の奏効確率の推定値は MLE と等しく標本奏効割合であり、MLE と異なるのは第 2 段階で試験が終了する場合である。

$$\hat{\pi} = \begin{cases} \frac{S}{n_1}, & S \leq r \\ \frac{\sum_{i=\max(r+1, S-n_2)}^{\min(S, n_1)} n_1 - 1 C_{i-1} n_2 C_{S-i}}{\sum_{j=\max(r+1, S-n_2)}^{\min(S, n_1)} n_1 C_j n_2 C_{S-j}}, & S > r \end{cases} \quad (9)$$

Jung and Kim は Rao-Blackwell の定理より、任意の不偏推定量 π^* に完備十分統計量 S を条件付けた期待値 $E[\pi^* | S=s]$ は UMVUE であることから、不偏推定量である第 1 段階の奏効割合 X_1/n_1 を累積奏効例数 S で条件付けた期待値 $E[X_1/n_1 | S=s]$ を計算することで、UMVUE を求めた。バイアスが存在する MLE に比べ、UMVUE はその名の通り母数の値に関わらず不偏推定量の中で分散が最も小さいという数理的な特徴がある。

無効による早期中止デザインにおける、UMVUE の正確な $100(1-\alpha)\%$ 信頼区間 (\hat{p}_L, \hat{p}_U) を、数値計算で求めることで構成することができる。ただし s は累積奏効例数 S の実現値である。

$$\Pr(\hat{\pi}(S) \geq \hat{\pi}(s) | \pi = \hat{\pi}_L) = \alpha/2 \quad (10)$$

$$\Pr(\hat{\pi}(S) \leq \hat{\pi}(s) | \pi = \hat{\pi}_U) = \alpha/2 \quad (11)$$

本稿ではこれを SAS/IML の最適化ルーチンを用いて算出した (豊泉 (2005))。

3.4 MLE と UMVUE の性質

本稿では (α, β) を $(0.05, 0.10)$ に固定した上で (π_0, π_1) を変化させたパターン A と、 (π_0, π_1) を $(0.2, 0.4)$ に固定した上で (α, β) を変化させたパターン B の 2 つの状況を設定する。パターン A では抗がん剤の奏効確率は一般的にそれほど高くないので、実際に計画されうる状況として、 $(\pi_0, \pi_1) = (0.1, 0.3), (0.2, 0.4), (0.3, 0.5)$ の 3 つを設定する。パターン B では、 $(\alpha, \beta) = (0.10, 0.10), (0.05, 0.20), (0.05, 0.10)$ の 3 つの状況を設定する。さらに、Simon 法においては最適化規準として期待症例数最小化規準とミニマックス規準の 2 つがあるため、12 通りのデザインについて比較を行った。

検討を行ったデザインはどれも定性的な傾向が同じであったため、本稿では典型例として、表 1 の場合における MLE と UMVUE の点推定値と信頼区間を図 1 に示す。このグラフは横軸が累積奏効例数 S であり、縦軸が対応した点推定値と 95% 信頼区間である。MLE は標本奏効

表 1 性能比較に用いる Simon 法による試験デザインの一例

最適化規準	π_0	π_1	α	β	n_1	n_2	r	R
期待症例数最小化規準	0.2	0.4	0.05	0.20	13	30	3	12

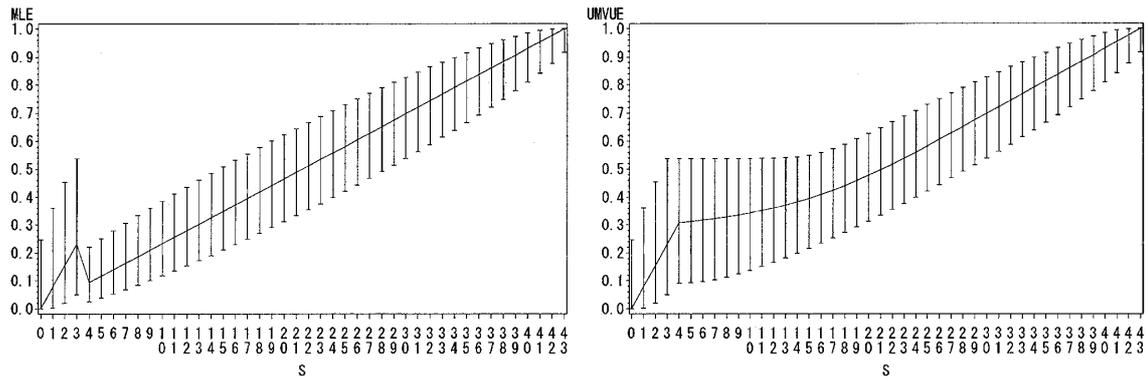


図1 MLEとUMVUEの点推定値と信頼区間

割合であるため試験終了段階ごとに累積奏効例数に対して直線的に増加する。予選合格の境界 ($S=4$) において、標本奏効割合の分母が第1段階と第2段階の例数が合わさり増大するため、推定値が一旦下がることに注意をされたい。

UMVUE は中間解析において予選合格したという情報を第2段階の推定値に反映させているため、試験終了段階に関わらず累積奏効例数に対し単調増加である。予選合格となり本選に進む場合、予選合格の境界例数 r における奏効割合 r/n_1 より大きい推定値を与えるため、予選敗退した場合に比べて本選に進んだ場合の推定値を高く評価するという観点からは妥当である。

累積奏効例数 S が小さい場合は、第1段階である程度の奏効が得られたにもかかわらず、第2段階ではほとんど奏効が得られず、各段階の結果に矛盾が生じた場合である。このとき UMVUE の信頼区間はこの矛盾した結果を反映して広くなる。また、真の奏効確率がかなり高い場合は、早期中止になる可能性がほとんどなく、 S は X_1+X_2 になり、MLE と UMVUE は $(X_1+X_2)/(n_1+n_2)$ で一致する。

4. 推定量の比較

4.1 比較基準

UMVUE は第1, 第2段階を通した不偏性を持つが、この性質が実務的な観点から必ずしも望ましいとは限らない。なぜなら、第II相試験で不合格という結果が得られた場合に奏効確率をわざわざ推定することはなく、推定が問題となるのは合格と判定される場合のみである。これは入学試験を行った大学が不合格者の能力を評価しようとしなないのと同じで、評価をしても得るものはないからである。

したがって、通常は第2段階で合格した条件下 ($S>R$) でのみ奏効確率の推定が必要になる。一步譲ったとしても、予選合格 ($S>r$) という条件までが意味ある限度である。このような実務的な観点から、本稿では合格または予選合格という条件付の下で、推定量の偏りと平均二乗誤差 (mean squared error; MSE) を評価することにする。また、それぞれの正確な信頼区間を、真の値を含む確率である被覆確率により評価する。

4.2 予選合格条件付バイアス

早期終了の状況を除いているため、推定値が小さな値となる場合が生じず、バイアスは MLE・UMVUE を問わず常に正である。条件付バイアスは、条件付期待値から真の奏効確率 π を引いたものである。MLE と UMVUE の予選合格条件付バイアスは (12), (13) 式で示される。

表2 性能比較の規準

1. 予選合格条件付バイアス	$E[\hat{\pi} S>r]-\pi$
2. 合格条件付バイアス	$E[\hat{\pi} S>R]-\pi$
3. 予選合格条件付平均二乗誤差	$MSE[\hat{\pi} S>r]$
4. 合格条件付平均二乗誤差	$MSE[\hat{\pi} S>R]$
5. 予選合格条件付被覆確率	$CP[\hat{\pi} S>r]=\sum_{S \in (\hat{\pi}_L, \hat{\pi}_U)} \Pr(S S>r)$
6. 合格条件付被覆確率	$CP[\hat{\pi} S>R]=\sum_{S \in (\hat{\pi}_L, \hat{\pi}_U)} \Pr(S S>R)$

$$E[\hat{\pi}|S>r]-\pi = \frac{\sum_{s=r+1}^n \left\{ \frac{s}{n} \cdot \sum_{i=\max(r+1, s-n_2)}^{\min(s, n_1)} C_i n_2 C_{s-i} \pi^s (1-\pi)^{n-s} \right\}}{\sum_{j=r+1}^{n_1} C_j \pi^j (1-\pi)^{n_1-j}} - \pi \quad (12)$$

$$E[\tilde{\pi}|S>r]-\pi = \frac{\sum_{s=r+1}^n \left\{ \pi^s (1-\pi)^{n-s} \sum_{i=\max(r+1, s-n_2)}^{\min(s, n_1)} C_{i-1} n_2 C_{s-i} \right\}}{\sum_{l=r+1}^{n_1} C_l \pi^l (1-\pi)^{n_1-l}} - \pi \quad (13)$$

これらの条件付バイアスを MLE のバイアスと UMVUE のバイアスの比 (bias ratio; BR) を用いて評価する. 2つのバイアスの比は真の奏効確率によらず一定であり, 評価が単純になる. 予選合格条件付バイアスの比 BR ($S>r$) は (14) 式に示すように, 最大の症例数 $n=n_1+n_2$ を第1段階の症例数 n_1 で割った定数となり, 真の奏効確率 π によらないことが数理的に導ける (付録 A 参照).

$$BR(S>r) = \frac{E[\tilde{\pi}|S>r]-\pi}{E[\hat{\pi}|S>r]-\pi} = \frac{n}{n_1} \quad (14)$$

2段階デザインにおいては $n_1 < n$ が常に成り立つため, 予選合格条件付バイアスは MLE に比べて UMVUE が常に大きく, MLE に対して $BR(S>r) = n/n_1$ 倍のバイアスを持つ. 第1段階の症例数 n_1 の大きさは最大の症例数 n の半分前後であるため, バイアスの比はデザインによるが 1.5~3 倍程度である. 表1に示した Simon 法の試験デザインにおける予選合格条件付バイアスを図2に示す. MLE と UMVUE の予選合格条件付バイアスは期待奏効確率 $\pi_1 = 0.4$ の下で 0.012 と 0.041 であり, 比は 3.3 倍である.

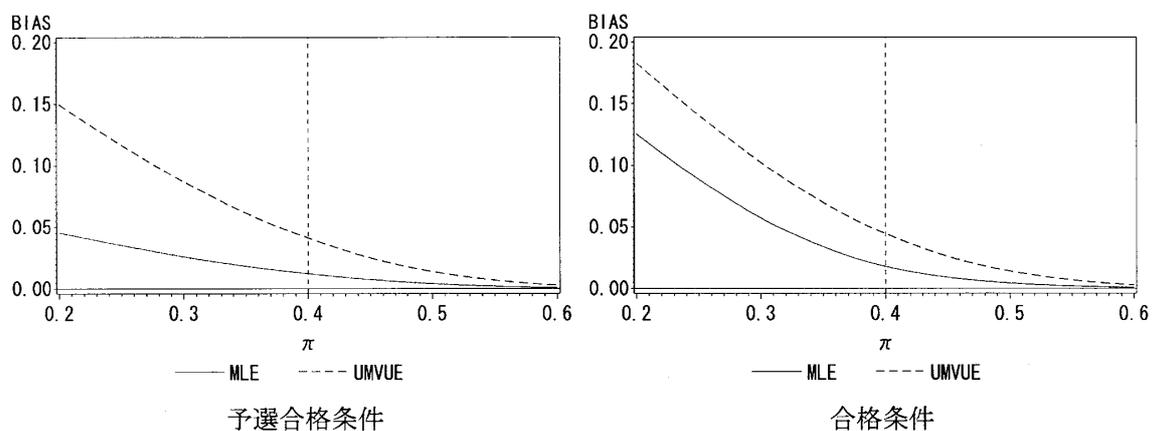


図2 条件付バイアス

4.3 合格条件付バイアス

合格条件付バイアスは (15), (16) 式で示される.

$$E[\hat{\pi} | S > R] - \pi = \frac{\sum_{s=R+1}^n \left\{ \frac{s}{n} \cdot \sum_{k=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_k n_2 C_{s-k} \pi^s (1-\pi)^{n-s} \right\}}{\sum_{i=R+1}^n \sum_{j=\max(r+1, i-n_2)}^{\min(i, n_1)} n_1 C_j n_2 C_{i-j} \pi^i (1-\pi)^{n-i}} - \pi \quad (15)$$

$$E[\tilde{\pi} | S > R] - \pi = \frac{\sum_{s=R+1}^n \left\{ \pi^s (1-\pi)^{n-s} \sum_{k=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_{k-1} n_2 C_{s-k} \right\}}{\sum_{i=R+1}^n \sum_{j=\max(r+1, i-n_2)}^{\min(i, n_1)} n_1 C_j n_2 C_{i-j} \pi^i (1-\pi)^{n-i}} - \pi \quad (16)$$

合格条件付バイアスの比 $BR(S > R)$ は (16) 式を (15) 式で割った値であり (17) 式で表される.

$$\begin{aligned} BR(S > R) &= \frac{E[\tilde{\pi} | S > R] - \pi}{E[\hat{\pi} | S > R] - \pi} \\ &= \frac{n \sum_{s=R+1}^n \left\{ \pi^s (1-\pi)^{n-s} \sum_{i=\max(r+1, s-n_2)}^{\min(s, n_1)} (i - n_1 \pi) n_1 C_i n_2 C_{s-i} \right\}}{n_1 \sum_{t=R+1}^n \left\{ \pi^t (1-\pi)^{n-t} (t - n\pi) \sum_{j=\max(r+1, t-n_2)}^{\min(t, n_1)} n_1 C_j n_2 C_{t-j} \right\}} \end{aligned} \quad (17)$$

合格条件と対応する各段階の奏効例数 X_1, X_2 の標本空間が複雑であり, バイアスの比は予選合格条件のように定数にはならず, 真の奏効確率 π に依存する. 合格条件付バイアスは数式での定式的な評価は困難であるが, MLE よりも UMVUE のバイアスが大きいことはパターン A・パターン B のどちらのデザインでも確認できた. 推定が重要となるのは, 合格と見なされる期待奏効確率 π_1 近傍におけるバイアスである. 期待奏効確率 π_1 における UMVUE のバイアスは, 約 1.5~2.5 倍 MLE のバイアスより大きい. 表 1 で示した試験デザインにおける合格条件付バイアスの結果を図 2 に示す. 期待奏効確率 $\pi_1 = 0.4$ における MLE と UMVUE のバイアスは 0.018 と 0.044 であり, 比は 2.4 倍である.

UMVUE は予選合格と合格のどちらの条件においても, MLE に比べてバイアスが大きい. これは, UMVUE が不偏性を満たすために MLE に比べて第 2 段階の推定値を過大にしており, 小さい奏効例数で試験が終了する場合は除かれる条件付の下では, MLE に比べて大きな正のバイアスを持つからである.

4.4 予選合格条件付平均二乗誤差

MLE $\hat{\pi}$ と UMVUE $\tilde{\pi}$ の予選合格条件付平均二乗誤差は (18), (19) 式のようにになる.

$$MSE[\hat{\pi} | S > r] = \sum_{s=r+1}^n \left\{ \left(\frac{s}{n} - \pi \right)^2 \frac{\sum_{i=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_i n_2 C_{s-i} \pi^s (1-\pi)^{n-s}}{\sum_{j=r+1}^{n_1} n_1 C_j \pi^j (1-\pi)^{n_1-j}} \right\} \quad (18)$$

$$\begin{aligned} &MSE[\tilde{\pi} | S > r] \\ &= \sum_{s=r+1}^n \left\{ \left(\frac{\sum_{i=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_{i-1} n_2 C_{s-i}}{\sum_{j=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_j n_2 C_{s-j}} - \pi \right)^2 \frac{\sum_{k=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_k n_2 C_{s-k} \pi^s (1-\pi)^{n-s}}{\sum_{l=r+1}^{n_1} n_1 C_l \pi^l (1-\pi)^{n_1-l}} \right\} \end{aligned} \quad (19)$$

2つの推定量の予選合格条件付平均二乗誤差は条件付バイアスと異なり, 真の奏効確率によって大きさや大小関係が変化する. 予選合格条件付平均二乗誤差は数式のまま評価することが困難なため, 数値計算により定量的に評価する. 奏効確率を横軸に, (18), (19) 式の平方根

($\sqrt{\text{MSE}}$) を縦軸にとったのが図3の左のグラフである.

閾値奏効確率 π_0 と期待奏効確率 π_1 の中間 $(\pi_0 + \pi_1)/2$ 以下において UMVUE の予選合格条件付平均二乗誤差が極端に大きくなっている. これは, UMVUE がぎりぎりでも予選合格した薬剤の奏効確率を過大評価しているからである. 逆に中間 $(\pi_0 + \pi_1)/2$ より大きい奏効確率においてはほぼ同じ大きさであり, 実務的に重要となる期待奏効確率 π_1 の近傍に於いては, わずかに MLE の方が大きい. 条件付平均二乗誤差の平方根は条件付バイアスに対して一桁大きいため, バイアスに比べてバラツキが大きい. これは症例数が数十例程度で少ないためである.

4.5 合格条件付平均二乗誤差

MLE と UMVUE の合格条件付平均二乗誤差は (20), (21) 式ようになる.

$$\text{MSE}[\hat{\pi}|S>R] = \sum_{s=R+1}^n \left\{ \left(\frac{s}{n} - \pi \right)^2 \frac{\sum_{k=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_k n_2 C_{s-k} \pi^s (1-\pi)^{n-s}}{\sum_{i=R+1}^n \sum_{j=\max(r+1, i-n_2)}^{\min(i, n_1)} n_1 C_j n_2 C_{i-j} \pi^i (1-\pi)^{n-i}} \right\} \quad (20)$$

$\text{MSE}[\hat{\pi}|S>R]$

$$= \sum_{s=R+1}^n \left\{ \left(\frac{\sum_{i=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_{i-1} n_2 C_{s-i}}{\sum_{j=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_j n_2 C_{s-j}} - \pi \right)^2 \frac{\sum_{k=\max(r+1, s-n_2)}^{\min(s, n_1)} n_1 C_k n_2 C_{s-k} \pi^s (1-\pi)^{n-s}}{\sum_{i=R+1}^n \sum_{j=\max(r+1, i-n_2)}^{\min(i, n_1)} n_1 C_j n_2 C_{i-j} \pi^i (1-\pi)^{n-i}} \right\} \quad (21)$$

(20), (21) 式の平方根 ($\sqrt{\text{MSE}}$) をとったのが図3の右のグラフである. 図3の左右のグラフを比較すると, 奏効確率が $(\pi_0 + \pi_1)/2$ 以下において, 予選合格条件よりも2つの合格条件付平均二乗誤差の差は縮まっているが, UMVUE の方が大きいことには変わりはない. また, 期待奏効確率 π_1 の近傍に於いては平均二乗誤差に大きな差があるとはいえない.

4.6 予選合格条件付被覆確率

推定が問題となるのは薬剤が合格する場合であるため, 真の被覆確率が期待奏効確率 π_1 近傍での被覆確率が名義信頼水準 $(1-\alpha)$ を保証するかどうか重要となる. 条件付被覆確率は信頼区間が真値を含む累積奏効例数 $S \in (\hat{\pi}_L, \hat{\pi}_U)$ についての条件付確率 $\text{Pr}(S|S>r)$ の和により求められる. MLE の信頼区間を $(\hat{\pi}_L, \hat{\pi}_U)$, UMVUE の信頼区間を $(\tilde{\pi}_L, \tilde{\pi}_U)$ で表すと, それぞれの信頼区間の予選合格条件付被覆確率は (22), (23) 式で与えられる.

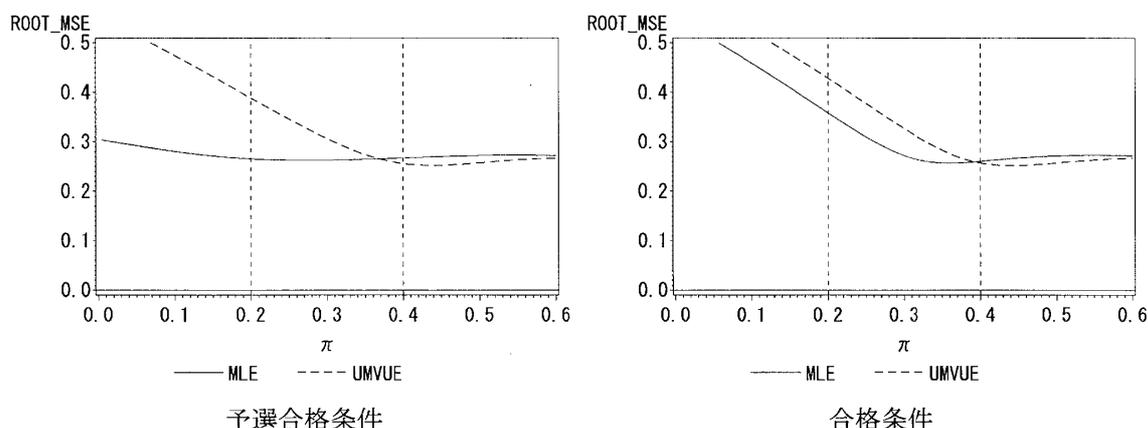


図3 条件付 $\sqrt{\text{MSE}}$

$$CP[\hat{\pi} | S > r] = \sum_{S \in (\hat{\pi}_L, \hat{\pi}_U)} \Pr(S | S > r) \quad (22)$$

$$CP[\hat{\pi} | S > r] = \sum_{S \in (\hat{\pi}_L, \hat{\pi}_U)} \Pr(S | S > r) \quad (23)$$

4.7 合格条件付被覆確率

合格条件付被覆確率は (24), (25) 式で与えられる.

$$CP[\hat{\pi} | S > R] = \sum_{S \in (\hat{\pi}_L, \hat{\pi}_U)} \Pr(S | S > R) \quad (24)$$

$$CP[\hat{\pi} | S > R] = \sum_{S \in (\hat{\pi}_L, \hat{\pi}_U)} \Pr(S | S > R) \quad (25)$$

期待奏効確率 π_1 付近の条件付被覆確率を数値計算した結果を図4に示す. 図4の横軸である奏効確率は, 期待奏効確率 π_1 を中心に, その前後 0.2 の範囲を示している.

図4の信頼区間の被覆確率によると, 期待奏効確率 π_1 付近では条件付被覆確率は常に名義信頼水準を上回るが, 小さい奏効確率である閾値奏効確率 π_0 に近づくにつれて名義信頼水準を大きく下回ることがある. これは小さい奏効確率の推定値となる場合を除いた条件付で評価しているからである.

第II相試験は数十例の症例数で行われることから, 図1に示したようにMLEとUMVUEの95%信頼区間は一般的に広くなる(点推定値 $\pm 10\% \sim 15\%$). 第2段階におけるMLEとUMVUEの95%信頼区間を比較すると, UMVUEの点推定値がMLEを上回ると同様に, UMVUEの信頼区間はMLEに比べて上限が大きくなり区間幅は広くなる. これによりUMVUEの信頼区間に真の奏効確率を含む累積奏効例数が増え, 被覆確率がMLEに比べて大きくなる. 例えば, 予選合格条件下の期待奏効確率0.4の場合, MLEの信頼区間が0.4を含む累積奏効例数は $s = \{11, 12, \dots, 24\}$ であるのに対して, UMVUEの信頼区間が含む累積奏効例数は $s = \{4, 5, \dots, 24\}$ であり, 被覆確率の差は $\sum_{s=4}^{10} \Pr(S=s | S > r) = 0.005$ となる. 真の奏効確率が0.5近傍である場合に, 2つの信頼区間が含む累積奏効例数の差が最も大きくなるため, 被覆確率の差も最も大きくなる. また, 確率関数が離散型であることにより, 被覆確率は名義信頼水準と比べて保守的であるが, 上記の理由によりUMVUEに比べてMLEの被覆確率の方がその保守性が緩和されており, より名義信頼水準に近い値となり, 区間幅はUMVUEに比べて狭くなることから, MLEの信頼区間はUMVUEに比べて望ましい性質を持っているといえる.

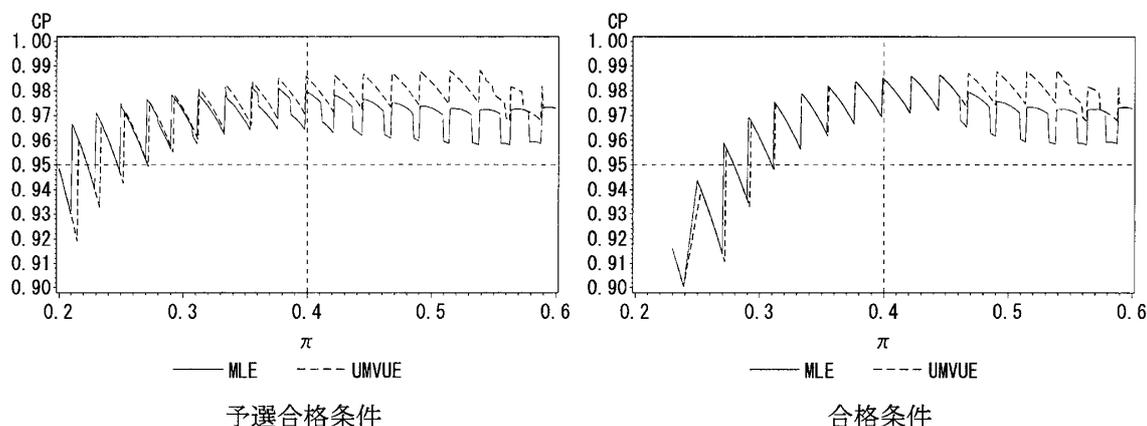


図4 条件付被覆確率

5. 考察と総括

UMVUE は不偏性を満たす推定量のクラスの中で分散が最小という優れた性質を持つ推定量である。しかしながら、統計学を現実の問題に応用する上では一般化した広い条件に対して最適な推定量が必ずしも好ましいとは限らない。個々の現実問題には様々な制約や条件が存在する。これは抗がん剤の第II相試験の問題においても例外ではない。

第1段階の推定量はMLEもUMVUEも等しく奏効した割合になる。第1段階の被験者数は10~20例であるため、奏効確率が閾値奏効確率以上であるにもかかわらず偶然にも奏効例数が小さくなった場合、2段階デザインでは早期中止となり、その結果として奏効確率は過小に評価されることになる。すなわち第1段階の推定量は負のバイアスを持つことになる。この第1段階の負のバイアスを代償するため、UMVUEは第2段階に限ればより大きな正のバイアスを持つことになる。

UMVUEの特徴である、第1段階、第2段階を通じて不偏性が成り立つことは第II相試験の推定の問題において必ずしも望ましい性質ではない。なぜならば、不偏性は同じような試験を繰り返し行い、良いときも悪いときも含めて平均的な成績を評価する場合にバイアスが無いという性質であるが、2段階デザインの場合、良い結果が得られたときしか推定を行わないからである。第II相試験の推定で推定量に求められる性質は、合格または予選合格といった条件における推定の精度である。この実務的な観点からMLEとUMVUEを評価した結果、MLEに比べてUMVUEはバイアスにより過大評価される傾向があり、平均二乗誤差は必ずしも小さくなく、信頼区間の被覆確率の点からも望ましいとはいえない。以上の考察より、我々は2段階デザインにおける奏効確率の推定には、UMVUEよりMLEを用いる方がよいと考える。

付録A 予選合格条件付バイアス比の導出

予選合格条件付バイアス比は、 X_1 と X_2 に分解すると、(26)式で示すように最大の症例数 n を第1段階の症例数 n_1 で割った定数となる。

$$\begin{aligned}
 BR(S > r) &= \frac{E[\hat{\pi} | S > r] - \pi}{E[\hat{\pi} | S > r] - \pi} \\
 &= \frac{\sum_{x_1=r+1}^{n_1} \sum_{x_2=0}^{n_2} \left\{ \frac{x_1}{n_1} {}_{n_1}C_{x_1} {}_{n_2}C_{x_2} \pi^{x_1+x_2} (1-\pi)^{n-x_1-x_2} \right\} - \pi \cdot \sum_{x_1=r+1}^{n_1} \left\{ {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n-x_1} \right\}}{\sum_{x_1=r+1}^{n_1} \sum_{x_2=0}^{n_2} \left\{ \frac{x_1+x_2}{n} {}_{n_1}C_{x_1} {}_{n_2}C_{x_2} \pi^{x_1+x_2} (1-\pi)^{n-x_1-x_2} \right\} - \pi \cdot \sum_{x_1=r+1}^{n_1} \left\{ {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n-x_1} \right\}} \\
 &= \frac{\sum_{x_1=r+1}^{n_1} \left(\frac{x_1}{n_1} - \pi \right) {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n-x_1}}{\sum_{x_1=r+1}^{n_1} \left(\frac{x_1+n_2\pi}{n} - \pi \right) {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n-x_1}} \\
 &= \frac{n}{n_1} \cdot \frac{\sum_{x_1=r+1}^{n_1} (x_1 - n_1\pi) {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n-x_1}}{\sum_{x_1=r+1}^{n_1} (x_1 - n_1\pi) {}_{n_1}C_{x_1} \pi^{x_1} (1-\pi)^{n-x_1}} = \frac{n}{n_1} \tag{26}
 \end{aligned}$$

謝辞

本研究は日本学術振興会の科学研究費 No. 16200022 の助成を得た。また、適切なコメントをいただいた査読者に心より感謝いたします。

参 考 文 献

- [1] Chang M. N., Wieand H. S. and Chang V. T. (1989). The bias of the sample proportion following a group sequential phase II clinical trial, *Statistics in Medicine*, **8**, 563-570.
- [2] Clopper C. J. and Pearson E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, **26**, 404-413.
- [3] Fleming T. R. (1982). One-sample multiple testing procedure for phase II clinical trials, *Biometrics*, **38**, 143-151.
- [4] 福田治彦, 新美三由紀, 石塚直樹 (2004). 米国 SWOG に学ぶがん臨床試験の実践, 医学書院.
- [5] Jung S. H. and Kim K. (2004). On the estimation of the binomial probability in multistage clinical trials, *Statistics in Medicine*, **23**, 881-896.
- [6] Lehmann E. L. (1983). *Theory of Point Estimation*, Wiley.
- [7] Simon R. (1989). Optimal two-stage designs for phase II clinical trials, *Controlled Clinical Trials*, **10**, 1-10.
- [8] 豊泉滋之 (2005). 2 段階デザインにおける奏効確率の一樣最小分散不偏推定量のプログラム作成, SAS Forum ユーザー会学術総会論文集, 153-160.