

試験治療効果の発現の遅れを考慮した重み付きログランク検定
- Staggered Entry 下における確率打ち切り法と群逐次法 -

吉田瑞樹

1 序文

がんワクチンの効果発現までには、薬剤投与後、一定の閾値時間の経過を要する。通常の解析手法では、薬剤投与直後から効果が発現すると仮定しており、生存関数の群間比較には、薬剤効果に比例ハザード性を仮定し、その仮定の下で漸近的に最適なログランク検定を用いる。しかし、試験薬剤ががんワクチンの場合、比例ハザード性の仮定は必ずしも成立しないため、生存関数の群間比較にログランク検定を用いると検出力の低下を招く。このような遅延型ワクチン効果の検出に適した解析手法として、重みを $1 - S_{\text{wv}}(t)$ とした重み付きログランク検定¹⁾が提案されている。ただし、 $S_{\text{wv}}(t)$ は時点 t の全群をプールした生存関数の推定値とする。

実際の臨床試験では、試験開始時に全症例を一斉に登録することはまれであり、一般には症例登録期間を設け、その期間内で順次症例登録を行う (staggered entry)。また、長期に及ぶ臨床試験では、無効性及び有効性の早期中止を目的として、試験途中に中間解析を行うことが多い。まず無効性中止目的の中間解析では、試験継続可否の判断を下すための統計学的指標として、条件付き検出力に基づく確率打ち切り法を用いることが多い。しかし、staggered entry 下で、生存関数の群間比較に重み付きログランク検定を用いる場合、スコア過程はブラウン運動の確率過程に収束しないため、条件付き検出力の算出に通常の方法 (Lan-Wittes 法²⁾) を用いることはできない。そこで、本研究では、重み付きログランク検定に基づく条件付き検出力の算出方法を提案する。シミュレーションに基づく方法と、漸近論に基づく解析的な方法の 2 通りの算出方法を提案する。また、ブラウン運動の性質を利用した方法論を利用できるように、重みを調整した重み付きログランク検定統計量を提案する。次に、有効性中止目的の中間解析では、多重性の調整方法として、 α 消費関数法³⁾に基づく群逐次法を用いることが多い。しかし、生存関数の群間比較に重み付きログランク検定を用いる場合、重みが暦時間に依存するため、検定統計量は独立増分構造にならず、無調整で α 消費関数法から算出した棄却限界値を用いると、第 I 種の過誤確率の増大を招く恐れがある。そこで、本研究では、無調整の重み付きログランク検定統計量に対して α 消費関数法を適用した際の第 I 種の過誤確率を定量的に評価する。さらに、第 I 種の過誤確率の正確な調整が可能な方法論を提案する。

2 想定する臨床試験

エンドポイントを生存期間とした 2 群比較臨床試験を実施する。対照群を標準治療、試験群を標準治療とがんワクチンの併用治療とする。一定の症例登録期間を設けて症例集積を行い、追跡調査終了後、重みを $1 - S_{\text{wv}}(t)$ とした重み付きログランク検定を行う。試験途中に無効性或いは有効性中止目的の中間解析を実施する。試験全体の過誤確率は片側 0.025 とする。

3 方法

3.1 各群の生存期間データに対して仮定するモデル

対照群は通常の殺細胞性の薬剤であるため、生存関数に対してハザードが $\gamma\lambda_0 t^{\gamma-1}$ のワイブル分布を仮定する。一方、試験群はがんワクチン併用群であるため、生存関数に対して変化点モデルを仮定する。つまり、治療効果の発現時期 ν までは、対照群と同様にハザードが $\gamma\lambda_0 t^{\gamma-1}$ の分布に従い、 ν 以降はハザードが $\gamma\lambda_1 t^{\gamma-1}$ の分布に従う。確率密度関数、生存関数及びハザード関数は、それぞれ以下の式で表される。また、 ν 以降のハザード比（対照群 / 試験群）は $HR = \lambda_0 / \lambda_1$ で表される。

$$\begin{array}{ll} t \leq \nu \text{ のとき} & t > \nu \text{ のとき} \\ \left\{ \begin{array}{l} f(t) = \gamma \lambda_0 t^{\gamma-1} \exp(-\lambda_0 t^\gamma) \\ S(t) = \exp(-\lambda_0 t^\gamma) \\ h(t) = \gamma \lambda_0 t^{\gamma-1} \end{array} \right. & \left\{ \begin{array}{l} f(t) = \exp(-(\lambda_0 - \lambda_1) * \nu^\gamma) * \gamma \lambda_1 t^{\gamma-1} \exp(-\lambda_1 t^\gamma) \\ S(t) = \exp(-(\lambda_0 - \lambda_1) * \nu^\gamma) * \exp(-\lambda_1 t^\gamma) \\ h(t) = \gamma \lambda_1 t^{\gamma-1} \end{array} \right. \end{array}$$

3.2 条件付き検出力に基づく確率打ち切り法

3.2.1 シミュレーションに基づく条件付き検出力

中間解析データに対して群ごとに分布（第3.1節）を当てはめ、分布のパラメータを推定する。推定した分布から、各症例に対して最終解析時点で観察されるはずの生存期間データを発生する。この完全なデータに対して重みを $1 - S_{\text{urv}}(t)$ とした重み付きログランク検定を行い、有意差の有無を判定する。これらのステップを1000回繰り返し、その内有意差が付いた回数の割合を条件付き検出力と定義する。

3.2.2 漸近論に基づく解析的な条件付き検出力

暦時間 t では、以下のスコア関数を用いて、生存関数の群間比較を行う。

$$U(t) = \sum_{i=1}^n \int_0^t Q(t, x) \{Z_i - E_Z(t, x)\} N_i(t, dx)$$

ただし、 x は登録後の経過時間、 $Q(t, x)$ は重みの推定値、 Z は群変数（ $Z = 0, 1$ ）、 $N_i(t, dx)$ は計数過程とする。 $U(t)$ より構築した検定統計量を $S(t)$ とすると、 $S(t)$ を与えた下で $S(t)$ が従う条件付き分布は漸近的に平均 $\mu(t)$ 、分散 $\sigma^2(t)$ の正規分布になるので、 $S(t) > c$ となる条件付き確率（条件付き検出力）は下式で表される⁴⁾。

$$\rho(t) = 1 - \Phi\left\{\frac{c - \mu(t)}{\sigma(t)}\right\}$$

ただし、 τ は最終解析の暦時間、 c は最終解析の棄却限界値、 $\Phi(x)$ は標準正規分布の累積分布関数とする。 $U(t)$ は独立増分構造でないため、ブラウン運動の確率過程に収束しないが、少なくともガウス過程に収束するため、条件付き分布の平均と分散は、下式のとおり、様々な統計量の関数で一意的に表される。

$$\mu(t) = f_1\{q_0(x), q(\tau, x)\} = f_1\{q_0(x), r(t), s_{\text{urv}0}(x), s_{\text{urv}1}(x), d_0(x), d_1(x)\}$$

$$\sigma^2(t) = f_2\{q_0(x), q(\tau, x)\} = f_2\{q_0(x), r(t), s_{\text{urv}0}(x), s_{\text{urv}1}(x), d_0(x), d_1(x)\}$$

ただし、 $f_1(\cdot), f_2(\cdot)$ は \cdot で規定される関数、 $q_0(x)$ は真の重み関数、 $q(\tau, x)$ は $Q(\tau, x)$ の収束値、 $r(t)$ は症例登録関数、 $s_{\text{urv}0}(x), s_{\text{urv}1}(x)$ は各群の生存関数、 $d_0(x), d_1(x)$ は各群の打ち切り関数とする。実際のデータ解析では、これらの統計量は一般に未知であるため、中間解析データよりパラメトリックに推定する。

3.2.3 重みを調整した重み付きログランク検定統計量

ブラウン運動の性質を満たすように、以下のスコア関数を定義する⁴⁾。

$$U_{\tau}(t) = \sum_{i=1}^n \int_0^t q(\tau, x) \{Z_i - E_Z(t, x)\} N_i(t, dx)$$

$U_{\tau}(t)$ は $U(t)$ の $Q(t, x)$ を $q(\tau, x)$ で置き換えたスコア関数である。 $q(\tau, x)$ は暦時間 t には依存しないため、 $U_{\tau}(t)$ は独立増分構造となり、帰無仮説下ではブラウン運動の確率過程に収束する。一方、対立仮説下では独立増分構造のガウス過程に収束する。

3.3 α 消費関数法に基づく群逐次法

時点 k の情報分数を t_k 、基準化された検定統計量を $B(t_k)$ 、棄却限界値を $b(t_k)$ 、 α 消費関数を $\alpha(t)$ とすると、 t_1, \dots, t_k で条件付けた下、以下の式を解くことで $b(t_k)$ を算出する³⁾。

$$\Pr(B(t_1) < b(t_1), \dots, B(t_{k-1}) < b(t_{k-1}), B(t_k) \geq b(t_k)) = \alpha(t_k) - \alpha(t_{k-1})$$

この方法では、群間比較に用いるスコア関数がブラウン運動の確率過程に収束することを前提とする。従って、重みが暦時間に依存しない場合（第3.2.3節）、第I種の過誤確率の正確な調整が可能になる。一方、重みが暦時間に依存する場合、第I種の過誤確率が增大する可能性がある。

4 シミュレーション研究

4.1 条件付き検出力に基づく確率打ち切り法

解析的な条件付き検出力を検定手法間で比較した結果を図1に示す。効果発現時期に遅れがある場合、重み付きログランク検定は最適ログランク検定（効果発現時期以降のデータに対するログランク検定を表す。効果発現時期が既知である場合、理論上最も検出力が高くなる検定手法）と同等の条件付き検出力を確保することができた。一方、通常のログランク検定の条件付き検出力は大幅に低下した。シミュレーションに基づく条件付き検出力を検定手法間で比較した場合も全く同様の傾向を示した（シミュレーションに基づく条件付き検出力は割愛した）。

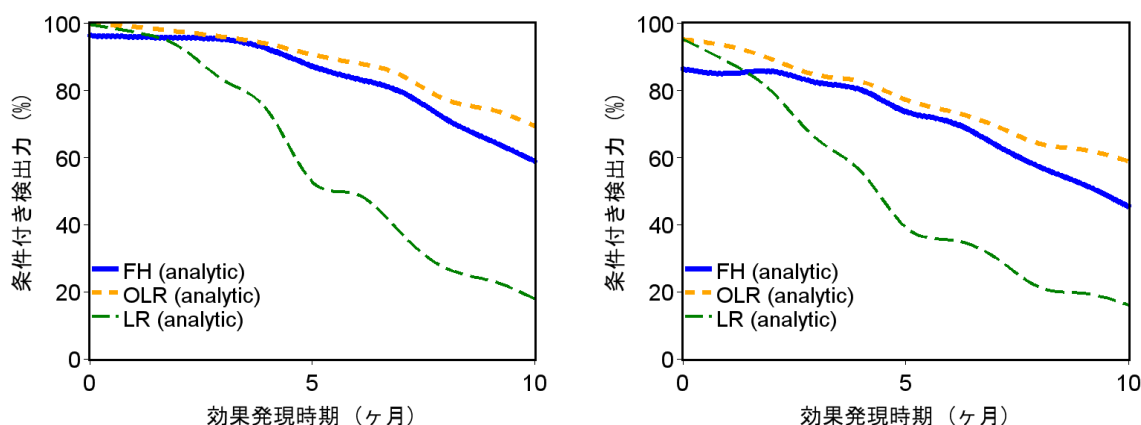


図1. 漸近論に基づく解析的な条件付き検出力の比較（片群100例、中間解析時点のイベント数: 50例、症例登録速度: 5例 / 月）

左図: 効果発現時期以降のハザード比が2.5の場合 右図: 効果発現時期以降のハザード比が2の場合

FH: 重み付きログランク検定、OLR: 最適ログランク検定、LR: ログランク検定

4.2 α 消費関数法に基づく群逐次法

試験全体の第I種の過誤確率を定量的に評価した結果を表1に示す。調整した重みを用いると、情報分数IIは共分散の2乗に一致し、第I種の過誤確率は0.025となった（共分散と第I種の過誤確率は割愛した）。情報分数Iと情報分数IIの乖離の程度は小さかった。無調整の重みを用いると、一斉登録の場合、情報分数IIIと共分散の2乗

は一致した。一斉登録でない場合、情報分数 III と共分散の 2 乗には乖離が見られ、症例登録速度が遅いほど乖離の程度は大きくなった。ただし、どの状況設定の下でも、第 I 種の過誤確率の増大は殆ど招かなかった。

表 1. 重み付きログランク検定に対して α 消費関数法 (O'Brien-Fleming 型) を適用した際の第 I 種の過誤確率

症例数 (片群)	中間解析時点		症例登録 速度 (例/月)	情報分数	情報分数	情報分数	共分散 の 2 乗 *4	第 I 種の 過誤確率 *5
	イベント数	イベント 発現割合		I *1	II *2	III *3		
100	50	0.25	1	0.222	0.219	0.227	0.219	0.02500
100	50	0.25	5	0.141	0.139	0.147	0.139	0.02500
100	50	0.25	50	0.026	0.026	0.028	0.026	0.02500
100	50	0.25	100	0.022	0.022	0.023	0.022	0.02500
100	50	0.25	200	0.021	0.021	0.022	0.022	0.02500
100	100	0.50	1	0.476	0.474	0.480	0.473	0.02501
100	100	0.50	5	0.398	0.395	0.403	0.396	0.02500
100	100	0.50	50	0.173	0.171	0.175	0.173	0.02500
100	100	0.50	100	0.169	0.168	0.171	0.170	0.02500
100	100	0.50	200	0.170	0.168	0.171	0.171	0.02500

*1: 重みを調整した重み付きログランク検定統計量から推定した情報分数 (推定した重みを用いた場合)

*2: 重みを調整した重み付きログランク検定統計量から推定した情報分数 (真の重みを用いた場合)

*3: 無調整の重み付きログランク検定統計量から推定した情報分数

*4: 無調整の重み付きログランク検定統計量から推定した共分散

*5: 無調整の重み付きログランク検定統計量を用いた場合の第 I 種の過誤確率

試験終了時の打ち切り症例の発生割合は約 10%

5 考察

本研究では、重み付きログランク検定に基づく条件付き検出力の算出方法を提案した。効果発現に遅れがある場合、提案した方法の利用が推奨される (図 1)。ただし、提案した方法は、変化点の推定精度に依存する。本シミュレーション研究では、効果発現時期の推定値の不確実性に対して頑健であることを示したが、中間解析時の情報量が少ない場合、効果発現時期の区間推定を行い、効果発現時期の 95%信頼下限値及び上限値に基づく条件付き検出力を補足的に算出すべきである。効果発現の遅れを考慮したベイズ流の予測検出力の利用も有用である。また、中間解析時のイベント数が十分にあって、効果発現時期以降のイベント数が少なければ、誤った意思決定を行う可能性があるため、中間解析実施前に盲検下で生存関数の形状や症例登録速度の確認を行うべきである。

本研究では、重み付きログランク検定に対する α 消費関数法の適用方法を提案した。一般的な状況下では、無調整で α 消費関数法を適用しても、実用上問題ないことを示した (表 1)。ただし、実際に無調整で α 消費関数法を適用する場合、試験開始時に本研究で行ったような事前検討を行う必要がある。また、事後的に試験全体の第 I 種の過誤確率の確認を行い、実際の影響度を解析報告書等に明記すべきである。

6 結論

本研究では、staggered entry 下で、重みを $1 - S_{\text{inv}}(t)$ とした重み付きログランク検定に基づく確率打ち切り法と群逐次法の提案を行った。提案した一連の手法は、中間解析を伴うがんワクチンの臨床試験を実施するための手段として、大変有益である。

参考文献

- 1) Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika* 1982; 69:553-566.
- 2) Lan KKG, Wittes J. The B-value: a tool for monitoring data. *Biometrics* 1988; 44:579-585.
- 3) Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; 70:659-663.
- 4) Lin DY, Yao Q, Ying Z. A general theory on stochastic curtailment for censored survival data. *Journal of the American Statistical Association* 1999; 94:510-521.