

論文の内容の要旨

論文題目

ブートストラップ・リサンプリング法を適用したゲートキーピングプロシジャによる多重性の調整

坂巻 顕太郎

1. 序文

臨床試験では単独で治療効果を決定付けるプライマリエンドポイントと副次的だが重要な効果指標のセカンダリエンドポイントを用いて効果の有無を判断する。判断は各エンドポイントの検定結果に基づいて行われるが、効果のないエンドポイントに対して効果ありと判断する誤り（タイプ I エラー）を抑える必要がある。複数のエンドポイントを用いる場合、いずれかの検定が有意となれば効果ありと判断することでタイプ I エラーが 1 回以上起こる確率（FWER : Family Wise Error Rate）が上昇する検定の多重性の問題を考慮しなければいけない。その際、プライマリエンドポイントがセカンダリエンドポイントより先に検定されるようエンドポイントの重要度（仮説の順序性）を考慮した多重性の調整方法が必要となる。この要求に応える検定手順にゲートキーピングプロシジャがある。

既存のゲートキーピングプロシジャには 2 つの問題がある。1 つは閉検定手順に基づくことで、M 個のエンドポイントに対して $2^M - 1$ 回の検定を行う煩雑さ、仮説の順序を考慮していない不要な検定による検出力の低下が問題となる。この問題を解消する検定手順は提案されていない。もう 1 つは検出力の向上のために用いるリサンプリング法の適用で、並べ替え法はサブセットピボタリティと同時交換可能性の仮定の妥当性、ブートストラップ法は検定への利用の困難さが問題となる。近年、並べ替え法の仮定を必要としないブートストラップ法を用いた検定方法が提案されたがゲートキーピングプロシジャに適用したものはまだない。

そこで本研究は、仮説の順序性を適切に考慮したゲートキーピングプロシジャの検定手順と各手順の検定にブートストラップ法を適用することを提案する。

2. 方法

ゲートキーピングプロシジャの 1 つであるパラレルゲートキーピングプロシジャは、重要度で区分した仮説の集合（ファミリー）のうち、重要度の高いファミリー 1 の帰無仮説を 1 つでも棄却したら、ファミリー 2 の帰無仮説の検定を行う手順である。ただしファミリー 1 の帰無仮説は以下の条件により検定結果が変わる。

- ・条件 A : ファミリー 2 に含まれる帰無仮説の検定結果に依存する。
- ・条件 B : ファミリー 2 に含まれる帰無仮説の検定結果に依存しない。

条件 B ではプライマリエンドポイントの検定が他の検定に先んずるため、望ましい手順となる。

ここでプライマリエンドポイントの帰無仮説 $\{H_1, \dots, H_m\}$ をファミリー 1、セカンダリエンドポイントの帰無仮説 $\{H_{m+1}, \dots, H_M\}$ をファミリー 2 とする。 H_i の p 値 $p(H_i)$ と仮説の重要度に比例して大きくなる重み $v(H_i)$ を用いた重み付き p 値を $p_v(H_i) = p(H_i)/v(H_i)$, $i = 1, \dots, M$ とする。ファミリー 1 の帰無仮説を重み付き p 値が小さい順に O_{11}, \dots, O_{1m} 、ファミリー 2 も同様に O_{21}, \dots, O_{2M-m} とする。このとき提案する検定手順は重み付き p 値の真の同時帰無分布 Q_{v_0} の下で条件 A では A1-A2、条件 B では B1-B3 となる。

A1 : ファミリー 1 に含まれる帰無仮説に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \{O_{11}, \dots, O_{1m}\}) \leq p_v(O_{11})) \leq \alpha$$

となれば O_{11} を棄却する。 $\{O_{12}, \dots, O_{1m}, O_{21}, \dots, O_{2M-m}\}$ を重み付き p 値が小さい帰無仮説から順に $\bar{O}_{21}, \dots, \bar{O}_{2M-1}$ 、 i 番目以降の帰無仮説の集合を $\bar{O}_{2(i)} = \{\bar{O}_{2i}, \dots, \bar{O}_{2M-1}\}$ とし、A2 に進む。 O_{11} を棄却しなければ検定を終える。

A2 : $1 \leq h \leq M-1$ において、 $\bar{O}_{2(h)} = \{\bar{O}_{2h}, \dots, \bar{O}_{2M-1}\}$ に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \bar{O}_{2(h)}) \leq p_v(\bar{O}_{2h})) \leq \alpha$$

となれば \bar{O}_{2h} を棄却する。 $h < M-1$ のとき A2 を繰り返す。 $h = M-1$ のとき検定を終える。 \bar{O}_{2h} を棄却しなければ検定を終える。

B1 : ファミリー1に含まれる帰無仮説に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \{O_{11}, \dots, O_{1m}\}) \leq p_v(O_{11})) \leq \alpha$$

となれば O_{11} を棄却する。 $O_{12}, \dots, O_{1m}, O_{21}, \dots, O_{2M-m}$ と並べた帰無仮説を順に $\tilde{O}_{21}, \dots, \tilde{O}_{2m-1}, \tilde{O}_{2m}, \dots, \tilde{O}_{2M-1}$ 、 i 番目以降の帰無仮説の集合を $\tilde{O}_{2(i)} = \{\tilde{O}_{2i}, \dots, \tilde{O}_{2M-1}\}$ とし、B2 に進む。 O_{11} を棄却しなければ検定を終える。

B2 : $1 \leq h \leq M-1$ において、 $\tilde{O}_{2(h)} = \{\tilde{O}_{2h}, \dots, \tilde{O}_{2M-1}\}$ に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \tilde{O}_{2(h)}) \leq p_v(\tilde{O}_{2h})) \leq \alpha$$

となれば \tilde{O}_{2h} を棄却する。 $h < M-1$ のとき B2 を繰り返す。 $h = M-1$ のとき検定を終える。 \tilde{O}_{2h} を棄却しないとき、 $\{\tilde{O}_{2h}, \dots, \tilde{O}_{2M-1}\}$ を重み付き p 値が小さい帰無仮説から順に $\check{O}_{21}, \dots, \check{O}_{2M-h}$ 、 i 番目以降の帰無仮説の集合を $\check{O}_{2(i)} = \{\check{O}_{2i}, \dots, \check{O}_{2M-h}\}$ とし、B3 に進む。

B3 : $1 \leq h' \leq M-h$ において、 $\check{O}_{2(h')} = \{\check{O}_{2h'}, \dots, \check{O}_{2M-h}\}$ に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \check{O}_{2(h')}) \leq p_v(\check{O}_{2h'})) \leq \alpha$$

となれば帰無仮説 $\check{O}_{2h'}$ を棄却する。 $h' < M-h$ のとき B3 を繰り返す。 $h' = M-h$ のとき検定を終える。 $\check{O}_{2h'}$ が棄却されなければ検定を終える。ただし、 $\check{O}_{2h'}$ がファミリー1の帰無仮説である場合、検定終了時に $\check{O}_{2h'}$ は棄却されなかったものとして扱う。

一般に未知である Q_{v_0} は推定する必要がある。提案するブートストラップ法の適用による推定は以下である。

群1、2から復元抽出したデータ $Y_1^b = \{Y_{1j}^b: j = 1, \dots, n_1\}$ 、 $Y_2^b = \{Y_{2j}^b: j = 1, \dots, n_2\}$ を $Y^b = \{Y_1^b, Y_2^b\}$ とし、B個のブートストラップサンプルを作成する。サンプルから各帰無仮説の重み付き p 値 $p_{vb}^B = \{p_{vb}^B(H_i): i = 1, \dots, M, b = 1, \dots, B\}$ を計算し、帰無仮説ごとに経験累積分布関数

$$Q_{vi}^B(z) = \frac{1}{B} \sum_{b=1}^B I(p_{vb}^B(H_i) \leq z)$$

を求める。重み付き p 値を変換した統計量 $\check{p}_{vb}^B(H_i) = q_{v0i}^{-1} Q_{vi}^B\{p_{vb}^B(H_i)\}$ を全サンプルで計算し、重み付き p 値の同時帰無分布を推定する。ただし、 q_{v0i} は $p_v(H_i)$ の真の周辺帰無分布である。

3. シミュレーション実験

3.1. 設定

エンドポイントの型、エンドポイント間の相関、治療効果の有無の3つが違う状況で、提案法が FWER を名義水準以下に保ち、検出力を上昇させるかをシミュレーション実験によって検討した。2群比較、各群100人、2つのプライマリエンドポイント、2つのセカンダリエンドポイントを用いる試験を想定した。各エンドポイントは全て正規分布、全て2項分布、2項分布と正規分布の組み合わせの3通りで発生させた。相関はピアソンの相関係数で定義し、2群で等しい場合は0、0.2、0.5、0.8、違う場合は

試験治療群と標準治療群で (0, 0.8)、(0.8, 0)となるようデータを発生させた。治療効果は全てのエンドポイントにおいて無しまたは有りの2通りとした。検定はt検定かFisherの直接確率検定を用い、片側検定を行った。各設定でのシミュレーションの繰り返し回数は2000回とした。

提案したパラレルゲートキーピングプロシジャにブートストラップ法を適用した場合の性能を評価した。ブートストラップ法の周辺分布は一様分布か並べ替え分布を利用し、リサンプリング回数を10000回とし同時帰無分布を推定した。性能の評価には、治療効果無しの場合はFWER、治療効果有りの場合は検出力を用いた。FWERは2.5%を名義水準とした。検出力は、各帰無仮説を棄却する確率、全て(任意)の誤った帰無仮説の棄却確率、全て(任意)のプライマリエンドポイントの誤った帰無仮説の棄却確率を用いた。比較のため、多重性を調整しない方法、ボンフェローニ法、閉検定手順に基づく既存の方法と提案したパラレルゲートキーピングプロシジャにボンフェローニ法と並べ替え法を適用した方法の性能も検討した。

3.2. 結果

全て正規分布、全て2項分布の2通りに関して治療効果無しの設定のFWERを表1に示す。パラレルゲートキーピングプロシジャは条件で結果が変わらなかったため条件Aを示す。提案手順において、ブートストラップ法に一様分布を用いた場合は正規分布、並べ替え分布を用いた場合は全ての設定で名義水準付近となった。閉検定手順、提案手順に並べ替え法を適用した場合も全ての設定で名義水準付近となった。その他の方法は過大か過小となり性能が低下した。2項分布の設定で過小となる傾向が顕著に見られた。相関が違ふ設定においても並べ替え法を適用した場合のFWERは保たれていた。

同様に治療効果有りの設定の検出力を表2に示す。パラレルゲートキーピングプロシジャは条件Bの結果を示す。閉検定手順と比較して、提案手順は全エンドポイント、全プライマリエンドポイントの棄却確率が5%程度高かった。並べ替え法を適用した場合は相関0.8の2項分布の設定において任意のプライマリエンドポイントの棄却確率が提案手順でわずかに高かった。提案手順において適用した検定方法を比較すると、相関0.8の設定、2項分布の設定では全エンドポイント、全プライマリエンドポイントの棄却確率がリサンプリング法で3-5%高かった。また、ブートストラップ法に一様分布を用いた場合は正規分布、並べ替え分布を用いた場合は全ての設定で検出力が高かった。その他の結果は表1、2から類推される結果であった。

4. 考察

仮説の順序性を適切に考慮したゲートキーピングプロシジャの検定手順と各手順の検定にブートストラップ法を適用することを提案した。提案法は3つの点で有用な方法であると考えられる。1つ目はエンドポイントの重要度の設定など臨床家の意見を必要とする場面で有効となる手順の明確さ、簡潔性である。2つ目は臨床試験の第一目的であるプライマリエンドポイントの棄却確率の向上である。これは提案した手順とリサンプリング法の適用が影響する。3つ目は2群間の不等分散性、多群比較などサブセットピボタリティが成立しない場合にも実装可能なブートストラップ法の柔軟性である。これらの特徴を持つ提案法は臨床試験に用いる際には有用な方法であると考えられる。

提案法がFWERを名義水準以下に保ち、検出力を向上させるかをシミュレーション実験によって評価した。FWERは全ての設定で名義水準以下となった。2群で相関が違ふ場合に並べ替え法とブートストラップ法の違いが見られなかったのは、エンドポイントが2つずつでは相関の影響が大きくないことが理由として考えられる。検出力は提案法でプライマリエンドポイントに関する棄却確率がより

高い値となった。その理由に、閉検定手順がセカンダリエンドポイントに過度な重みを与え、仮説の順序性を適切に考慮していないことが考えられる。これらのことから提案法の有用性が示唆される。また、リサンプリング法の適用により検出力は向上した。本研究の設定において、並べ替え法は仮定を満たす適切な検定方法であるため、ブートストラップ法が同等の検出力を示したことは周辺分布が特定されている場合の適用が有効であることを示唆する。

シミュレーション実験から提案するブートストラップ法を適用したゲートキーピングプロシジャは FWER を名義水準以下に保ち検出力を向上することがわかった。特に、相関が高く 2 項分布の設定では検出力の向上が顕著であった。

表 1 FWER

	相関 (試験標準)	シングルステップ		閉検定手順		提案手順			
		Un	Bon	Bon	Per	Bon	Per	Boot1	Boot2
正規分布	(0,0)	8.4(4.5)	2.0(1.6)	2.3	2.3	2.3	2.3	2.3	2.3
	(0.8,0.8)	4.5(3.4)	1.7(1.7)	1.9	2.4	1.9	2.4	2.3	2.3
	(0,0.8)	7.4(4.6)	2.4(1.9)	2.3	2.5	2.3	2.5	2.4	2.5
	(0.8,0)	7.4(4.5)	2.0(1.7)	2.0	2.1	2.0	2.1	2.1	2.1
2項分布	(0,0)	6.7(3.5)	1.4(1.3)	1.6	2.0	1.6	2.0	1.6	2.0
	(0.8,0.8)	4.6(3.1)	1.5(1.4)	1.6	2.5	1.6	2.5	1.9	2.5
	(0,0.8)	6.0(3.6)	1.6(1.4)	1.8	2.1	1.8	2.5	1.9	2.5
	(0.8,0)	6.6(3.2)	1.8(1.6)	1.8	2.1	1.8	2.4	1.9	2.4

括弧内: プライマリエンドポイントに関する帰無仮説を誤って棄却する確率

Un: 多重性を調整しない方法, Bon: ボンフェローニ法

Per: 並べ替え法, Boot1: ブートストラップ法(一様分布), Boot2: ブートストラップ法(並べ替え分布)

表 2 検出力 (条件 B)

	相関	シングルステップ		閉検定手順		提案手順				
		Un	Bon	Bon	Per	Bon	Per	Boot1	Boot2	
正規分布	0	All	41.1	11.2	31.4	31.1	35.2	35.2	35.4	35.4
		All Primary	63.3	45.0	49.0	48.9	55.4	55.3	55.6	55.4
		Any	99.8	97.0	90.8	90.6	90.8	90.6	90.9	90.7
		(Any Primary)	(95.8)	(88.8)						
0.8	All	66.0	37.4	57.9	61.0	61.3	64.0	64.2	64.1	
	All Primary	72.8	58.2	61.9	65.5	66.5	70.2	70.4	70.2	
	Any	91.1	78.6	79.2	81.2	79.2	81.2	81.5	81.2	
	(Any Primary)	(86.4)	(77.1)							
2項分布	0	All	40.9	11.5	32.4	33.8	35.4	38.0	35.4	38.0
		All Primary	62.7	46.1	50.4	52.2	55.2	58.4	55.2	58.3
		Any	99.9	97.4	91.0	93.2	91.0	93.2	91.0	93.2
		(Any Primary)	(95.5)	(89.5)						
0.8	All	66.5	37.4	58.9	62.8	61.8	65.8	64.9	65.8	
	All Primary	72.6	58.7	62.4	67.3	66.3	71.7	70.2	71.5	
	Any	91.4	79.2	79.6	83.9	79.6	84.7	82.7	84.2	
	(Any Primary)	(87.2)	(77.4)							

All: 全エンドポイントの棄却確率, All Primary: 全プライマリエンドポイントの棄却確率

Any: 任意のエンドポイントの棄却確率, Any Primary: 任意のプライマリエンドポイントの棄却確率

Un: 多重性を調整しない方法, Bon: ボンフェローニ法

Per: 並べ替え法, Boot1: ブートストラップ法(一様分布), Boot2: ブートストラップ法(並べ替え分布)