

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

目次

要旨	4
1. 序文.....	5
1.1. 複合エンドポイントと OLS・GLS 検定.....	8
1.2. エンドポイントごとの検定と多重性の調整.....	10
1.3. 仮説の順序性を考慮した多重性の調整.....	13
1.4. ゲートキーピングプロシジャの問題点.....	15
1.5. 研究の目的.....	17
2. 方法.....	18
2.1. 想定する状況.....	18
2.2. 多重性の調整方法.....	19
2.2.1. シングルステップ法.....	19
2.2.2. 閉検定手順.....	20
2.2.3. 閉検定手順を用いたゲートキーピングプロシジャ.....	22
2.2.3.1. シリアルゲートキーピングプロシジャ.....	22
2.2.3.2. パラレルゲートキーピングプロシジャ.....	25
2.2.3.3. ボンフェローニ法を適用したゲートキーピングプロシジャ.....	27
2.2.3.4. 並べ替え法を適用したゲートキーピングプロシジャ.....	31
2.2.3.5. 並べ替え法を適用したゲートキーピングプロシジャの問題点.....	32
2.3. 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.....	33
2.3.1. 提案するゲートキーピングプロシジャ.....	33
2.3.1.1. シリアルゲートキーピングプロシジャ.....	34
2.3.1.2. パラレルゲートキーピングプロシジャ.....	37
2.3.2. 周辺分布を用いたブートストラップ法による同時帰無分布の推定.....	41
3. シミュレーション実験.....	42
3.1. 状況設定.....	43
3.2. 評価方法.....	46
3.3. シミュレーション実験の結果.....	47

4. 考察.....	78
5. 結論.....	85
6. 謝辞.....	86
文献.....	86
附録.....	98

図表

表 1 シリアルゲートキーピングプロシジャの決定行列.....	29
表 2 パラレルゲートキーピングプロシジャ (条件 A) の決定行列.....	30
表 3 パラレルゲートキーピングプロシジャ (条件 B) の決定行列.....	31
表 4 共通帰無仮説の重み.....	45
表 5 FWER (相関構造が等しい場合)	47
表 6 FWER (相関構造が違う場合)	47
表 7 仮説ごとの検出力 (連続変数、条件 A)	49
表 8 多重性を考慮した検出力 (連続変数、条件 A)	50
表 9 仮説ごとの検出力 (連続変数、条件 B)	51
表 10 多重性を考慮した検出力 (連続変数、条件 B)	52
表 11 仮説ごとの検出力 (2 値変数、条件 A)	59
表 12 多重性を考慮した検出力 (2 値変数、条件 A)	60
表 13 仮説ごとの検出力 (2 値変数、条件 B)	61
表 14 多重性を考慮した検出力 (2 値変数、条件 B)	62
表 15 仮説ごとの検出力 (2 値変数と連続変数、条件 A)	69
表 16 多重性を考慮した検出力 (2 値変数と連続変数、条件 A)	70
表 17 仮説ごとの検出力 (2 値変数と連続変数、条件 B)	71
表 18 多重性を考慮した検出力 (2 値変数と連続変数、条件 B)	72
図 1 閉じた仮説族.....	21
図 2 シリアルゲートキーピングプロシジャ.....	23
図 3 ファミリー1に含まれる帰無仮説の検定に関する共通帰無仮説.....	24
図 4 ファミリー2に含まれる帰無仮説の検定に関する共通帰無仮説.....	24
図 5 パラレルゲートキーピングプロシジャ.....	26
図 6 ファミリー1に含まれる帰無仮説の検定に関する共通帰無仮説.....	26
図 7 ファミリー2に含まれる帰無仮説の検定に関する共通帰無仮説.....	27
図 8 仮説族ごとの検出力 (連続変数、条件 A)	53
図 9 多重性を考慮した検出力 (連続変数、条件 A)	54
図 10 仮説族ごとの検出力 (連続変数、条件 B)	55

図 11	多重性を考慮した検出力（連続変数、条件 B）	56
図 12	仮説族ごとの検出力（2 値変数、条件 A）	63
図 13	多重性を考慮した検出力（2 値変数、条件 A）	64
図 14	仮説族ごとの検出力（2 値変数、条件 B）	65
図 15	多重性を考慮した検出力（2 値変数、条件 B）	66
図 16	仮説族ごとの検出力（2 値変数と連続変数、条件 A）	73
図 17	多重性を考慮した検出力（2 値変数と連続変数、条件 A）	74
図 18	仮説族ごとの検出力（2 値変数と連続変数、条件 B）	75
図 19	多重性を考慮した検出力（2 値変数と連続変数、条件 B）	76

要旨

複数のプライマリエンドポイントとセカンダリエンドポイントを用いて治療効果を評価する際、エンドポイントの重要度を考慮して多重性を調整することが重要である。その中でも仮定が少なく検出力がより高い方法に関心があることが多い。本研究では、仮説の順序性を適切に考慮した検定手順となるゲートキーピングプロシジャ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。シミュレーション実験により提案法が **FWER** を名義水準以下に保ち、既存の方法に比べて検出力が高い方法であることを示した。

1. 序文

臨床試験では治療効果を評価するために複数のエンドポイントが設定される。急性呼吸窮迫症候群（ARDS: Acute Respiratory Distress Syndrome）を対象とする臨床試験の例では、死亡率、人工呼吸器の使用期間、集中治療室の使用期間、Quality of Life (QOL) の4つのエンドポイントを設定することが考えられる¹。ARDSは臨床的に重症状態の患者に突然起こる呼吸不全で、人工呼吸器の装着が余儀なく、死亡率が40%と予後の悪い疾患である²。そのため、臨床的に十分な治療効果があるというエビデンスを与える、死亡率、人工呼吸器の使用期間の2つがプライマリエンドポイントとなる。また、治療効果の副次的な側面を表す、集中治療室の使用期間、QOLはセカンダリエンドポイントとなる。このように、複数のプライマリエンドポイント、複数のセカンダリエンドポイントを用いて治療効果を評価する場合がある。

複数のプライマリエンドポイントを設定する目的は2つに分けられる。1つ目はARDSのように疾患の特徴が多面的である場合である。特に、アルツハイマー病³、リウマチ⁴、関節炎⁵、神経痛⁶、疼痛⁷⁻¹¹などの疾患は、症状改善といったソフトなエンドポイント1つで治療効果を十分に定義しきれないため、複数のエンドポイントによる治療効果の評価が求められる。アルツハイマー病では、認知機能、臨床評価、ADL (Activities of Daily Life) などにより治療効果を評価する³。2つ目は複数の疾患に対して同時に評価を行う場合である。例えば、循環器疾患として脳卒中、心筋梗塞と複数の部位を評価の対象とすることがある¹²。また、複数のセカンダリエンドポイ

ントを設定する目的は、治療効果の副次的な側面、疾患に対する作用機序などを多面的に評価することである。アルツハイマー病に対する治療効果であれば、介護者負担の減少などを副次的に評価する¹³。スタチンの治療効果であれば、コレステロールの血中濃度の低下以外にも血管保護作用など様々な作用機序が評価される¹⁴。

一方で、一般の臨床試験でのプライマリエンドポイントは1つであることが多い。これはプライマリエンドポイントとセカンダリエンドポイントの特徴¹⁵⁻¹⁷に起因している。プライマリエンドポイントは、治療効果を明瞭に定義し、臨床的な意思決定に基づいて評価可能な変数が選択される。治療効果の有無の決定もしくは決定するには情報不足かの判断に利用し、研究のサンプルサイズや検出力を決めるものである¹⁶。セカンダリエンドポイントは、治療効果を表す重要なものではあるが単独では十分なエビデンスを示せない変数である。主に、プライマリエンドポイントの理解を高めるもの、プライマリエンドポイントとしては適切な検出力を担保できないもの、治療のメカニズムを説明するもの、複合エンドポイントの要素などが選択される¹⁷。他の治療との差別化を図るためにも利用される。これらの特徴から、ICH (International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use) ガイドラインにおいて、治療効果を確実に説明するプライマリエンドポイントを1つ選択し、治療効果の有無を判断することを推奨している¹⁶。その理由に検定の多重性の問題がある。

検定の多重性は、複数回の検定を行うことによって治療効果ありと誤って判断する

確率が個々の検定の有意水準より大きくなることをいう¹⁸。例えば、エンドポイントに関する検定が1回でも有意となれば治療効果ありと判断する場合、治療効果のない互いに独立な4つのエンドポイントに対して有意水準5%で検定を繰り返すと、4回の検定のうち1回以上タイプIエラーが起こる確率（FWER；FamilyWise Error Rate）は18.5%となる。タイプIエラーとは治療効果なしという仮説（帰無仮説）が正しいにもかかわらず誤って帰無仮説を棄却して治療効果ありと判断するエラーのことである。このような問題を避けるため、1つのプライマリエンドポイントに関する検定により治療効果の有無を判断する。セカンダリエンドポイントに関する検定の多重性を考慮しない理由は、プライマリエンドポイントの結果の解釈や将来の研究課題として利用されるため、多重性を考慮する必要がないとされているからである^{17,19-22}。

しかし、医薬品の開発費が高騰している現状²³において、人的、時間的、金銭的なコストが掛かる臨床試験で検証的な結果が1つのみであるのは効率が悪い。そのため、ARDSの例のように複数のプライマリエンドポイントによって治療効果を評価する臨床試験が今後増えていくことが考えられる。また、規制当局の要求に対して新たな知見や付加的な知見を与えるセカンダリエンドポイントは重要な治療効果を表す²⁴ため、効果・効能のラベリングに寄与する結果として扱われる機会が増えると考えられる。一例としてPRO（Patient Reported Outcome）やQOLが挙げられる²⁵。複数のプライマリエンドポイントとセカンダリエンドポイントが治療効果の評価やラベリングに寄与する問題に対して、Changは事前に設定した検定を全てプロトコルに明示

すれば多重性を考慮する必要はないと述べている^{26,27}。しかし、適切に多重性を考慮しないことで重大な判断の誤りを犯すことがある²⁸ため、複数回の検定を行うことが避けられない場合は多重性を考慮することが求められる^{16,29-32}。また、FDA (Food and Drug Administration) は FWER を名義水準以下に保つ検定方法・手順を用いた結果のみのラベリングを許している³³。そのため、セカンダリエンドポイントを含めて適切に多重性を考慮する検定方法・手順が求められている。

1.1. 複合エンドポイントと OLS・GLS 検定

多重性の問題を避けるために、複数のエンドポイントを1つにまとめた複合エンドポイントを用いることがある^{4,34,35}。脳卒中と心筋梗塞をまとめた心血管疾患¹⁸、ACR (American College of Rheumatology) コアセットの圧痛関節数、腫脹関節数などをまとめた ACR20 が有名である⁴。複合エンドポイントの利用により1回の検定で治療効果の有無を判断するため多重性を考慮する必要性がなくなる。また、発症が稀な疾患では1つのエンドポイントに絞って検定するより検出力が上がる可能性もある¹¹。しかし、構成するエンドポイントのうち1つだけに効果がある、効果のあるエンドポイントと有害な影響のあるエンドポイントが混ざるなどの場合は、適切な結果の解釈が行えないことが指摘されている^{35,36}。また、区別すべきプライマリエンドポイントとセカンダリエンドポイントをまとめることで治療効果を過度に大きく見積もる問題も指摘されている³⁷。

これと同様の問題を持ち合わせた多重性を考慮する方法に、O'Brien が提案した OLS (Ordinary Least Squares) 検定、GLS (Generalized Least Squares) 検定がある³⁸。これらの方法は、正規分布に従うエンドポイントの平均を標準偏差で基準化した治療効果 (エフェクトサイズ) の和を検定統計量に利用する方法である。OLS 検定では全てのエンドポイントについて同じ重み、GLS 検定ではエンドポイント間の相関を考慮した重みを用いてエフェクトサイズの和を求める。検定の帰無仮説は「全てのエンドポイントに対して治療効果がない」であり、仮説の棄却が「いずれかのエンドポイントに対して治療効果がある」となるため、OLS 検定、GLS 検定はグローバル検定である。GLS 検定はエンドポイントの相関構造によって負の重みを用いて検定統計量を構成する可能性があるため、Tang らが拡張法を提案している³⁹。また、連続量だけでなく 2 値変数や生存時間にも用いることができる検定方法が提案されている⁴⁰⁻⁴²。これらは検定統計量を複合する検定方法として解釈できるため、複合エンドポイントと同様のメリット、デメリットがある¹¹。また、エンドポイントごとにエフェクトサイズの大きさや方向性が違うと FWER を名義水準以下に保てないという問題もある⁴³。

セカンダリエンドポイントを含めて治療効果の定義やラベリングを考える際は、個々のエンドポイントに関する検定を行うことが重要である。そのため、複合エンドポイントや OLS 検定などを用いた多重性の調整は有効でない。

1.2. エンドポイントごとの検定と多重性の調整

エンドポイントごとに検定を行う際に利用する多重性の調整方法に、FWER を名義水準以下に保つように調整した有意水準（調整有意水準）を各検定に用いる方法⁴⁴と閉検定原理に基づいた閉検定手順⁴⁵を利用する方法がある。調整有意水準を用いる方法は2つに分けられる。1つはシングルステップ法で、各エンドポイントの検定結果が互いに影響せず、全てのエンドポイントを同時に検定する方法である。各検定で求まる p 値のみを利用するボンフェローニ法⁴⁴、シダック法⁴⁶、帰無仮説のもとでの検定統計量や p 値の同時分布(同時帰無分布)を利用するリサンプリング法⁴⁷がある。もう1つはステップワイズ法で、各エンドポイントの検定結果によって決まる検定の順序からエンドポイントを1つずつ連続的に検定する方法である。ホルム法⁴⁸、ホッフベルグ法⁴⁹、ホムメル法⁵⁰がある。

ボンフェローニ法は、仮説数の逆数を有意水準に掛けて調整有意水準を求める方法である。有意水準に掛ける重みを仮説ごとに変える拡張法も提案されている⁴⁴。この方法は検定統計量や p 値の同時帰無分布の構造によらず FWER が名義水準以下となり、計算が簡便であるためよく用いられる。シダック法は、1回の検定でタイプ I エラーが起こらない確率の積から調整有意水準を求める方法である。FWER が名義水準以下となるのは全てのエンドポイントが独立か多変量正規分布に従うときであり、 t 分布や他の分布ではさらなる条件が必要である。ボンフェローニ法よりも一様に検出力は高いがその違いは大きくない⁵¹。リサンプリング法は、並べ替え法やブートストラッ

プ法により推定した同時帰無分布から調整有意水準を求める方法である。データ生成分布に対する仮定が少なく、同時帰無分布の相関構造や離散性を利用しているため、ボンフェローニ法やシダック法より検出力が高くなる。特に、エンドポイントが多い、エンドポイント間の相関が高いときは性能が良いことが知られている⁵¹。並べ替え法は正確に同時帰無分布を求める方法であるが、サブセットピボタリティ⁴⁷と同時交換可能性⁵²の仮定が必要である。サブセットピボタリティは、全ての帰無仮説が正しい場合でも一部の帰無仮説が正しい場合でも同時帰無分布の構造が変わらないことである。同時交換可能性はエンドポイントの同時分布の構造が群間で等しいことである。これらの仮定が成立していないと FWER は名義水準に保たれない可能性がある。ブートストラップ法は漸近的な性質はよいがサンプル数が少ないときは性能を確認する必要がある¹¹。また、リサンプリング法はコンピュータに対する負荷の大きさ、打ち切りを伴う生存時間など複雑なデータにおけるモデルの確認の必要性などのデメリットが指摘されている⁵³。

ホルム法は、 p 値が小さい帰無仮説から順に検定を行う方法である。ボンフェローニ法と同様、調整有意水準を求める重みを変えた拡張ホルム法も提案されている⁴⁴。同時帰無分布の構造によらず FWER を名義水準以下に保つ方法で、ボンフェローニ法より一様に検出力が高く、計算が簡単であるためよく使われる。ホッフベルグ法は、 p 値が大きい帰無仮説から順に調整有意水準を用いて検定を行う方法である。各仮説の検定はホルム法と対応する調整有意水準を用いる⁵⁴が、検定順序の違いからホルム

法より検出力が高くなる。シムズの不等式により正当化され、エンドポイントが独立か正の相関のときのみ FWER が名義水準以下となる⁵⁵。条件が満たされないときに FWER を小さくする修正ホッフベルグ法も提案されている⁵⁶。シムズの不等式に基づくその他の方法にホムメル法がある。ホムメル法は、ホッフベルグ法より大きい調整有意水準を用いるため検出力がより高くなる。

閉検定手順は、各エンドポイントに関する帰無仮説とそれらの積仮説から定義する仮説の集合により導かれる帰無仮説の構造を利用する検定手順である。積仮説の検定にはグローバル検定を用い、OLS 法や GLS 法、シングルステップ法が利用される^{57,58}。

シングルステップ法は積仮説を定義する帰無仮説に対して用い、いずれかの帰無仮説を棄却すれば積仮説を棄却する。その他のグローバル検定にシムズ法がある⁵⁹。ステップワイズ法は閉検定手順として導くことが可能である⁵¹。

セカンダリエンドポイントを含めて多重性の調整を行うとき、調整有意水準を利用する方法や閉検定手順を用いることの問題は、仮説の順序性を考慮していないことである。つまり、プライマリエンドポイントが棄却されたときにのみセカンダリエンドポイントが意味を持つという順序性を考慮していないということである。シングルステップ法の問題はプライマリエンドポイントとセカンダリエンドポイントの検定を同時に行うため多重性を過度に調整してしまうことである。ステップワイズ法では、プライマリエンドポイントより先にセカンダリエンドポイントが検定される可能性が問題となる。閉検定手順はプライマリエンドポイントとセカンダリエンドポイント

の積仮説の検定でシングルステップ法と同様の問題がある。臨床試験の第一目的は、プライマリエンドポイントの検定が有意であることなので、多重性の調整方法は仮説の順序性を考慮する必要がある⁶⁰⁻⁶⁴。

1.3. 仮説の順序性を考慮した多重性の調整

仮説の順序性を考慮した方法には固定順検定 (Fixed Sequence) 法⁶⁵がある。固定順検定法は事前に決めた順番で帰無仮説の検定を行い、2 番目以降の検定は前の帰無仮説が棄却されたら行う方法である。そのため、プライマリエンドポイントから順に検定を行い、次にセカンダリエンドポイントの検定を行うことが可能となる。しかし、複数のプライマリエンドポイントがある場合、事前に決めた順番が適切でないとプライマリエンドポイントの中で明らかに誤っている帰無仮説が棄却されない可能性がある⁵¹。この問題を解消する検定手順にゲートキーピングプロシジャがある。

ゲートキーピングプロシジャは重要度によりまとめた帰無仮説の集合 (ファミリー)のうち、重要度が高いファミリーの帰無仮説から順に検定を行う方法ある^{1,51,53, 66,67}。ファミリーに含める帰無仮説は、プライマリエンドポイント、セカンダリエンドポイントと重要度によって分けることができる。固定順検定法と違う点はファミリー内の帰無仮説に関して事前に順序性を仮定しないことである。ファミリーに含まれる帰無仮説をゲートキーパーと言い、ゲートキーパーが棄却されるかにより次のファミリーに含まれる帰無仮説の検定を行うかを決める^{1,51,53,66,67} 検定手順である。多重性の調整

には、積仮説の検定に対してゲートキーパーを考慮する閉検定手順が用いられる^{51,53}。

ゲートキーパーの種類によりシリアルゲートキーピングプロシジャ、パラレルゲートキーピングプロシジャに分けられる。2つのゲートキーピングプロシジャの折衷であるツリーゲートキーピングプロシジャ⁶⁸⁻⁷⁰も提案されている。

シリアルゲートキーピングプロシジャは、ファミリーに含まれる全ての帰無仮説を棄却するとき、次のファミリーの検定を行う方法である^{66,67}。ファミリー内の仮説が同様に重要な意味を持ち、全ての検定が有意でないと次のファミリーの検定に意味を持たない場合に用いる⁵¹。ファミリー内の各帰無仮説が単独でゲートキーパーとなる。

ゲートキーパーを考慮する閉検定手順では、プライマリエンドポイントとセカンダリエンドポイントの積仮説に対してプライマリエンドポイントのみの積仮説と同様の検定を行う。この検定方法を用いることで、プライマリエンドポイントの検定がセカンダリエンドポイントの検定結果に依存しないようになる。

パラレルゲートキーピングプロシジャは、ファミリーに含まれるいずれかの帰無仮説を棄却するとき、次のファミリーの検定を行う方法である¹。ファミリー内の全ての検定が有意でなくとも下位のファミリーの検定結果が解釈可能な場合に用いる⁵¹。

ファミリー内の全ての帰無仮説がまとめて1つのゲートキーパーとなる。ゲートキーパーを考慮する閉検定手順では、全プライマリエンドポイントとセカンダリエンドポイントの積仮説に対して全プライマリエンドポイントの積仮説と同様の検定を行う。

また、プライマリエンドポイントの検定を行うかは、条件 A、B のもとでの積仮説の

検定結果により決定する。条件 A では積仮説をどのように検定してもよいが、条件 B ではプライマリエンドポイントの検定統計量や p 値を用いてシングルステップ法により積仮説の検定を行う必要がある。条件 B は、プライマリエンドポイントの検定がセカンダリエンドポイントの検定結果に依存しないため、条件 A より望ましい検定手順となる。Dmitrienko らは積仮説の検定にボンフェローニ法を用いた場合の詳細な検定手順を提案している^{71,72}。

1.4. ゲートキーピングプロシジャの問題点

ゲートキーピングプロシジャを臨床試験に用いる際、検出力の向上から、積仮説の検定にボンフェローニ法よりリサンプリング法を適用することが望ましい。しかし、既存の方法を用いることには2つの問題がある。1つはゲートキーピングプロシジャが閉検定手順を用いている点である。閉検定手順は、M 個のエンドポイントに対して計 $2^M - 1$ 個の仮説の検定を行わなければいけないため、リサンプリング法の計算負荷を含めた手順の煩雑さが問題となる。Grechanovsky らは、閉検定手順では検定回数を減らすショートカット法を導くことが可能であると述べており⁷³、Dmitrienko ら⁷⁴、Hommel ら⁷⁵がボンフェローニ法を適用する場合のショートカット法を提案している。しかし、リサンプリング法を適用する場合のゲートキーピングプロシジャのショートカット法はまだ提案されていない。

また、閉検定手順を用いたパラレルゲートキーピングプロシジャは仮説の順序性を

適切に考慮していないという問題がある。パラレルゲートキーピングプロシジャは、全てのプライマリエンドポイントの積仮説を棄却すれば少なくとも1つのプライマリエンドポイントを棄却する検定手順である。しかし、同時帰無分布の構造によりプライマリエンドポイントとセカンダリエンドポイントの積仮説が棄却されないことで、棄却されるべきプライマリエンドポイントが棄却されないことがある。この原因は、閉検定手順がセカンダリエンドポイントを考慮した積仮説の検定を常に行うことにある。この問題を解消し、仮説の順序性を適切に考慮したパラレルゲートキーピングプロシジャはまだ提案されていない。

もう1つの問題はリサンプリング法を適用した場合の同時帰無分布の推定である。並べ替え法はデータ生成分布に帰無分布を想定しているため、帰無仮説の全ての組み合わせに関する積仮説それぞれにおいて同時帰無分布の推定を行う必要があり、時間と効率の観点で問題である。Westfallらはサブセットピボタリティが成立すれば1回の同時帰無分布の推定により全ての仮説の検定が行えることを示している⁷⁶。しかし、Dudoitら^{77,78}、Romanoら⁷⁹、van der Laanら⁸⁰はサブセットピボタリティを仮定すること自体に問題がある可能性を指摘している。また、並べ替え法は同時交換可能性も仮定している⁵²が、Huangら⁸¹、Hsu⁸²は、同時交換可能性の仮定の間違いがエンドポイントの検定結果に影響する問題を指摘している。例えば、群間でエンドポイントの周辺分布が等しいという治療効果なしの帰無仮説が正しくても、群間で同時分布の相関構造が違うことにより治療効果ありと誤って判断する可能性がある。一方、ブー

トストラップ法は真のデータ生成分布に基づく同時分布の推定方法であるため、同時帰無分布の推定は困難であり、検定への利用に関する議論がなされてこなかった⁸³。しかし、近年になり、遺伝子統計学の分野で Dudoit ら⁷⁷、van der Laan ら⁸⁰が、サブセットピボタリティや同時交換可能性の仮定を必要としないブートストラップ法による同時帰無分布の推定方法を提案した。1つは帰無仮説における検定統計量や p 値の平均と分散を用いる方法⁷⁷、もう1つは帰無仮説における検定統計量や p 値の周辺分布を用いる方法⁸⁰である。これらは平均、分散、周辺分布を特定できれば、漸近的に FWER が名義水準以下となる^{78,80}方法である。また、前者に比べて、後者は一様に検出力が高いことが示されている^{78,80,84}。ブートストラップ法は1回の同時帰無分布の推定で全ての検定が行えるため、臨床試験でも有効な方法であると考えられるが、これらの方法をゲートキーピングプロシジャに適用したものはない。

1.5. 研究の目的

複数のプライマリエンドポイントとセカンダリエンドポイントを用いて治療効果を評価する場合、仮説の順序性を考慮した多重性の調整方法を利用することが重要である。いくつかの閉検定手順による検定手順はシングルステップ法の繰り返しとして導くことが可能であり^{78,79,85}、これを拡張することで簡潔かつ適切に仮説の順序性を考慮したゲートキーピングプロシジャを導くことが可能である。このような検定手順はまだ提案されていない。その中でも、仮定が少なく検出力が高い検定方法を用いた

多重性の調整方法を提案することは臨床試験の目的を達成するためには意義がある。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。本論文の構成は、第2章で既存の多重性の調整方法と提案法を紹介し、第3章でシミュレーション実験の結果を示し、第4章で考察する。

2. 方法

2.1. 想定する状況

ランダム化2群比較試験において、試験治療に対する標準治療の効果を複数のプライマリエンドポイントと複数のセカンダリエンドポイントを用いて評価する状況を想定する。

試験治療であれば1、標準治療であれば2をとる変数を X 、 M 個のエンドポイントを $Y = \{Y_i : i=1, \dots, M\}$ と表す。ただし、 $\{Y_i : i=1, \dots, m\}$ はプライマリエンドポイント、 $\{Y_i : i=m+1, \dots, M\}$ はセカンダリエンドポイントで、連続変数もしくは2値変数いずれかを想定する。群1には n_1 人、群2には n_2 人が割り付けられ、個人のエンドポイントを $Y_{11k}, i = 1, \dots, M, k = 1, \dots, n_1, Y_{12k}, i = 1, \dots, M, k = 1, \dots, n_2$ とする。各群の治療効果は、連続変数であれば平均、2値変数であればリスクによって定義し、群 j における平均とリスクは $\mu_j = \{\mu_{ij} : i=1, \dots, M, j=1, 2\}$ 、 $p_j = \{p_{ij} : i=1, \dots, M, j=1, 2\}$ と表す。治療効果は差の指標 $\delta = \{\delta_i : i=1, \dots, M\}$ を用い、連続量であれば $\delta_i = \mu_{i1} - \mu_{i2}$ 、2値であれば $\delta_i = p_{i1} - p_{i2}$ と

なる。検定の帰無仮説は $H = \{H_i : \delta_i \geq 0, i=1, \dots, M\}$ 、対立仮説は $H_A = \{H_{Ai} : \delta_i < 0, i=1, \dots, M\}$ とする。検定は、連続変数は t 検定、2 値変数は Fisher の直接確率検定によりを行い、仮説の棄却は p 値によって判断する。

2.2. 多重性の調整方法

本節では未調整 p 値を用いた多重性の調整方法について述べる。未調整 p 値は各帰無仮説の検定で計算される p 値で、 $p = \{p(H_i) : i=1, \dots, M\}$ と表す。ただし、調整 p 値 $\tilde{p} = \{\tilde{p}(H_i) : i = 1, \dots, M\}$ も同時に導出する。調整 p 値はある多重性の調整方法のもとで H_i を棄却できる最小の FWER の名義水準のことである⁴⁵。未調整 p 値が調整有意水準以下のとき帰無仮説を棄却する、ある検定手順のもとで帰無仮説を棄却することは、調整 p 値が未調整の有意水準以下のとき当該の帰無仮説を棄却することと同義である。また、調整 p 値は 1 回の検定で用いる通常の p 値と同様の解釈ができ、有意かどうかの 2 値判断より多くの情報を与える。

既存の方法として、2.2.1 節でシングルステップ法、2.2.2 節で閉検定手順、2.2.3 節で閉検定手順を用いたゲートキーピングプロシジャを記述する。2.2.4 節では提案するブートストラップ法を適用するゲートキーピングプロシジャについて紹介する。

2.2.1. シングルステップ法

シングルステップ法で用いる調整有意水準 $c(M; \alpha)$ は、帰無仮説の数 M 、未調整 p

値の同時帰無分布 Q_0 、FWER の名義水準 α により定義され、(1)式を満たす。

$$\Pr_{Q_0}\{\min_{i=1,\dots,M} p(H_i) \leq c(M; \alpha)\} \leq \alpha \quad (1)$$

シングルステップ法の 1 つであるボンフェローニ法の調整有意水準は、(1)式とボンフェローニの不等式から導かれ、 $w(H_i)\alpha$ である。ただし、 $w(H_i)$ は H_i の重要度が高くなるほど大きくなる重みで、 $\sum_{i=1}^M w(H_i) = 1$ である。ボンフェローニ法の調整 p 値は $p(H_i)/w(H_i)$ である。

2.2.2. 閉検定手順

閉検定手順とは、基本帰無仮説を誘導する全ての共通帰無仮説とその基本帰無仮説が棄却されたときのみ当該の帰無仮説を棄却する検定手順である。簡単のため、帰無仮説が $\{H_1, H_2, H_3, H_4\}$ の状況のもとで説明する。

閉検定手順で考慮する帰無仮説の種類は、基本帰無仮説、共通帰無仮説、包括帰無仮説がある。基本帰無仮説は $H_i, i=1, \dots, 4$ である。共通帰無仮説はいくつかの基本帰無仮説の積仮説のことで、2 つの基本帰無仮説からなる場合であれば、 $H_{i\bar{i}} = H_i \cap H_{\bar{i}}, i=1, \dots, 4, \bar{i}=1, \dots, 4, i \neq \bar{i}$ である。包括帰無仮説は全ての基本帰無仮説の積仮説 $H_{1234} = H_1 \cap H_2 \cap H_3 \cap H_4$ であり、共通帰無仮説の 1 つである。これらの帰無仮説で定義される仮説の集合を仮説族という。仮説族に含まれる帰無仮説がその帰無仮説の積仮説を全て含んでいるとき、その仮説族を閉じた仮説族という。閉じた仮説族について、基本帰無仮説を底辺にとり、2 つの帰無仮説から導かれる積仮説を下から順番に表したも

のが図 1 である。

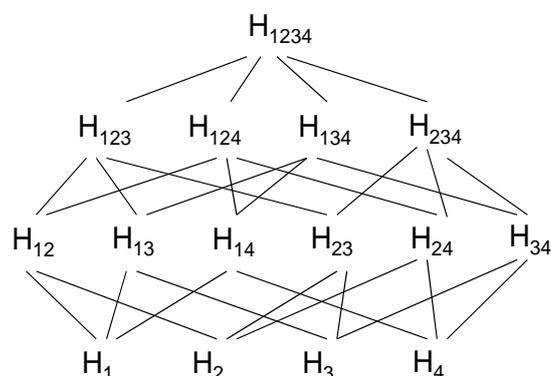


図 1 閉じた仮説族

一般に閉じた仮説族は基本帰無仮説を誘導する全ての共通帰無仮説を含んでいる。誘導するとは、ある帰無仮説の真であることが仮説族に含まれる他の帰無仮説も真であることを導くことをいう。例えば、 H_{123} は $\{H_{12}, H_{13}, H_{23}, H_1, H_2, H_3\}$ を誘導する。閉検定手順のもとで FWER が α 以下となるように H_i を棄却するには、閉じた仮説族の中で H_i を誘導する全ての帰無仮説と H_i を有意水準 α で棄却したとき H_i を棄却すればよい。多重性を調整した H_i の検定に必要な共通帰無仮説の検定は重複するので、図 1 の閉じた仮説族の階層性を利用した以下の手順で検定を行う。

手順 1 : H_{1234} の検定を有意水準 α で行う。

手順 2 : H_{1234} の検定が有意であれば、 H_{123} 、 H_{124} 、 H_{134} 、 H_{234} をそれぞれ有意水準 α で検定を行う。 H_{1234} の検定が有意でなければいずれの仮説も検定を行わない。

手順 3 : H_{12} を誘導する共通帰無仮説 (H_{123} 、 H_{124}) の検定が全て有意であったとき、

H_{12} の検定を有意水準 α で行う。共通帰無仮説の検定が 1 つでも有意でなけ

れば H_{12} の検定を行わない。 H_{13} 、 H_{14} 、 H_{23} 、 H_{24} 、 H_{34} も同様の手順で検定を行う。

手順 4: H_1 を誘導する共通帰無仮説 (H_{12} 、 H_{13} 、 H_{14}) の検定が全て有意であったとき、

H_1 の検定を有意水準 α で行う。共通帰無仮説の検定が 1 つでも有意でなければ H_1 の検定を行わない。 H_2 、 H_3 、 H_4 も同様の手順で検定を行う。

H_i の調整 p 値は H_i を誘導する共通帰無仮説の p 値と H_i の p 値の最大値である。例えば

H_1 の調整 p 値は、

$$\tilde{p}(H_1) = \max(p(H_1), p(H_{12}), p(H_{13}), p(H_{14}), p(H_{123}), p(H_{124}), p(H_{134}), p(H_{1234}))$$

となる。 $p(\cdot)$ は括弧内の共通帰無仮説の検定に対する p 値である。

2.2.3. 閉検定手順を用いたゲートキーピングプロシジャ

2.2.3.1 節、2.2.3.2 節でシリアルゲートキーピングプロシジャ、パラレルゲートキーピングプロシジャの考え方と閉検定手順を用いた検定手順について、2.2.3.3 節でボンフェローニ法を適用したゲートキーピングプロシジャ、2.2.3.4 節で並べ替え法を適用したゲートキーピングプロシジャ、2.2.3.5 節で並べ替え法を適用した際に生じる問題について記述する。

2.2.3.1. シリアルゲートキーピングプロシジャ

重要度が高いファミリーから 1,2,.. と順位をつける。シリアルゲートキーピングプロ

シジヤは、 d 番目のファミリーに含まれる全ての帰無仮説が棄却されたとき、 $d+1$ 番目のファミリーに含まれる帰無仮説の検定を行う検定手順である。閉検定手順のもとでは、複数のファミリーの帰無仮説を誘導する共通帰無仮説に対し、ゲートキーパーとなる最も上位のファミリーに含まれる帰無仮説の積仮説に対する検定を行う。簡単のため、プライマリエンドポイントの帰無仮説 $\{H_1, H_2\}$ をファミリー1、セカンダリエンドポイントの帰無仮説 $\{H_3, H_4\}$ をファミリー2とした状況のもとで説明する。

H_1, H_2 ともに棄却したとき、 H_3, H_4 の検定を行う検定手順であり、図 2 のように表される。

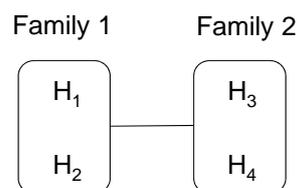


図 2 シリアルゲートキーピングプロシジヤ

H_1, H_2 はそれぞれが H_3, H_4 の検定を行うかを決めるゲートキーパーである。

閉検定手順のもとで、検定はまずファミリー1 の帰無仮説を誘導する共通帰無仮説に関して行う。ファミリー1 に含まれる帰無仮説を誘導する共通帰無仮説に注目し、閉じた仮説族を用いて表現したものが図 3 である。

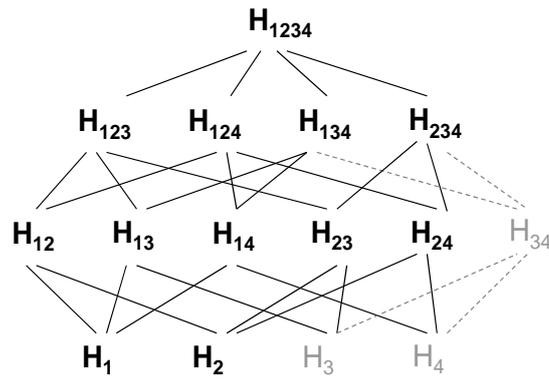


図 3 ファミリー1に含まれる帰無仮説の検定に関する共通帰無仮説

これらの共通帰無仮説の検定は H_3 、 H_4 の真偽に依存しないため、 H_1 、 H_2 とともに誘導する共通帰無仮説の検定は H_{12} の検定、 H_1 (H_2) のみを誘導する共通帰無仮説の検定は H_1 (H_2) の検定と同様に行う。例えば、 H_{1234} であれば H_{12} の検定を用いる。このようにゲートキーパーを考慮して共通帰無仮説の検定を行う。

ファミリー1の帰無仮説を全て棄却したとき、ファミリー2の帰無仮説を誘導する共通帰無仮説に関する検定を行う。その際、ファミリー1の帰無仮説を誘導する共通帰無仮説は全て棄却されている。閉じた仮説族の中でファミリー2の帰無仮説の検定に関する共通帰無仮説に注目したものが図4である。

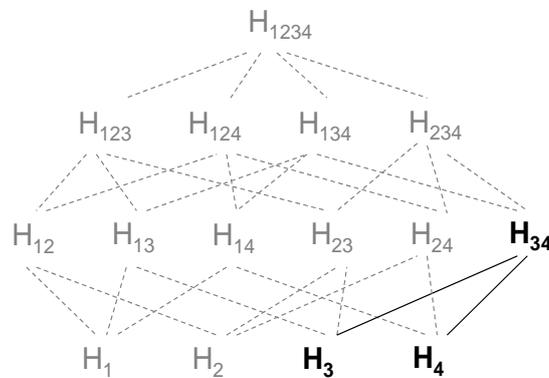


図 4 ファミリー2に含まれる帰無仮説の検定に関する共通帰無仮説

ファミリー2 の帰無仮説のみで定義される共通帰無仮説は積仮説に対する検定をそのまま行う。

2.2.3.2. パラレルゲートキーピングプロシジャ

パラレルゲートキーピングプロシジャは、 d 番目のファミリーに含まれる帰無仮説を1つでも棄却すれば、 $d+1$ 番目のファミリーに含まれる帰無仮説の検定を行う検定手順である。閉検定手順のもとでは、ゲートキーパーとなる全ての帰無仮説を誘導し、かつ複数のファミリーの帰無仮説を誘導する共通帰無仮説に対して、ゲートキーパーより下位のファミリーの帰無仮説を除いて定義される積仮説と同様の検定を行う。また、複数のファミリーの帰無仮説を誘導し、ゲートキーパーを含まない共通帰無仮説を、どの基本帰無仮説の未調整 p 値により棄却するかで基本帰無仮説の検定を行うかを判断する。ただし、以下の2つの条件により d 番目のファミリーの帰無仮説の検定を行うかの判断が異なる。

- ・条件 A : $d+1$ 番目以降のファミリーに含まれる帰無仮説の検定結果に依存する。
- ・条件 B : $d+1$ 番目以降のファミリーに含まれる帰無仮説の検定結果に依存しない。

簡単のため、プライマリエンドポイントの帰無仮説 $\{H_1, H_2\}$ をファミリー1、セカンダリエンドポイントの帰無仮説 $\{H_3, H_4\}$ をファミリー2 とした状況のもとで説明する。

H_1, H_2 いずれかを棄却したとき H_3, H_4 の検定を行う検定手順であり、図5のように表される。 $\{H_1, H_2\}$ の集合が H_3, H_4 の検定を行うかを定めるゲートキーパーとなる。

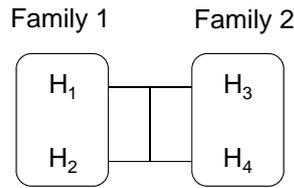


図 5 パラレルゲートキーピングプロシジャ

閉検定手順のもとで、検定はまずファミリー1の帰無仮説を誘導する共通帰無仮説
 に関して行う。ファミリー1に含まれる帰無仮説を誘導する共通帰無仮説に注目し、
 閉じた仮説族を用いて表現したものが図 6 である。

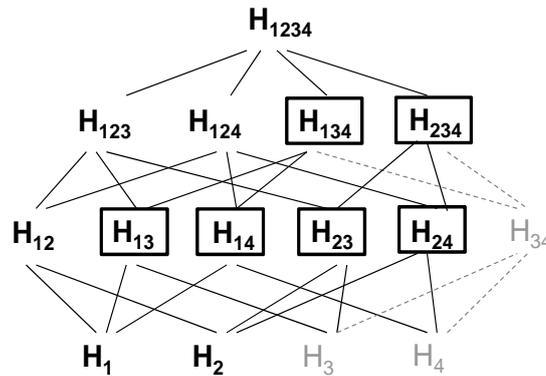


図 6 ファミリー1に含まれる帰無仮説の検定に関する共通帰無仮説

枠で囲まれているのは H_1 、 H_2 のどちらか一方のみを誘導する共通帰無仮説である。
 枠で囲まれていない共通帰無仮説の検定は H_3 、 H_4 の真偽に依存しないため、 H_1 、 H_2
 とともに誘導する共通帰無仮説の検定は H_{12} の検定と同様の検定を行う。枠で囲まれた
 共通帰無仮説の検定は条件 A では積仮説に対する検定をそのまま行う。条件 B では
 シングルステップ法を用いて検定を行い、 H_1 、 H_2 の未調整 p 値により共通帰無仮説を
 棄却しなければ基本帰無仮説 H_1 、 H_2 の検定は行わないと判断する。

ファミリー1のいずれかの帰無仮説を棄却したとき、ファミリー2の帰無仮説を誘

導する共通帰無仮説に関する検定を行う。いくつかの共通帰無仮説はファミリー1の帰無仮説のみを用いて検定する。閉じた仮説族の中でファミリー2の帰無仮説の検定に関する共通帰無仮説に注目したのが図7である。

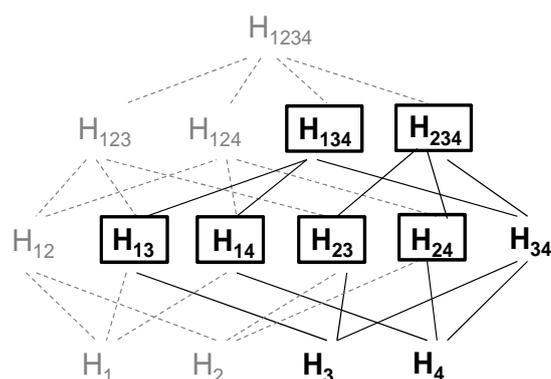


図7 ファミリー2に含まれる帰無仮説の検定に関する共通帰無仮説

ファミリー2の帰無仮説の検定では、枠で囲まれた共通帰無仮説は H_1 、 H_2 、 H_3 、 H_4 いずれの未調整 p 値で棄却されてもよいため、条件Bの場合では再度、積仮説の検定を全ての未調整 p 値を用いて行う。枠で囲まれてない共通帰無仮説はファミリー2の帰無仮説のみで定義されるので、条件によらず積仮説に対する検定をそのまま行う。

2.2.3.3. ボンフェローニ法を適用したゲートキーピングプロシジャ

共通帰無仮説が誘導する基本帰無仮説の集合を I とする。 I に含まれる帰無仮説に対して、仮説の重要度が高くなるほど大きくなる重み $w_I = \{w_I(H_i) : H_i \in I\}$ を、 $\sum_{H_i \in I} w_I(H_i) = 1$ となるように定義する。ボンフェローニ法によるグローバル検定は、 $H_i \in I$ のいずれかの帰無仮説について $p(H_i) \leq w_I(H_i)\alpha$ となれば共通帰無仮説を棄却する

検定方法である。共通帰無仮説の検定の p 値は $\min_{H_i \in I} p(H_i) / w_I(H_i)$ である。

シリアルゲートキーピングプロシジャの場合、ファミリー1の帰無仮説を誘導する共通帰無仮説の検定ではファミリー1の帰無仮説だけに重みを与え、ファミリー2の帰無仮説のみを誘導する共通帰無仮説の検定ではファミリー2の帰無仮説だけに重みを与える。ここで、2.2.3.1節の例をもとに各共通帰無仮説の検定を決定行列 (Decision Matrix) と呼ばれるものを用いてまとめたのが表1である。

ここではファミリー内の帰無仮説の重要度が同じとして各帰無仮説に与える重みを0.5とした。 H_{1234} などファミリー1の帰無仮説を全て誘導する共通帰無仮説の検定は H_{12} と、 H_{134} などのファミリー1の帰無仮説を誘導する共通帰無仮説の検定は H_1 、 H_2 いずれかの検定と同様に定義される。ファミリー2の帰無仮説のみを誘導する共通帰無仮説の検定では特別な考慮は行っていない。Implied Hypothesisにおけるそれぞれの列の最大値が対応する各帰無仮説の調整 p 値である。

表 1 シリアルゲートキーピングプロシジャの決定行列

Intersection hypothesis	p-value	Implied hypothesis			
		H ₁	H ₂	H ₃	H ₄
H ₁₂₃₄	$p(H_{1234}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{1234})$	$p(H_{1234})$	$p(H_{1234})$	$p(H_{1234})$
H ₁₂₃	$p(H_{123}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{123})$	$p(H_{123})$	$p(H_{123})$	0
H ₁₂₄	$p(H_{124}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{124})$	$p(H_{124})$	0	$p(H_{124})$
H ₁₂	$p(H_{12}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{12})$	$p(H_{12})$	0	0
H ₁₃₄	$p(H_{134}) = p(H_1)$	$p(H_{134})$	0	$p(H_{134})$	$p(H_{134})$
H ₁₃	$p(H_{13}) = p(H_1)$	$p(H_{13})$	0	$p(H_{13})$	0
H ₁₄	$p(H_{14}) = p(H_1)$	$p(H_{14})$	0	0	$p(H_{14})$
H ₁	$p(H_1) = p(H_1)$	$p(H_1)$	0	0	0
H ₂₃₄	$p(H_{234}) = p(H_2)$	0	$p(H_{234})$	$p(H_{234})$	$p(H_{234})$
H ₂₃	$p(H_{23}) = p(H_2)$	0	$p(H_{23})$	$p(H_{23})$	0
H ₂₄	$p(H_{24}) = p(H_2)$	0	$p(H_{24})$	0	$p(H_{24})$
H ₂	$p(H_2) = p(H_2)$	0	$p(H_2)$	0	$p(H_2)$
H ₃₄	$p(H_{34}) = \min(p(H_3)/0.5, p(H_4)/0.5)$	0	0	$p(H_{34})$	$p(H_{34})$
H ₃	$p(H_3) = p(H_3)$	0	0	$p(H_3)$	0
H ₄	$p(H_4) = p(H_4)$	0	0	0	$p(H_4)$

パラレルゲートキーピングプロシジャの場合、ファミリー1の帰無仮説を全て誘導する共通帰無仮説の検定では、ファミリー1の帰無仮説のみ重みを与える。その他の共通帰無仮説の検定に関しては誘導する基本帰無仮説に対して重みを与える。ここで、2.2.3.2節の例をもとに条件Aの場合の各共通帰無仮説の検定をまとめたものが表2である。

表 2 パラレルゲートキーピングプロシジャ (条件 A) の決定行列

Intersection hypothesis	p-value	Implied hypothesis			
		H ₁	H ₂	H ₃	H ₄
H ₁₂₃₄	$p(H_{1234}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{1234})$	$p(H_{1234})$	$p(H_{1234})$	$p(H_{1234})$
H ₁₂₃	$p(H_{123}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{123})$	$p(H_{123})$	$p(H_{123})$	0
H ₁₂₄	$p(H_{124}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{124})$	$p(H_{124})$	0	$p(H_{124})$
H ₁₂	$p(H_{12}) = \min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{12})$	$p(H_{12})$	0	0
H ₁₃₄	$p(H_{134}) = \min(p(H_1)/0.5, p(H_3)/0.25, p(H_4)/0.25)$	$p(H_{134})$	0	$p(H_{134})$	$p(H_{134})$
H ₁₃	$p(H_{13}) = \min(p(H_1)/0.5, p(H_3)/0.5)$	$p(H_{13})$	0	$p(H_{13})$	0
H ₁₄	$p(H_{14}) = \min(p(H_1)/0.5, p(H_4)/0.5)$	$p(H_{14})$	0	0	$p(H_{14})$
H ₁	$p(H_1) = p(H_1)$	$p(H_1)$	0	0	0
H ₂₃₄	$p(H_{234}) = \min(p(H_2)/0.5, p(H_3)/0.25, p(H_4)/0.25)$	0	$p(H_{234})$	$p(H_{234})$	$p(H_{234})$
H ₂₃	$p(H_{23}) = \min(p(H_2)/0.5, p(H_3)/0.5)$	0	$p(H_{23})$	$p(H_{23})$	0
H ₂₄	$p(H_{24}) = \min(p(H_2)/0.5, p(H_4)/0.5)$	0	$p(H_{24})$	0	$p(H_{24})$
H ₂	$p(H_2) = p(H_2)$	0	$p(H_2)$	0	$p(H_2)$
H ₃₄	$p(H_{34}) = \min(p(H_3)/0.5, p(H_4)/0.5)$	0	0	$p(H_{34})$	$p(H_{34})$
H ₃	$p(H_3) = p(H_3)$	0	0	$p(H_3)$	0
H ₄	$p(H_4) = p(H_4)$	0	0	0	$p(H_4)$

ここでは同一ファミリーの帰無仮説のみを誘導する共通帰無仮説の検定では与える重みを 0.5 とし、それ以外の共通帰無仮説の検定ではファミリー1 の帰無仮説に 0.5、その他の帰無仮説で残りの 0.5 を割り振ることとした。H₁₂₃₄ などファミリー1 の帰無仮説を全て誘導する共通帰無仮説の検定は H₁₂ の検定と同様に定義され、それ以外の共通帰無仮説の検定では特別な考慮は行っていない。

条件 B の場合、ファミリー1 の帰無仮説の検定が表 3 のようになる。

表 3 パラレルゲートキーピングプロシジャ (条件 B) の決定行列

Intersection hypothesis	p-value	Implied hypothesis			
		H ₁	H ₂	H ₃	H ₄
H ₁₂₃₄	$p(H_{1234})=\min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{1234})$	$p(H_{1234})$	$p(H_{1234})$	$p(H_{1234})$
H ₁₂₃	$p(H_{123})=\min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{123})$	$p(H_{123})$	$p(H_{123})$	0
H ₁₂₄	$p(H_{124})=\min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{124})$	$p(H_{124})$	0	$p(H_{124})$
H ₁₂	$p(H_{12})=\min(p(H_1)/0.5, p(H_2)/0.5)$	$p(H_{12})$	$p(H_{12})$	0	0
H ₁₃₄	$p(H_{134})=\min(p(H_1)/0.5, p(H_3)/0.25, p(H_4)/0.25)$	$p(H_{134})$	0	$p(H_{134})$	$p(H_{134})$
H ₁₃	$p(H_{13})=\min(p(H_1)/0.5, p(H_3)/0.5)$	$p(H_{13})$	0	$p(H_{13})$	0
H ₁₄	$p(H_{14})=\min(p(H_1)/0.5, p(H_4)/0.5)$	$p(H_{14})$	0	0	$p(H_{14})$
H ₁	$p^*(H_1)=p(H_1)/0.5$	$p^*(H_1)$	0	0	0
H ₂₃₄	$p(H_{234})=\min(p(H_2)/0.5, p(H_3)/0.25, p(H_4)/0.25)$	0	$p(H_{234})$	$p(H_{234})$	$p(H_{234})$
H ₂₃	$p(H_{23})=\min(p(H_2)/0.5, p(H_3)/0.5)$	0	$p(H_{23})$	$p(H_{23})$	0
H ₂₄	$p(H_{24})=\min(p(H_2)/0.5, p(H_4)/0.5)$	0	$p(H_{24})$	0	$p(H_{24})$
H ₂	$p^*(H_2)=p(H_2)/0.5$	0	$p^*(H_2)$	0	$p(H_2)$
H ₃₄	$p(H_{34})=\min(p(H_3)/0.5, p(H_4)/0.5)$	0	0	$p(H_{34})$	$p(H_{34})$
H ₃	$p(H_3)=p(H_3)$	0	0	$p(H_3)$	0
H ₄	$p(H_4)=p(H_4)$	0	0	0	$p(H_4)$

ファミリー1 の帰無仮説を誘導する共通帰無仮説がファミリー1 の帰無仮説の未調整 p 値を用いて棄却されないとファミリー1 の帰無仮説を検定しないことを反映した表となっている。H₁ の検定は、H₁₃₄、H₁₃、H₁₄ の検定で $p(H_1)/0.5 \leq \alpha$ となるとき行われる。そのため H₁ の検定の p 値は $p^*(H_1)=p(H_1)/0.5$ となる。 $p^*(H_2)$ も同様である。

2.2.3.4. 並べ替え法を適用したゲートキーピングプロシジャ

共通帰無仮説が誘導する基本帰無仮説に対応するインデックスの集合を S とし、 $Y_S=\{Y_i:i \in S\}$ とする。群 1 のデータ $\{Y_{S1i}:i = 1, \dots, n_1\}$ と群 2 のデータ $\{Y_{S2i}:i = 1, \dots, n_2\}$ を

$Y_{S11}, \dots, Y_{S1n_1}, Y_{S21}, \dots, Y_{S2n_2}$ の順に布置した観測データを、群の区別なく無作為に並べ替え、 n_1 番目までを群 1、それ以降を群 2 としたサンプル $Y_{S11}^a, \dots, Y_{S1n_1}^a, Y_{S21}^a, \dots, Y_{S2n_2}^a$ を作成する。サンプルは全ての並べ替えに関して作成する。全ての並べ替えの数が多いときは、その中から無作為に A 個のサンプルを作成する。各サンプルから各帰無仮説に対応する未調整 p 値、 $p_a^A = \{p_a^A(H_i): i = 1, \dots, M, a = 1, \dots, A\}$ を計算し、未調整 p 値の同時帰無分布 $Q_{P,S}$ を推定する。このとき、調整有意水準 $c(S, \alpha)$ は(2)式となる。

$$c(S, \alpha) = \max\{c: \Pr_{Q_{P,S}}\{\min_{i \in S} P(H_i) \leq c\} \leq \alpha\} \quad (2)$$

共通帰無仮説はいずれかの $i \in S$ に関して $p(H_i) \leq c(S, \alpha)$ となれば棄却する。共通帰無仮説の p 値は $\Pr_{Q_{P,S}}\{\min_{i \in S} P(H_i) \leq \min_{i \in S} p(H_i)\}$ である。

重みを利用したボンフェローニ法に対応させた並べ替え法では、未調整 p 値を $p(H_i)/w_I(H_i)$ と置き換えた検定を行う。 $I = \{H_i: i \in S\}$ である。

検定手順は 2.2.3.3 節と同様であるので省略する。

2.2.3.5. 並べ替え法を適用したゲートキーピングプロシジャの問題点

共通帰無仮説の検定に並べ替え法を適用する場合、 $p(H_{12}) \leq \alpha$ となっても、

$$\tilde{p}(H_1) = \max(p(H_1), p(H_{12}), p(H_{13}), p(H_{14}), p(H_{123}), p(H_{124}), p(H_{134}), p(H_{1234})) \geq \alpha$$

$$\tilde{p}(H_2) = \max(p(H_2), p(H_{12}), p(H_{23}), p(H_{24}), p(H_{123}), p(H_{124}), p(H_{234}), p(H_{1234})) \geq \alpha$$

となる可能性がある。同時帰無分布の構造によって、

$$\min(p(H_{13}), p(H_{14}), p(H_{134}), p(H_{23}), p(H_{24}), p(H_{234})) \geq p(H_{12})$$

となるためである。 $p(H_{12}) \leq \alpha$ となればプライマリエンドポイントのいずれかは棄却されることがパラレルゲートキーピングプロシジャの考え方である。しかし、閉検定手順を用いたパラレルゲートキーピングプロシジャでは $\{H_{13}, H_{14}, H_{134}, H_{23}, H_{24}, H_{234}\}$ の全てに関してファミリー2の帰無仮説 H_3, H_4 が寄与するため、その考え方を満たさない。ファミリー1の帰無仮説の検定で $\{H_{13}, H_{14}, H_{134}\}$ 、 $\{H_{23}, H_{24}, H_{234}\}$ のいずれかを H_1 (H_2)と同様の検定を行うことでこの問題は改善される。しかし、閉検定手順では H_1, H_2 のどちらが真かを区別しないため改善方法が利用できず、仮説の順序性を適切に考慮できないという問題が起こる。

2.3. 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

2.3.1節で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を提案する。

2.3.2節で van der Laan らが提案した周辺分布を用いたブートストラップ法による同時帰無分布の推定法を記述する。本研究では他のブートストラップ法より検出力が高いことが示されている周辺分布を用いる方法に限定する。

2.3.1. 提案するゲートキーピングプロシジャ

本節では、2.2.1節で紹介したシングルステップ法を各手順の検定に用いるゲートキーピングプロシジャを提案する。仮説の重要度に従った重みを利用するボンフェローニ法にならい、重み付き p 値の真の同時帰無分布 Q_{v_0} を利用したシングルステップ法

のもとで手順を説明する。重み付き p 値は、帰無仮説の重要度に応じた重みを未調整 p 値にかけたもので、重みを $v = \{v(H_i): i = 1, \dots, M\}$ とすると、 $p_v = \{p(H_1)/v(H_1), \dots, p(H_M)/v(H_M)\} = \{p_v(H_1), \dots, p_v(H_M)\}$ である。Hommel のショートカット法⁷⁴を参考に $\min_{i=1, \dots, m} v(H_i) \geq \sum_{i=m+1}^M v(H_i)$ とする。リサンプリング法を用いる場合、 $\sum_{i=1}^M v(H_i) \leq 1$ と基準化する必要はないため、各手順で重みを変更する必要はない。

Q_{v_0} は未知であることが多いため、推定方法として並べ替え法や 2.3.2 節のブートストラップ法を利用する。また、提案する手順はシングルステップ法の繰り返しであるので、各手順の検定にボンフェローニ法を用いることも可能である。

2.3.1.1. シリアルゲートキーピングプロシジャ

ファミリー1 の帰無仮説 $\{H_1, \dots, H_m\}$ を重み付き p 値が小さい順に O_{11}, \dots, O_{1m} 、対応する重み付き p 値を $p_v(O_{1i}), i = 1, \dots, m$ とする。このとき、重み付き p 値が i 番目以降である帰無仮説の集合を $O_{1(i)} = \{O_{1i}, \dots, O_{1m}\}$ で表す。同様にファミリー2 の帰無仮説 $\{H_{m+1}, \dots, H_M\}$ を重み付き p 値が小さい順に O_{21}, \dots, O_{2M-m} 、対応する重み付き p 値を $p_v(O_{2i}), i = 1, \dots, M - m$ 、重み付き p 値が i 番目以降である帰無仮説の集合を $O_{2(i)} = \{O_{2i}, \dots, O_{2M-m}\}$ と表す。重み付き p 値の確率変数を $P_v = \{P_v(H_1), \dots, P_v(H_M)\}$ 、実現値を $p_v = \{p_v(H_1), \dots, p_v(H_M)\}$ とし、 P_v の真の同時帰無分布を Q_{v_0} 、 Q_{v_0} のもとでのある事象が起こる確率を $\Pr_{Q_{v_0}}$ と表すとき、提案する検定手順は以下で表される。

手順 1 : ファミリー 1 に含まれる帰無仮説に対し、

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

となれば帰無仮説 O_{11} を棄却して手順 2 に進む。 O_{11} を棄却しなければ検定を終える。

手順 2 : $2 \leq h \leq m$ において、棄却されていないファミリー 1 に含まれる帰無仮説

$O_{1(h)} = \{O_{1h}, \dots, O_{1m}\}$ に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in O_{1(h)}) \leq p_v(O_{1h})) \leq \alpha$$

となれば帰無仮説 O_{1h} を棄却する。 $h < m$ のとき手順 2 を繰り返し、 $h = m$ のとき手順 3 に進む。 O_{1h} を棄却しなければ検定を終える。

手順 3 : ファミリー 2 に含まれる帰無仮説に対し、

$$\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in O_{2(1)}) \leq p_v(O_{21})) \leq \alpha$$

となれば帰無仮説 O_{21} を棄却して手順 4 に進む。 O_{21} を棄却しなければ検定を終える。

手順 4 : $2 \leq h \leq M - m$ において、棄却されていないファミリー 2 に含まれる帰無仮説

$O_{2(h)} = \{O_{2h}, \dots, O_{2M-m}\}$ に対し、

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

となれば、帰無仮説 O_{2h} を棄却する。 $h < M - m$ のとき手順 4 を繰り返し、 $h = M - m$ のとき検定を終える。 O_{2h} を棄却しなければ検定を終える。

ファミリー 1 の帰無仮説の調整 p 値は、

$$\max_{l=1,\dots,h} (\Pr_{Q_{v_0}} (\min(P_v(H_i): H_i \in O_{1(l)}) \leq p_v(O_{1l})))$$

ファミリー2 の帰無仮説の調整 p 値は、

$$\max_{l'=1,\dots,h} \left(\begin{array}{l} \max_{l=1,\dots,m} (\Pr_{Q_{v_0}} (\min(P_v(H_i): H_i \in O_{1(l)}) \leq p_v(O_{1l}))), \\ \Pr_{Q_{v_0}} (\min(P_v(H_i): H_i \in O_{2(l')}) \leq p_v(O_{2l'})) \end{array} \right)$$

である。

簡単のため、プライマリエンドポイントに関する帰無仮説 H_1 、 H_2 をファミリー1、セカンダリエンドポイントに関する帰無仮説 H_3 、 H_4 をファミリー2、 $p_v(H_1) \leq p_v(H_2)$ 、 $p_v(H_3) \leq p_v(H_4)$ であるときに検定手順を適用する場合を説明する。

ファミリー1 の帰無仮説について、

$$\Pr_{Q_{v_0}} (\min(P_v(H_1), P_v(H_2)) \leq p_v(H_1)) \leq \alpha$$

となれば H_1 を棄却し H_2 の検定を行う。 H_1 を棄却できなければ検定を終える。次に、

$$\Pr_{Q_{v_0}} (P_v(H_2) \leq p_v(H_2)) \leq \alpha$$

となれば H_2 を棄却しファミリー2 の帰無仮説の検定を行う。 H_2 を棄却できなければ検定を終える。

次に、ファミリー2 の帰無仮説について、

$$\Pr_{Q_{v_0}} (\min(P_v(H_3), P_v(H_4)) \leq p_v(H_3)) \leq \alpha$$

となれば H_3 を棄却し H_4 の検定を行う。 H_3 を棄却できなければ検定を終える。最後に、

$$\Pr_{Q_{v_0}} (P_v(H_4) \leq p_v(H_4)) \leq \alpha$$

となれば、 H_4 を棄却し検定を終える。そうでなければ H_4 を棄却せずに検定を終える。

この検定手順は、Hommel らが提案した閉検定手順に基づくシリアルゲートキーピングプロシジャのショートカット法⁷⁴と一致する。

2.3.1.2. パラレルゲートキーピングプロシジャ

条件 A における提案する手順は以下で表される。

手順 1 : ファミリー 1 に含まれる帰無仮説に対し、

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

となれば帰無仮説 O_{11} を棄却する。ファミリー 1 の残りの帰無仮説とファミリー 2 の帰無仮説 $\{O_{12}, \dots, O_{1m}, O_{21}, \dots, O_{2M-m}\}$ を、重み付き p 値が小さい順に $\bar{O}_{2i}, \dots, \bar{O}_{2M-1}$ 、対応する重み付き p 値を $p_v(\bar{O}_{2i}), i = 1, \dots, M-1$ 、重み付き p 値が i 番目以降である帰無仮説の集合を $\bar{O}_{2(i)} = \{\bar{O}_{2i}, \dots, \bar{O}_{2M-1}\}$ として、手順 2 に進む。 O_{11} を棄却しなければ検定を終える。

手順 2 : $1 \leq h \leq M-1$ において、前の手順までに棄却されていない帰無仮説

$$\bar{O}_{2(h)} = \{\bar{O}_{2h}, \dots, \bar{O}_{2M-1}\}, \text{ に対し、}$$

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

となれば帰無仮説 \bar{O}_{2h} を棄却する。 $h < M-1$ のとき手順 2 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。

O_{11} の調整 p 値は、

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

その他の帰無仮説の調整 p 値は、

$$\max_{l=1, \dots, h} \left(\Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in O_{1(1)}) \leq p_v(O_{11})), \right. \\ \left. \Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \bar{O}_{2(l)}) \leq p_v(\bar{O}_{2l})) \right)$$

である。

条件 B における提案する手順は以下で表される。

手順 1 : ファミリー-1 に含まれる帰無仮説に対し、

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

となれば帰無仮説 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 $\tilde{O}_{21}, \dots, \tilde{O}_{2m-1}, \tilde{O}_{2m}, \dots, \tilde{O}_{2M-1}$ 、対応する重み付き p 値を $p_v(\tilde{O}_{2i}), i = 1, \dots, M-1$ 、i 番目以降の仮説の集合を $\tilde{O}_{2(i)} = \{\tilde{O}_{2i}, \dots, \tilde{O}_{2M-1}\}$ として、手順 2 に進む。 O_{11} を棄却しなければ検定を終える。

手順 2 : $1 \leq h \leq M-1$ において、前の手順までに棄却されていない帰無仮説

$$\tilde{O}_{2(h)} = \{\tilde{O}_{2h}, \dots, \tilde{O}_{2M-1}\}, \text{ に対し、}$$

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

となれば帰無仮説 \tilde{O}_{2h} を棄却する。 $h < M-1$ のとき手順 2 を繰り返す。 $h = M-1$ のとき検定を終える。 \tilde{O}_{2h} を棄却できないとき、 $\{\tilde{O}_{2h}, \dots, \tilde{O}_{2M-1}\}$ を重み付き p 値が小さい順に $\check{O}_{21}, \dots, \check{O}_{2M-h}$ 、対応する重み付き p 値を $p_v(\check{O}_{2i}), i = 1, \dots, M-h$ 、重み付き p 値が i 番目以降である帰無仮説の集合を $\check{O}_{2(i)} = \{\check{O}_{2i}, \dots, \check{O}_{2M-h}\}$ として、手順 3 に進む。

手順 3 : $1 \leq h' \leq M-h$ において、前の手順までに棄却されなかった帰無仮説

$\bar{O}_{2(h')} = \{\bar{O}_{2h'}, \dots, \bar{O}_{2M-h}\}$ に対し、

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

となれば帰無仮説 $\bar{O}_{2h'}$ を棄却する。 $h' < M-h$ のとき手順 3 を繰り返す。 $h' = M-h$

のとき検定を終える。 $\bar{O}_{2h'}$ が棄却されなければ検定を終える。ただし、 $\bar{O}_{2h'}$ が

ファミリー1 の帰無仮説である場合、検定終了時に $\bar{O}_{2h'}$ は棄却されなかったも

のとして扱う。

O_{11} の調整 p 値は、

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

その他の帰無仮説の調整 p 値は、

$$\max_{l=1, \dots, h} \left(\begin{array}{l} \Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in O_{1(1)}) \leq p_v(O_{11})), \\ \Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \bar{O}_{2(l)}) \leq p_v(\bar{O}_{2l})) \end{array} \right)$$

ただし、検定順序によってファミリー2 の帰無仮説の調整 p 値は、

$$\max_{l'=1, \dots, h'} \left(\begin{array}{l} \Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in O_{1(1)}) \leq p_v(O_{11})), \\ \max_{l=1, \dots, h-1} \Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \bar{O}_{2(l)}) \leq p_v(\bar{O}_{2l})), \\ \Pr_{Q_{v_0}}(\min(P_v(H_i): H_i \in \bar{O}_{2(l')}) \leq p_v(\bar{O}_{2l'})) \end{array} \right)$$

となる。

簡単のため、プライマリエンドポイントに関する帰無仮説 H_1 、 H_2 をファミリー1、

セカンダリエンドポイントに関する帰無仮説 H_3 、 H_4 をファミリー2、 $p_v(H_1) \leq$

$p_v(H_3) \leq p_v(H_2) \leq p_v(H_4)$ であるときに検定手順を適用する場合を説明する。

条件 A の場合、ファミリー1 の帰無仮説について、

$$\Pr_{Q_{v_0}}(\min(P_v(H_1), P_v(H_2)) \leq p_v(H_1)) \leq \alpha$$

となれば H_1 を棄却し H_2 とファミリー2 の帰無仮説の検定を行う。 H_1 を棄却できなければ検定を終える。次に、重み付き p 値が小さい順番からファミリー2 の帰無仮説である H_3 に関して、

$$\Pr_{Q_{v_0}}(\min(P_v(H_2), P_v(H_3), P_v(H_4)) \leq p_v(H_3)) \leq \alpha$$

となれば H_3 を棄却し残った帰無仮説の検定を行う。 H_3 を棄却できなければ検定を終える。同様に、

$$\Pr_{Q_{v_0}}(\min(P_v(H_2), P_v(H_4)) \leq p_v(H_2)) \leq \alpha$$

となれば H_2 を棄却し H_4 の検定を行う。 H_2 を棄却できなければ検定を終える。最後に、

$$\Pr_{Q_{v_0}}(P_v(H_4) \leq p_v(H_4)) \leq \alpha$$

となれば H_4 を棄却し検定を終える。そうでなければ H_4 を棄却せずに検定を終える。

条件 B の場合、ファミリー1 の帰無仮説について、

$$\Pr_{Q_{v_0}}(\min(P_v(H_1), P_v(H_2)) \leq p_v(H_1)) \leq \alpha$$

となれば H_1 を棄却し H_2 とファミリー2 の帰無仮説の検定を行う。 H_1 を棄却できなければ検定を終える。次に、ファミリーが上位である順番から H_2 に関して、

$$\Pr_{Q_{v_0}}(\min(P_v(H_2), P_v(H_3), P_v(H_4)) \leq p_v(H_2)) \leq \alpha \quad (\#)$$

となれば H_2 を棄却しファミリー2 の帰無仮説の検定を行う。

$$\Pr_{Q_{v_0}}(\min(P_v(H_3), P_v(H_4)) \leq p_v(H_3)) \leq \alpha$$

となれば H_3 を棄却し H_4 の検定を行う。 H_3 を棄却できなければ検定を終える。最後に、

$$\Pr_{Q_{v_0}}(P_v(H_4) \leq p_v(H_4)) \leq \alpha$$

となれば H_4 を棄却し検定を終える。そうでなければ H_4 を棄却せずに検定を終える。

ただし、(#)1において H_2 を棄却しなければ、 H_2 、 H_3 、 H_4 のなかで最も重み付き p 値が小さい H_3 に関して

$$\Pr_{Q_{v_0}}(\min(P_v(H_2), P_v(H_3), P_v(H_4)) \leq p_v(H_3)) \leq \alpha$$

となれば H_3 を棄却し H_2 と H_4 の検定を行う。 H_3 を棄却できなければ検定を終える。

同様に重み付き p 値が小さい順から

$$\Pr_{Q_{v_0}}(\min(P_v(H_2), P_v(H_4)) \leq p_v(H_2)) \leq \alpha$$

となれば H_2 を棄却し H_4 の検定を行う。ただし、 H_2 はファミリー1 の帰無仮説なので検定終了時に棄却されなかったとする。 H_2 を棄却できなければ検定を終える。最後に

$$\Pr_{Q_{v_0}}(P_v(H_4) \leq p_v(H_4)) \leq \alpha$$

となれば、 H_4 を棄却し検定を終える。そうでなければ H_4 を棄却せずに検定を終える。

条件 B のように検定すれば、ファミリー1 の帰無仮説の検定がファミリー2 の帰無仮説の検定結果に依存しないようになる。

2.3.2. 周辺分布を用いたブートストラップ法による同時帰無分布の推定

群ごとに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。各サンプルから各帰無仮説に対応する重み付き p 値 $p_{vb}^B = \{p_{vb}^B(H_i): i = 1, \dots, M, b = 1, \dots, B\}$ を計算する。

p_{vb}^B を用いて、重み付き p 値の周辺分布の経験累積分布関数、

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

を求める。これを使って重み付き p 値を変換した統計量、

$$\check{p}_{vb}^B(H_i) = q_{v0i}^{-1} Q_{vi}^B(p_{vb}^B(H_i))$$

を計算し、重み付き p 値の同時帰無分布 Q_{v0} を求める。ただし、 q_{v0i} は $p_v(H_i)$ の真の周辺帰無分布である。

周辺分布を利用する場合、以下のいずれかの条件が成り立てば少なくとも漸近的には FWER を名目水準以下に保つことが van der Laan らによって示されている^{78,80}。

- 周辺分布が正確に特定されている。
- 周辺分布の漸近分布が特定されている。
- 帰無仮説のもとで周辺分布に関する並べ替え分布が特定されている。

多くの場合いずれかの条件が仮定できるので、重み付き p 値の周辺分布は特定される。

3. シミュレーション実験

多重性の調整に対する提案法の性能を評価するためにシミュレーション実験を行った。性能の比較として、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パラレルゲートキーピングプロシジャ、提案したパラレルゲートキーピングプロシジャを比較した。シリアルゲートキーピングプロシジャは提案した手順と閉検定手順を一致させることが可能であるため、シミュレーション実験を行わなかった。

3.1. 状況設定

先行研究で行っているシミュレーション実験⁸⁶を参考に、ランダム化2群比較試験を想定してシミュレーション実験の状況を設定した。各群100人ずつ、2つのプライマリエンドポイント $\{Y_1, Y_2\}$ 、2つのセカンダリエンドポイント $\{Y_3, Y_4\}$ を用いて治療効果を評価する場合を想定した。

治療効果については、

- 1) 全てのエンドポイントで治療効果なし（帰無仮説）
- 2) 全てのエンドポイントで治療効果あり（対立仮説）

の2つの状況を想定した。

各エンドポイントは、

- I) $\{Y_1, Y_2, Y_3, Y_4\}$ が正規分布に従う連続変数
- II) $\{Y_1, Y_2, Y_3, Y_4\}$ が2項分布に従う2値変数
- III) $\{Y_1, Y_3\}$ が2項分布に従う2値変数、 $\{Y_2, Y_4\}$ が正規分布に従う連続変数

とし、

- i) 2群間で相関構造が等しい
- ii) 2群間で相関構造が違う

を想定した。i)のとき、全てのエンドポイント間の相関がピアソンの相関係数で0、0.2、0.5、0.8となるようデータを発生させた。ii)のとき、試験治療群と標準治療群における全てのエンドポイント間の相関が(0, 0.8)、(0.8, 0)となるようデータを発生さ

せた。ただし、2 値変数と連続変数の相関が 0.8 となることはないため、i)では 0.8 を除き、ii)では 0.8 から 0.5 に置き換えた。相関のある 2 値変数、相関のある 2 値変数と連続変数は、多変量正規分布から発生させた正規乱数を 2 値変数に変換する方法を利用した⁸⁷。治療効果がない場合、試験治療群 $\{Y_{11}, Y_{12}, Y_{13}, Y_{14}\}$ と標準治療群 $\{Y_{21}, Y_{22}, Y_{23}, Y_{24}\}$ のデータは、連続変数の周辺分布が平均 0、分散 1 の正規分布、2 値変数の周辺分布が確率 0.3 のベルヌーイ分布となる多変量正規分布から発生させた。治療効果がある場合、試験治療群 $\{Y_{11}, Y_{12}, Y_{13}, Y_{14}\}$ のデータは、連続変数の周辺分布が平均 -0.4、分散 1 の正規分布、2 値変数の周辺分布が確率 0.13 のベルヌーイ分布となる多変量正規分布から発生させた。標準治療群 $\{Y_{21}, Y_{22}, Y_{23}, Y_{24}\}$ のデータは、治療効果がない場合と同様に発生させた。設定根拠は多重性を調整しない方法においてエンドポイントごとの仮説が棄却される確率が 80%となることである。

エンドポイント $\{Y_1, Y_2, Y_3, Y_4\}$ に対応する帰無仮説 $\{H_1, H_2, H_3, H_4\}$ の検定は、連続変数なら t 検定、2 値変数なら Fisher の直接確率検定を用い、片側検定を行うこととした。多重性を調整しない方法では未調整 p 値、シングルステップ法とゲートキーピングプロシジャでは調整 p 値が 2.5%以下のとき当該の帰無仮説を棄却することとした。

ボンフェローニ法により調整 p 値を求める際の各帰無仮説の重みは、 $\{w(H_1), w(H_2), w(H_3), w(H_4)\} = \{0.4, 0.4, 0.1, 0.1\}$ とした。

ゲートキーピングプロシジャのファミリーは、 H_1, H_2 をファミリー 1、 H_3, H_4 をファミリー 2 とした。閉検定手順では、共通帰無仮説の検定にボンフェローニ法と並べ

替え法を適用した。ボンフェローニ法の重みは Dmitrienko らの先行研究⁵¹を参考に、表4のように設定した。並べ替え法でも表4の重みを使い、ボンフェローニ法と対応した検定を行った。リサンプリング回数は10000回として同時帰無分布を推定した。

表 4 共通帰無仮説の重み

	H ₁	H ₂	H ₃	H ₄
H ₁₂₃₄	0.5	0.5	0	0
H ₁₂₃	0.5	0.5	0	0
H ₁₂₄	0.5	0.5	0	0
H ₁₂	0.5	0.5	0	0
H ₁₃₄	0.5	0	0.25	0.25
H ₁₃	0.5	0	0.5	0
H ₁₄	0.5	0	0	0.5
H ₂₃₄	0	0.5	0.25	0.25
H ₂₃	0	0.5	0.5	0
H ₂₄	0	0.5	0	0.5
H ₃₄	0	0	0.5	0.5

提案した検定手順では、ブートストラップ法に加え、ボンフェローニ法、並べ替え法を各手順の検定に適用した。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た。これは、正規分布に従う変数に対する検定の未調整 p 値（重み付き p 値）の分布が一様分布に従うこと、帰無仮説のもとで並べ替え分布を用いることが妥当であることが理由である。リサンプリング回数は10000回として同時帰無分布を推定した。ボンフェローニ法と対応させるように、重みは $\{v(H_1), v(H_2), v(H_3), v(H_4)\} = \{0.4, 0.4, 0.1, 0.1\}$ とした。

各設定でのシミュレーション回数は2000回とした。FWERのシミュレーションエ

ラーは 0.35% である。解析には R2.13.1 を用いた。附録に解析プログラムを記載した。

3.2. 評価方法

全てのエンドポイントで治療効果なしの設定は、FWER の名義水準を 2.5% として性能を評価した。ただし、パラレルゲートキーピングプロシジャにおける FWER は、プライマリエンドポイントの帰無仮説を誤って棄却する確率と同値なので、多重性を調整しない方法とボンフェローニ法ではプライマリエンドポイントの帰無仮説を誤って棄却する確率も用いて性能を評価した。

全てのエンドポイントで治療効果ありの設定は、大きく 3 つの検出力を用いて性能を評価した。1 つ目は各帰無仮説を棄却する確率を検出力として用いた。これは 1 回の検定の際に用いる通常の検出力と同じである。2 つ目は多重性を考慮する際に用いる検出力で、全ての誤った帰無仮説を棄却する確率 (All Pairs Power)、任意の誤った帰無仮説を棄却する確率 (Any Pairs Power)、誤った帰無仮説の中で棄却された仮説の割合である平均検出力 (Mean power) を用いた。3 つ目はプライマリエンドポイントに注目した検出力として、プライマリエンドポイントに関する全ての誤った帰無仮説を棄却する確率 (All Primary-Endpoint-Pairs Power)、プライマリエンドポイントに関する任意の誤った帰無仮説が棄却される確率 (Any Primary-Endpoint Pairs Power) を性能評価に用いた。これらは多重性を調整する際に試験の成功確率を評価すべきという Bretz らの指摘⁸⁸に基づいて設定した。

3.3. シミュレーション実験の結果

全てのエンドポイントで治療効果なしの設定での FWER を表 5、表 6 に示す。

条件 A、B によって結果が変わらなかったため条件 B は省略する。

表 5 FWER (相関構造が等しい場合)

	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
連続変数	0	8.4(4.5)	2.0(1.6)	2.3	2.3	2.3	2.3	2.3	2.3
	0.2	7.4(4.2)	2.0(1.7)	2.1	2.2	2.1	2.2	2.2	2.1
	0.5	6.4(4.1)	2.0(1.7)	2.1	2.1	2.1	2.1	2.1	2.1
	0.8	4.5(3.4)	1.7(1.7)	1.9	2.4	1.9	2.4	2.3	2.3
2値変数	0	6.7(3.5)	1.4(1.3)	1.6	2.0	1.6	2.0	1.6	2.0
	0.2	6.9(3.4)	1.4(1.2)	1.6	1.9	1.6	1.9	1.6	1.8
	0.5	6.3(3.6)	1.6(1.3)	1.7	2.2	1.7	2.2	1.8	2.1
	0.8	4.6(3.1)	1.5(1.4)	1.6	2.5	1.6	2.5	1.9	2.5
2値変数と 連続変数	0	7.8(4.1)	1.8(1.6)	1.9	1.9	1.9	2.2	1.9	2.2
	0.2	8.0(4.5)	1.8(1.5)	2.1	2.1	2.1	2.4	2.1	2.4
	0.5	7.5(4.4)	1.9(1.7)	2.3	2.4	2.3	2.8	2.4	2.8

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 6 FWER (相関構造が違う場合)

	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
連続変数	(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値変数	(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
2値変数と 連続変数	(0,0.5)	7.8(4.3)	2.2(1.8)	2.2	2.1	2.2	2.5	2.2	2.5
	(0.5,0)	8.3(4.2)	1.8(1.6)	1.8	1.8	1.8	2.2	1.8	2.2

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)

相関構造が等しい場合に注目すると、多重性を調整しない方法は、全ての設定で **FWER**、プライマリエンドポイントを誤って棄却する確率ともに **2.5%**を大きく上回った。ボンフェローニ法は常に保守的で、特に **2** 値変数を含む場合は顕著であった。ボンフェローニ法を適用したパラレルゲートキーピングプロシジャは、閉検定手順、提案した手順によらず、全て連続変数で相関が高い場合、**2** 値変数を含む場合は常に保守的であった。提案した手順のうち、ボンフェローニ法と一様分布を用いたブートストラップ法、並べ替え法と並べ替え分布を用いたブートストラップ法の挙動がほぼ等しかった。**2** 値変数と連続変数の設定では、提案法で **2.5%**を超えたものもあるが、シミュレーションエラーの範疇であった。また、相関構造が違う場合に注目すると、並べ替え法での **FWER** の上昇は見られなかった。多重性の調整方法による違いは、相関構造が等しい場合と傾向は同じだった。

全てのエンドポイントで治療効果あり、全て連続変数、相関構造が等しい設定での検出力を、表 7 から表 10、図 8 から図 11 に示す。相関構造が違う場合は相関構造が等しい場合と同様の傾向を示したため、以下全ての設定で省略する。

表 7 仮説ごとの検出力 (連続変数、条件 A)

		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
相関 0	H ₁ (Family1)	79.3	67.2	73.0	73.2	75.1	74.9	75.2	75.0
	H ₂ (Family1)	79.8	66.6	73.7	73.7	75.2	75.0	75.5	75.1
	H ₃ (Family2)	78.7	50.7	66.3	66.1	63.5	63.5	64.0	63.6
	H ₄ (Family2)	81.0	49.8	68.0	67.8	64.6	65.0	64.8	65.0
相関 0.2	H ₁ (Family1)	79.5	67.2	73.0	73.2	74.4	74.2	74.6	74.2
	H ₂ (Family1)	79.8	67.5	73.4	73.6	74.8	74.7	75.2	74.8
	H ₃ (Family2)	78.4	50.4	64.9	65.4	63.1	63.3	63.5	63.3
	H ₄ (Family2)	81.4	49.9	67.4	67.5	64.6	64.8	65.0	64.8
相関 0.5	H ₁ (Family1)	79.5	67.4	73.2	73.9	74.4	75.1	75.2	75.0
	H ₂ (Family1)	80.0	67.3	73.0	74.0	74.7	75.2	75.4	75.0
	H ₃ (Family2)	78.8	50.0	65.2	66.0	64.2	65.1	65.3	65.0
	H ₄ (Family2)	81.1	49.4	67.2	68.4	65.1	66.0	65.8	65.8
相関 0.8	H ₁ (Family1)	79.5	67.8	72.5	74.9	73.9	75.8	76.1	75.8
	H ₂ (Family1)	79.7	67.6	72.8	75.3	74.0	76.1	76.3	76.1
	H ₃ (Family2)	79.3	48.6	66.8	70.2	66.7	69.3	69.3	69.1
	H ₄ (Family2)	80.0	48.9	67.2	70.7	66.7	68.9	69.2	68.9

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 8 多重性を考慮した検出力 (連続変数、条件 A)

		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
相関 0	All	41.1	11.2	38.4	38.2	38.7	38.6	39.0	38.6
	All Primary	63.3	45.0	55.9	56.1	59.6	59.3	59.9	59.4
	Any (Any Primary)	99.8 (95.8)	97.0 (88.8)	90.8	90.6	90.8	90.6	90.9	90.7
	Mean	79.7	58.6	70.2	70.2	69.6	69.6	69.9	69.7
相関 0.2	All	47.2	17.2	44.5	44.6	44.9	44.6	45.1	44.6
	All Primary	65.6	48.4	58.0	58.2	60.8	60.4	61.2	60.5
	Any (Any Primary)	98.9 (93.6)	93.7 (86.4)	88.3	88.4	88.3	88.4	88.6	88.5
	Mean	79.8	58.8	69.7	69.9	69.2	69.2	69.6	69.3
相関 0.5	All	56.6	26.5	52.7	53.4	53.4	53.9	54.0	53.8
	All Primary	68.8	52.4	61.6	62.6	64.5	65.0	65.3	65.0
	Any (Any Primary)	96.4 (90.6)	87.0 (82.2)	84.5	85.3	84.5	85.3	85.3	85.1
	Mean	79.8	58.5	69.7	70.6	69.6	70.4	70.5	70.2
相関 0.8	All	66.0	37.4	62.0	64.3	63.1	64.5	64.8	64.5
	All Primary	72.8	58.2	66.0	69.0	68.6	70.8	71.0	70.8
	Any (Any Primary)	91.1 (86.4)	78.6 (77.1)	79.2	81.2	79.2	81.2	81.5	81.2
	Mean	79.6	58.2	69.8	72.8	70.3	72.5	72.7	72.5

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

Mean: 平均検出力

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 9 仮説ごとの検出力 (連続変数、条件 B)

		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
相関 0	H ₁ (Family1)	79.3	67.2	69.8	69.7	73.0	72.8	73.2	72.8
	H ₂ (Family1)	79.8	66.6	70.0	69.8	73.1	73.2	73.4	73.2
	H ₃ (Family2)	78.7	50.7	66.3	66.1	63.5	63.5	64.0	63.6
	H ₄ (Family2)	81.0	49.8	68.0	67.8	64.6	65.0	64.8	65.0
相関 0.2	H ₁ (Family1)	79.5	67.2	69.8	69.5	72.5	72.7	72.8	72.8
	H ₂ (Family1)	79.8	67.5	70.3	70.0	72.7	72.8	73.3	72.9
	H ₃ (Family2)	78.4	50.4	64.9	65.4	63.1	63.3	63.5	63.3
	H ₄ (Family2)	81.4	49.9	67.4	67.5	64.6	64.8	65.0	64.8
相関 0.5	H ₁ (Family1)	79.5	67.4	70.5	71.0	73.0	73.9	74.1	73.8
	H ₂ (Family1)	80.0	67.3	70.1	70.7	72.9	73.9	74.0	73.7
	H ₃ (Family2)	78.8	50.0	65.2	66.0	64.2	65.1	65.3	65.0
	H ₄ (Family2)	81.1	49.4	67.2	68.4	65.1	66.0	65.8	65.8
相関 0.8	H ₁ (Family1)	79.5	67.8	70.4	73.0	72.5	75.5	75.8	75.5
	H ₂ (Family1)	79.7	67.6	70.7	73.6	73.2	75.9	76.1	75.8
	H ₃ (Family2)	79.3	48.6	66.8	70.2	66.7	69.3	69.3	69.1
	H ₄ (Family2)	80.0	48.9	67.2	70.7	66.7	68.9	69.2	68.9

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 10 多重性を考慮した検出力 (連続変数、条件 B)

	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 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 (Any Primary)	99.8 (95.8)	97.0 (88.8)	90.8	90.6	90.8	90.6	90.9	90.7
	Mean	79.7	58.6	68.5	68.3	68.6	68.6	68.8	68.7
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.2	All	47.2	17.2	38.2	38.1	41.5	41.8	41.9	41.8
	All Primary	65.6	48.4	51.7	51.5	56.8	57.0	57.5	57.1
	Any (Any Primary)	98.9 (93.6)	93.7 (86.4)	88.3	88.4	88.3	88.4	88.6	88.5
	Mean	79.8	58.8	68.1	68.1	68.2	68.4	68.7	68.5
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.5	All	56.6	26.5	47.1	47.8	50.8	51.7	51.9	51.6
	All Primary	68.8	52.4	56.0	56.6	61.4	62.5	62.7	62.3
	Any (Any Primary)	96.4 (90.6)	87.0 (82.2)	84.5	85.3	84.5	85.3	85.3	85.1
	Mean	79.8	58.5	69.7	70.6	69.6	70.4	70.5	70.2
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 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 (Any Primary)	91.1 (86.4)	78.6 (77.1)	79.2	81.2	79.2	81.2	81.5	81.2
	Mean	79.6	58.2	68.8	71.9	69.8	72.4	72.6	72.3

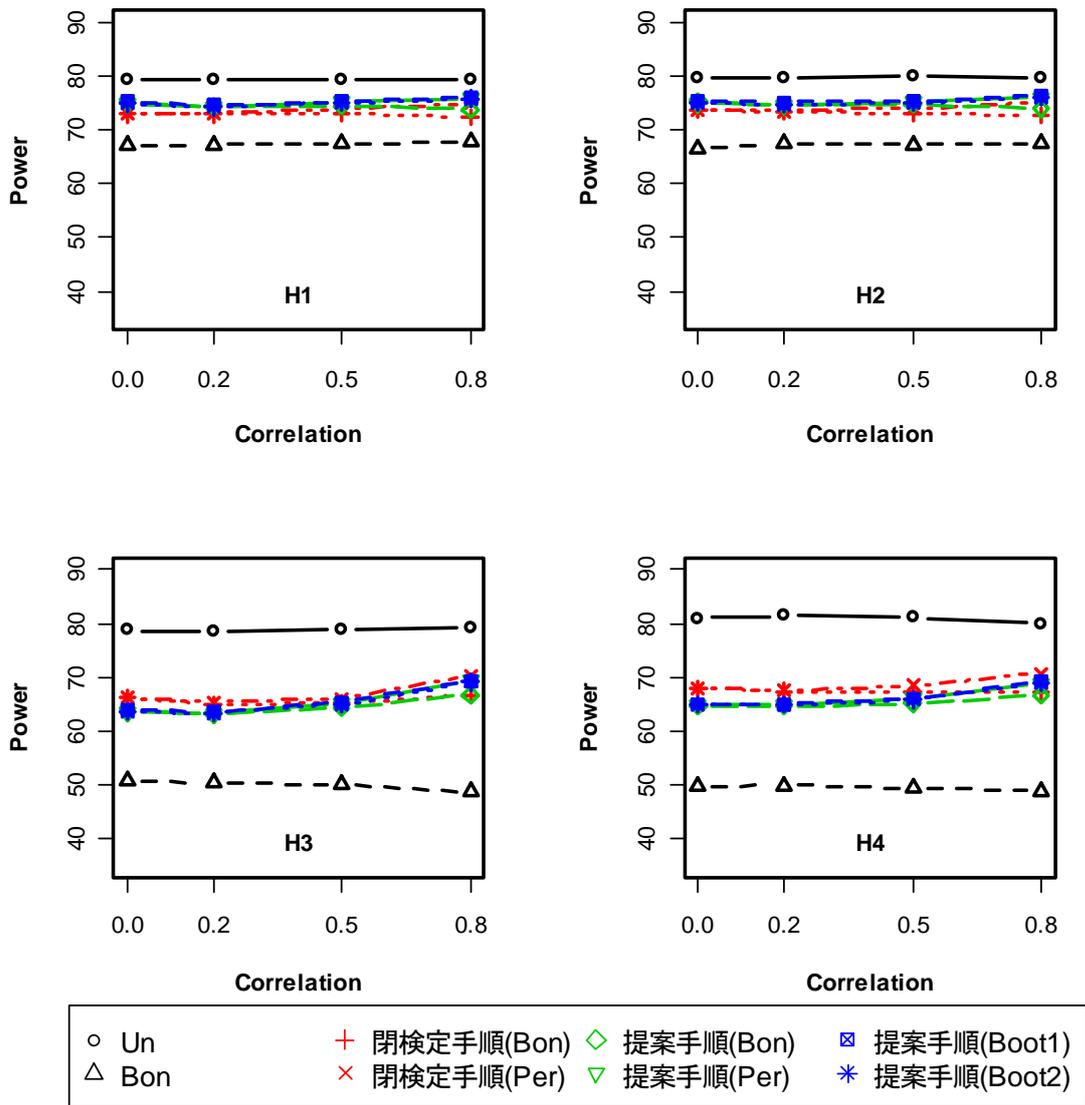
All: 全エンドポイントの棄却確率, All Primary: 全プライマリエンドポイントの棄却確率

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

Mean: 平均検出力

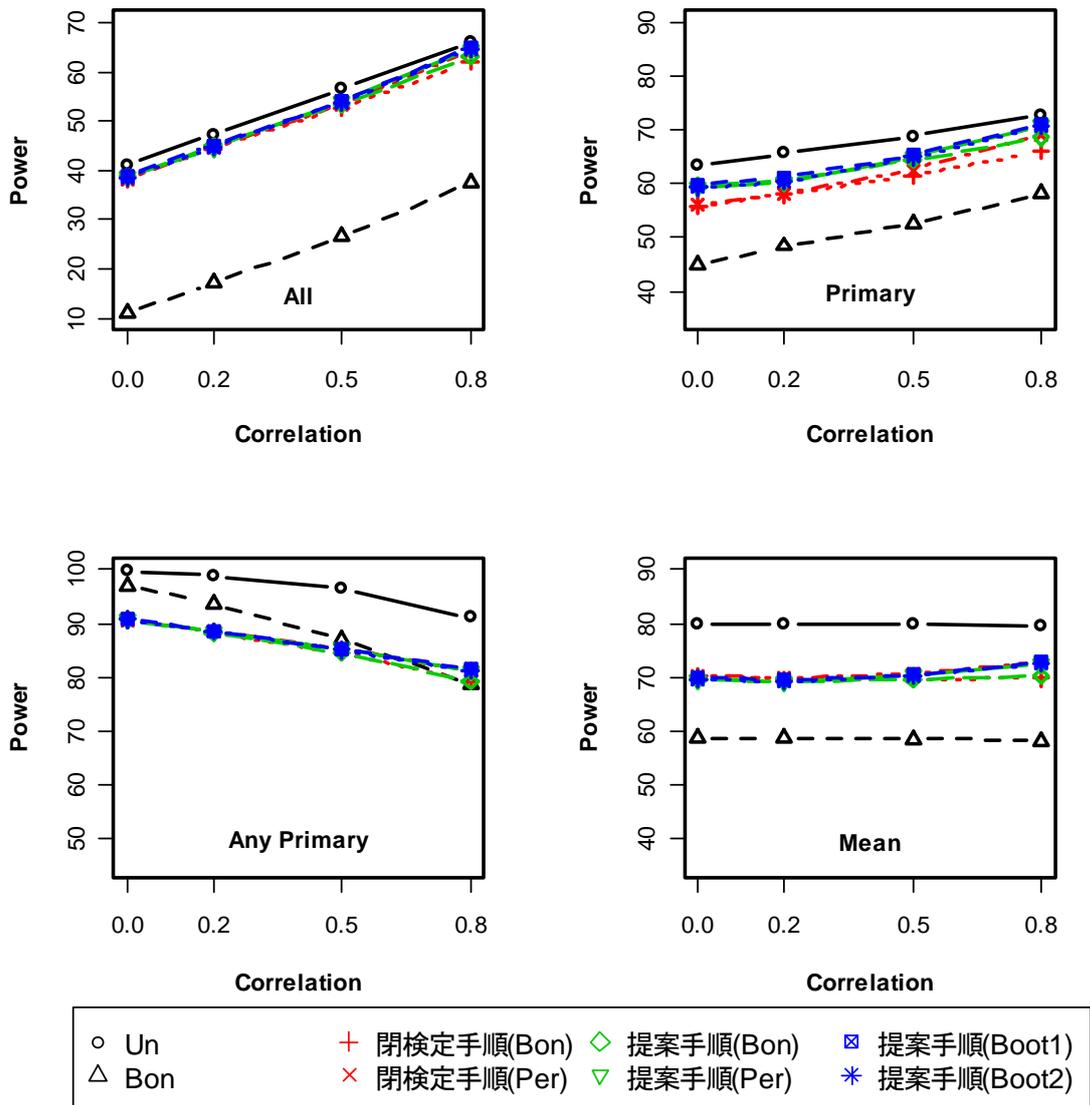
Un: 多重性を調整しない方法, Bon: ボンフェローニ法

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)



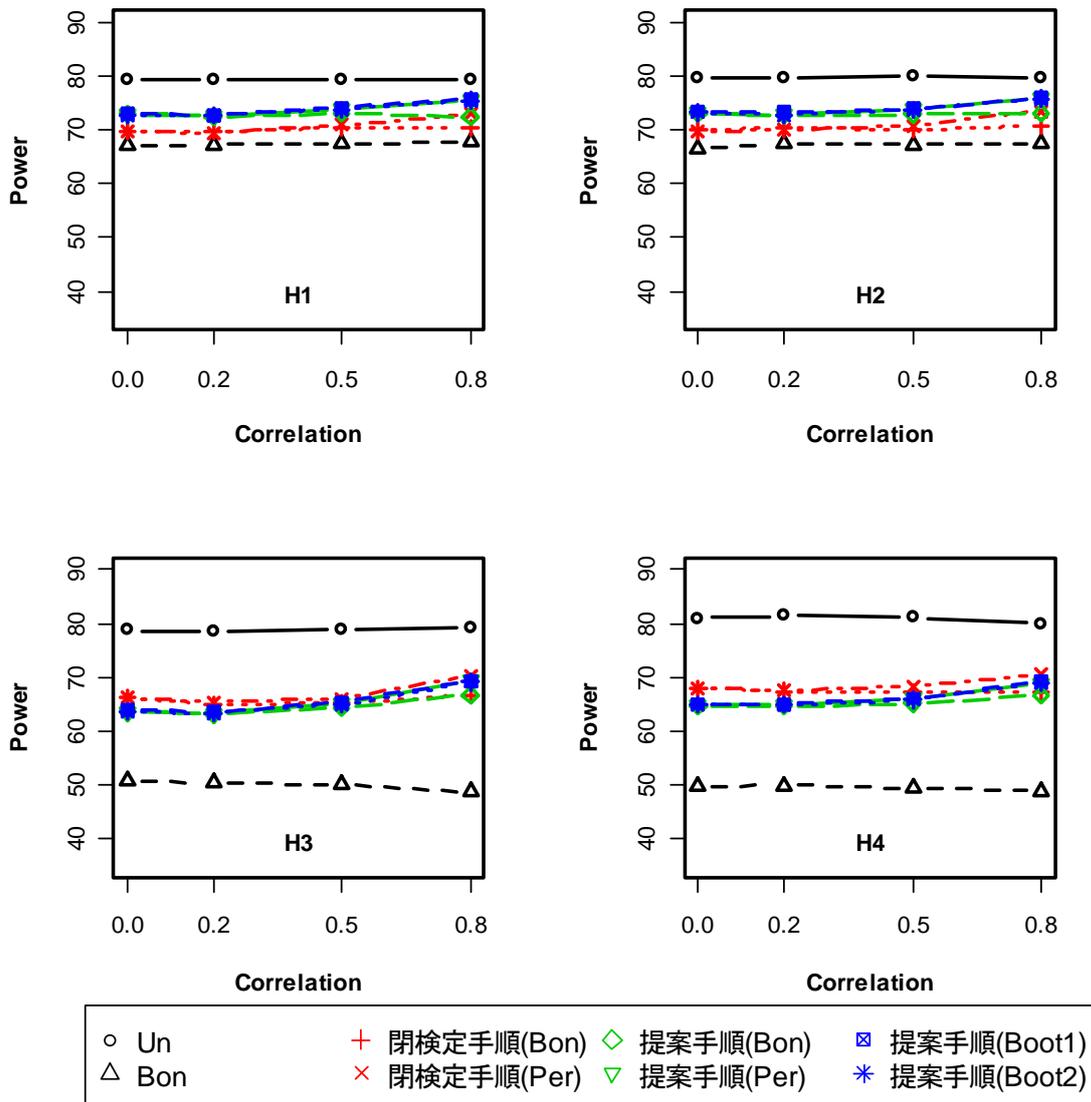
Un : 多重性を調整しない方法, Bon : ボンフェローニ法, Per : 並べ替え法
 Boot1 : ブートストラップ法 (一様分布)
 Boot2 : ブートストラップ法 (並べ替え分布)

図 8 仮説族ごとの検出力 (連続変数、条件 A)



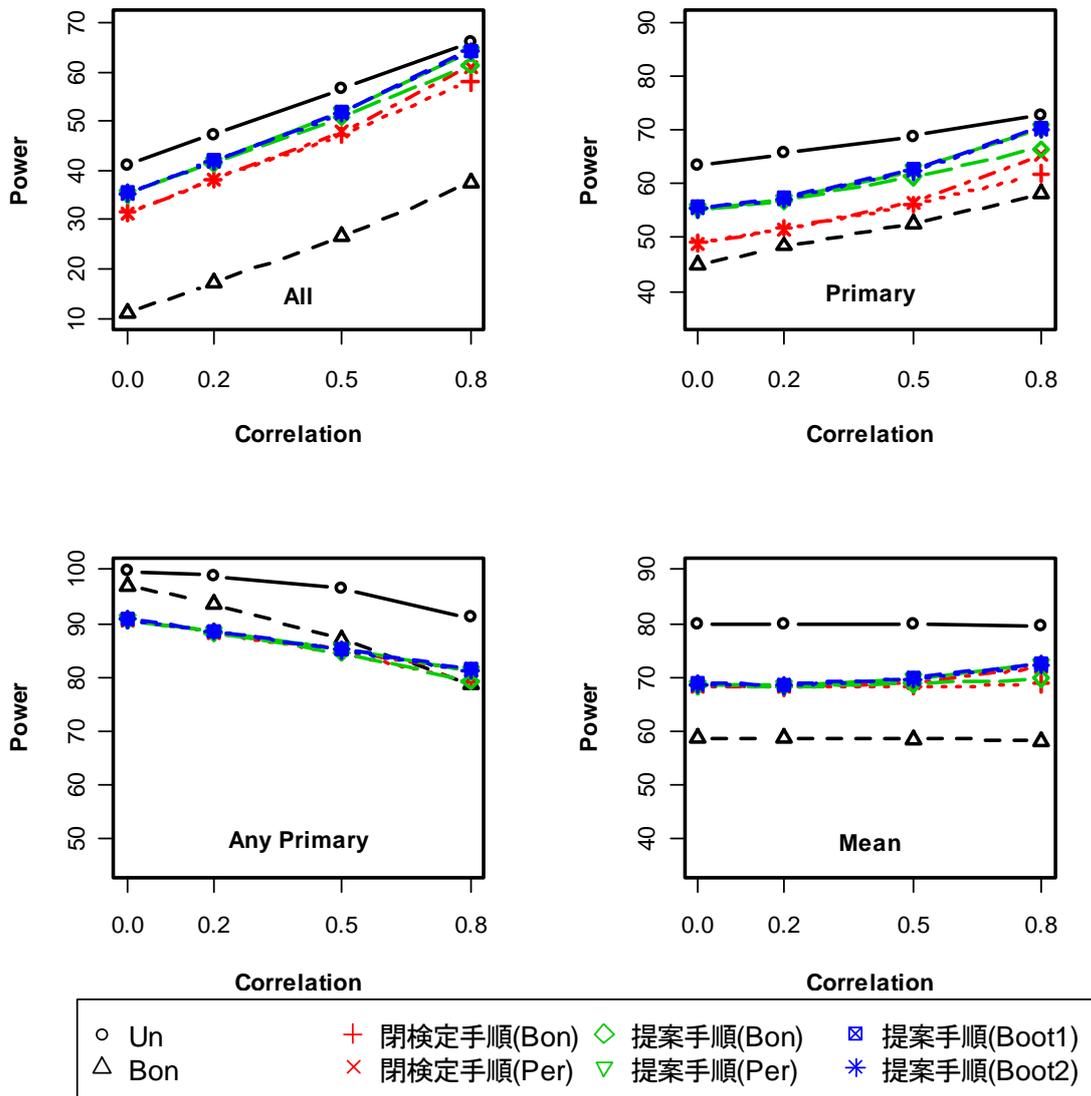
Un : 多重性を調整しない方法, Bon : ボンフェローニ法, Per : 並べ替え法
 Boot1 : ブートストラップ法 (一様分布)
 Boot2 : ブートストラップ法 (並べ替え分布)
 All : 全エンドポイントの棄却確率
 Primary : 全プライマリエンドポイントの棄却確率
 Any Primary : 任意のプライマリエンドポイントの棄却確率
 Mean : 平均検出力

図 9 多重性を考慮した検出力 (連続変数、条件 A)



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)

図 10 仮説族ごとの検出力 (連続変数、条件 B)



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)
 All : 全エンドポイントの棄却確率
 Primary : 全プライマリ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
 Mean : 平均検出力

図 11 多重性を考慮した検出力 (連続変数、条件 B)

表 7、表 9、図 8、図 10 の仮説ごとの検出力に注目する。多重性を調整しない方法と比較してボンフェローニ法は著しく低い値となった。同様に、ボンフェローニ法を適用したパラレルゲートキーピングプロシジャとの比較において、閉検定手順ではプライマリエンドポイントの検出力が条件 A で 6%、条件 B で 10%、セカンダリエンドポイントの検出力が 13-15%低下した。提案した手順ではプライマリエンドポイントの検出力が条件 A で 4-5%、条件 B で 6-7%、セカンダリエンドポイントの検出力が 15-16%低下した。また、提案した手順において、ボンフェローニ法、並べ替え法を適用した場合を比較すると、相関が低いときにボンフェローニ法がわずかながら検出力が大きく、相関 0.8 ではエンドポイント、条件によらず並べ替え法が 2-3%大きくなった。提案した手順に一様分布を用いたブートストラップ法を適用した場合、ボンフェローニ法、その他リサンプリング法を適用した場合より常に検出力が大きかった。

表 8、表 10、図 9、図 11 の多重性を考慮した検出力に注目する。ボンフェローニ法を適用したパラレルゲートキーピングプロシジャにおいて、閉検定手順と提案した手順を比較すると、条件 A では全エンドポイントを棄却する確率の違いは大きくなかったが、全プライマリエンドポイントを棄却する確率は 3-4%提案した手順で大きかった。条件 B では全エンドポイントを棄却する確率は 3-4%、全プライマリエンドポイントを棄却する確率は 5-6%提案した手順で大きかった。任意のプライマリエンドポイントを棄却する確率に違いはなかった。提案した手順において、ボンフェローニ法と並べ替え法を比較すると、相関 0.8 のとき、条件 B では全エンドポイント、全プ

ライマリエンドポイントを棄却する確率が 3-4%並べ替え法で上昇した。ブートストラップ法でも同様の上昇が見られた。検定手順、適用した検定方法の違いによる検出力の違いは、仮説ごとの検出力と同様の傾向が見られた。

全て 2 値変数の設定の検出力を、表 11 から表 14、図 12 から図 15 に示す。

表 11 仮説ごとの検出力 (2 値変数、条件 A)

		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
相関 0	H ₁ (Family1)	79.7	67.7	74.4	76.6	75.8	77.6	75.8	77.5
	H ₂ (Family1)	78.5	67.8	73.7	75.8	75.0	77.2	75.0	77.1
	H ₃ (Family2)	82.8	51.1	68.7	72.7	66.0	70.1	66.2	70.1
	H ₄ (Family2)	80.2	48.8	66.8	70.6	63.4	67.6	63.4	67.5
相関 0.2	H ₁ (Family1)	79.7	67.7	73.8	75.8	75.1	76.6	75.6	76.6
	H ₂ (Family1)	78.3	68.5	73.9	75.4	75.1	76.3	75.3	76.4
	H ₃ (Family2)	82.1	51.3	68.5	71.0	65.8	69.0	66.2	68.8
	H ₄ (Family2)	80.0	50.8	66.8	69.7	64.5	67.3	64.8	67.5
相関 0.5	H ₁ (Family1)	79.7	67.7	73.6	76.2	74.9	77.2	76.1	77.2
	H ₂ (Family1)	80.3	69.7	74.5	76.9	75.2	77.8	76.1	77.6
	H ₃ (Family2)	81.3	50.7	67.2	71.4	65.5	69.7	67.1	69.5
	H ₄ (Family2)	80.3	49.5	65.9	70.3	64.3	68.4	65.9	68.4
相関 0.8	H ₁ (Family1)	79.7	67.7	73.1	77.7	73.9	78.0	76.7	77.8
	H ₂ (Family1)	80.1	68.4	73.2	78.2	74.1	78.8	76.9	78.3
	H ₃ (Family2)	80.3	48.4	67.4	73.9	66.0	71.3	69.4	71.0
	H ₄ (Family2)	80.9	49.1	68.0	74.4	67.1	72.1	70.0	72.0

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 12 多重性を考慮した検出力 (2 値変数、条件 A)

	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0	All	40.9	11.5	39.1	40.2	39.4	40.8	39.4	40.6
	All Primary	62.7	46.1	57.0	59.2	59.9	61.7	59.9	61.5
	Any (Any Primary)	99.9 (95.5)	97.4 (89.5)	91.0	93.2	91.0	93.2	91.0	93.2
	Mean	80.3	58.9	70.9	74.0	70.1	73.2	70.1	73.1
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.2	All	47.4	18.0	45.2	46.7	45.7	46.9	45.9	46.9
	All Primary	64.8	49.3	58.7	61.3	61.3	62.7	61.6	62.8
	Any (Any Primary)	99.2 (93.2)	94.6 (87.0)	89.0	90.0	89.0	90.2	89.3	90.2
	Mean	80.0	59.6	70.8	73.0	70.2	72.3	70.5	72.3
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.5	All	55.9	26.5	53.1	55.0	53.3	55.3	54.3	55.2
	All Primary	68.7	53.8	62.5	65.5	64.5	67.2	65.9	67.0
	Any (Any Primary)	97.0 (91.4)	88.5 (83.6)	85.6	87.8	85.6	87.8	86.4	87.8
	Mean	80.4	59.4	70.3	73.7	70.0	73.3	71.3	73.2
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.8	All	66.5	37.4	63.1	66.1	63.5	66.3	65.6	66.1
	All Primary	72.6	58.7	66.6	71.2	68.3	72.2	71.0	72.0
	Any (Any Primary)	91.4 (87.2)	79.2 (77.4)	79.6	84.7	79.6	84.7	82.7	84.2
	Mean	80.3	58.4	70.4	76.0	70.3	75.1	73.3	74.8

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

Mean: 平均検出力

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 13 仮説ごとの検出力 (2 値変数、条件 B)

		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
相関 0	H ₁ (Family1)	79.7	67.7	70.7	72.2	73.5	76.0	73.5	75.9
	H ₂ (Family1)	78.5	67.8	70.7	72.0	72.8	75.5	72.9	75.5
	H ₃ (Family2)	82.8	51.1	68.7	72.7	66.0	70.1	66.2	70.1
	H ₄ (Family2)	80.2	48.8	66.8	70.6	63.4	67.6	63.4	67.5
相関 0.2	H ₁ (Family1)	79.7	67.7	70.7	72.2	73.3	75.5	73.9	75.4
	H ₂ (Family1)	78.3	68.5	71.5	72.4	74.0	75.4	74.2	75.4
	H ₃ (Family2)	82.1	51.3	68.5	71.0	65.8	69.0	66.2	68.8
	H ₄ (Family2)	80.0	50.8	66.8	69.7	64.5	67.3	64.8	67.5
相関 0.5	H ₁ (Family1)	79.7	67.7	70.7	73.3	73.3	76.4	74.7	76.4
	H ₂ (Family1)	80.3	69.7	72.5	74.2	74.2	77.1	75.1	77.0
	H ₃ (Family2)	81.3	50.7	67.2	71.4	65.5	69.7	67.1	69.5
	H ₄ (Family2)	80.3	49.5	65.9	70.3	64.3	68.4	65.9	68.4
相関 0.8	H ₁ (Family1)	79.7	67.7	70.7	75.2	72.5	77.8	76.3	77.5
	H ₂ (Family1)	80.1	68.4	71.3	75.6	73.5	78.5	76.5	78.1
	H ₃ (Family2)	80.3	48.4	67.4	73.9	66.0	71.3	69.4	71.0
	H ₄ (Family2)	80.9	49.1	68.0	74.2	67.1	72.1	70.0	72.0

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 14 多重性を考慮した検出力 (2 値変数、条件 B)

	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 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 (Any Primary)	99.9 (95.5)	97.4 (89.5)	91.0	93.2	91.0	93.2	91.0	93.2
	Mean	80.3	58.9	69.2	71.9	68.9	72.3	69.0	72.3
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.2	All	47.4	18.0	39.8	41.0	43.0	45.1	43.4	45.1
	All Primary	64.8	49.3	53.1	55.0	58.3	60.7	58.8	60.7
	Any (Any Primary)	99.2 (93.2)	94.6 (87.0)	89.0	90.0	89.0	90.2	89.3	90.2
	Mean	80.0	59.6	69.4	71.3	69.4	71.8	69.8	71.8
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 0.5	All	55.9	26.5	48.3	50.1	51.0	54.1	52.3	54.1
	All Primary	68.7	53.8	57.6	60.3	62.0	65.6	63.5	65.6
	Any (Any Primary)	97 (91.4)	88.5 (83.6)	85.6	87.7	85.6	87.8	86.4	87.8
	Mean	80.4	59.4	69.1	72.3	69.3	72.9	70.7	72.8
	Un	シングル ステップ		閉検定手順		提案手順			
		Bon	Bon	Per	Bon	Per	Boot1	Boot2	
相関 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 (Any Primary)	91.4 (87.2)	79.2 (77.4)	79.6	83.9	79.6	84.7	82.7	84.2
	Mean	80.3	58.4	69.3	74.7	69.8	75.0	73.1	74.7

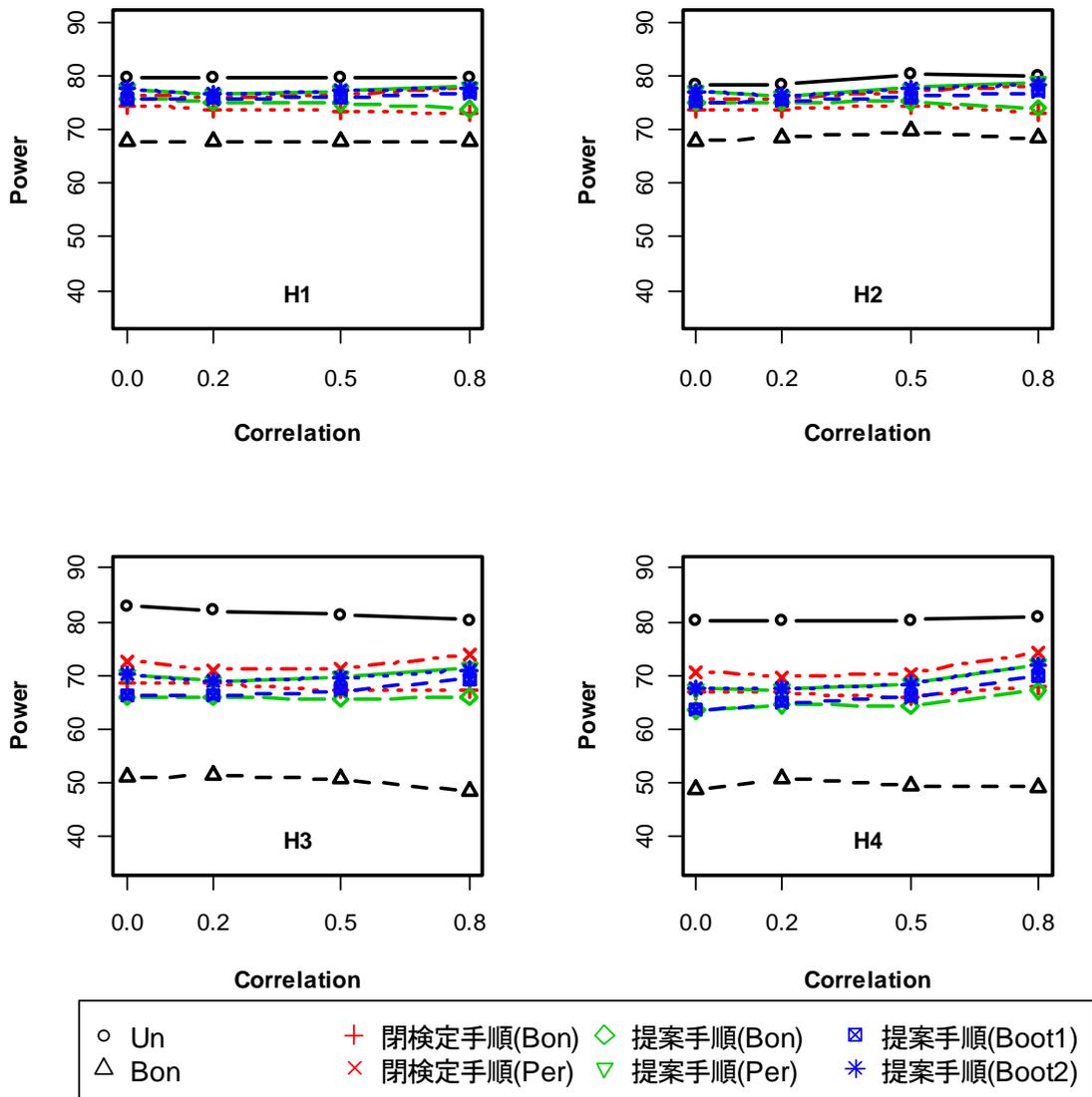
All: 全エンドポイントの棄却確率, All Primary: 全プライマリエンドポイントの棄却確率

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

Mean: 平均検出力

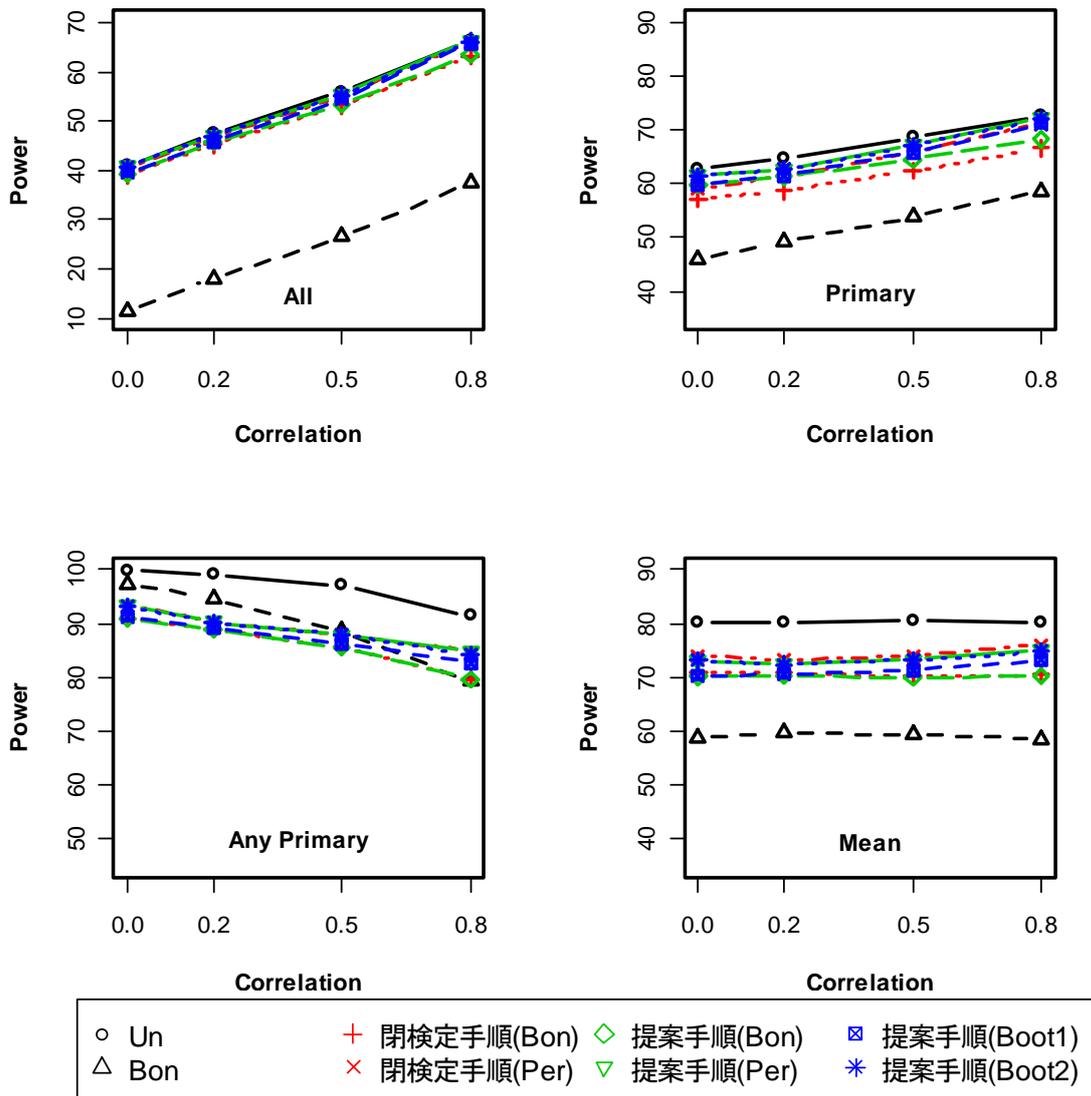
Un: 多重性を調整しない方法, Bon: ボンフェローニ法

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)



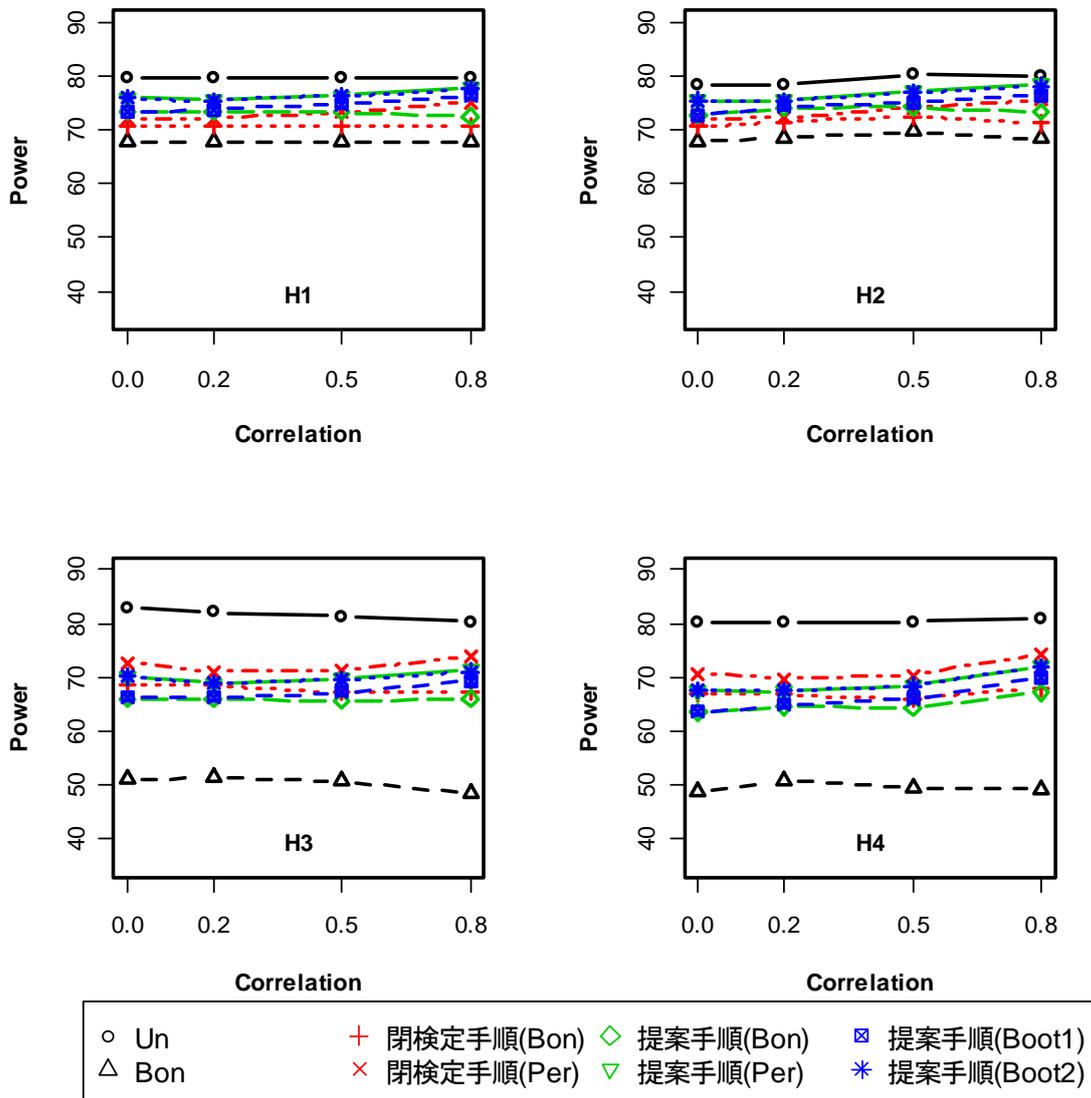
Un : 多重性を調整しない方法, Bon : ボンフェローニ法, Per : 並べ替え法
 Boot1 : ブートストラップ法 (一様分布)
 Boot2 : ブートストラップ法 (並べ替え分布)

図 12 仮説族ごとの検出力 (2 値変数、条件 A)



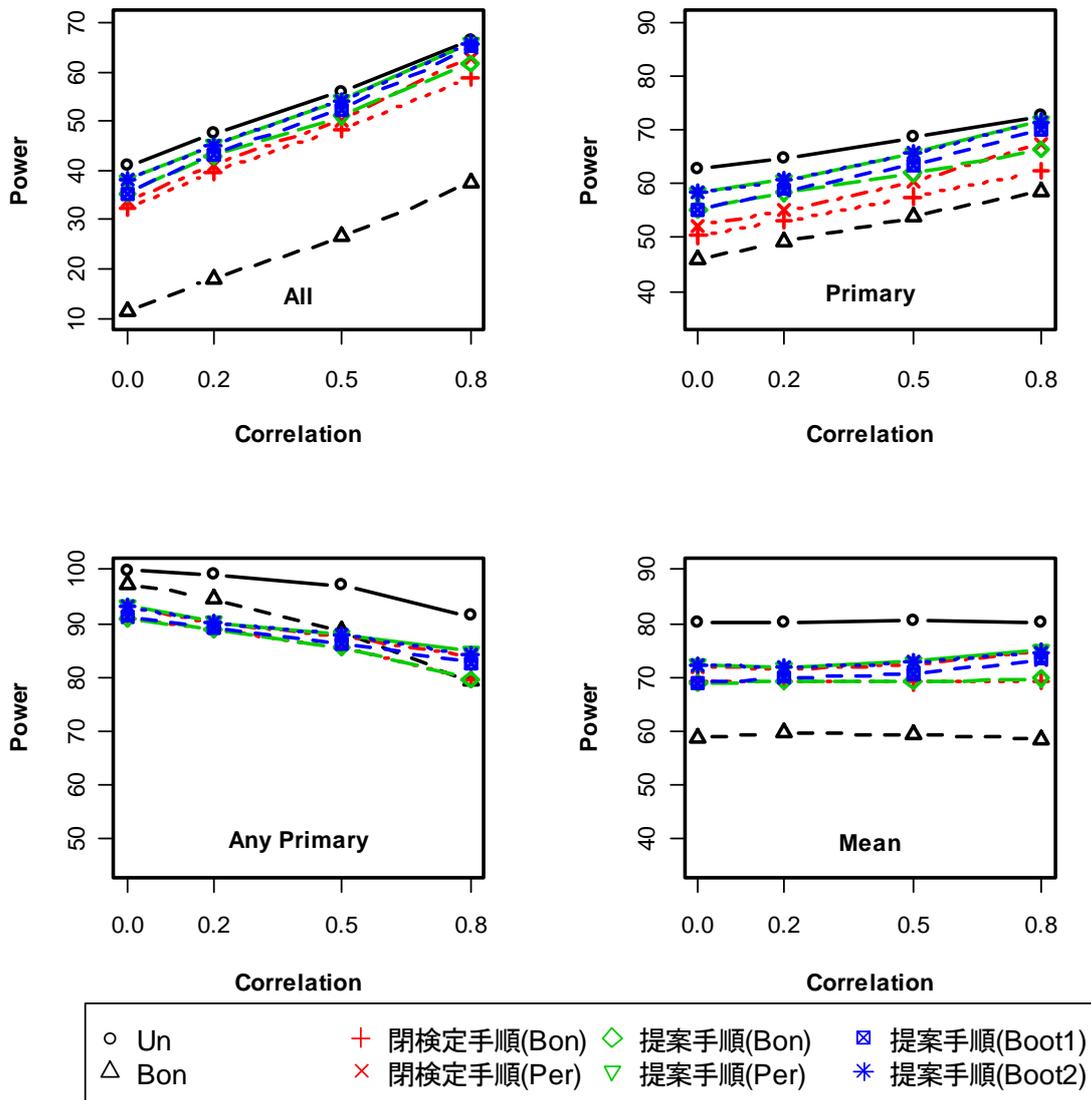
Un : 多重性を調整しない方法, Bon : ボンフェローニ法, Per : 並べ替え法
 Boot1 : ブートストラップ法 (一様分布)
 Boot2 : ブートストラップ法 (並べ替え分布)
 All : 全エンドポイントの棄却確率
 Primary : 全プライマリエンドポイントの棄却確率
 Any Primary : 任意のプライマリエンドポイントの棄却確率
 Mean : 平均検出力

図 13 多重性を考慮した検出力 (2 値変数、条件 A)



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)

図 14 仮説族ごとの検出力 (2 値変数、条件 B)



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)
 All : 全エンドポイントの棄却確率
 Primary : 全プライマリ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
 Mean : 平均検出力

図 15 多重性を考慮した検出力 (2 値変数、条件 B)

表 11、表 13、図 12、図 14 の仮説ごとの検出力に注目する。多重性を調整しない方法と比較してボンフェローニ法は著しく低い値となった。同様に、ボンフェローニ法を適用したパラレルゲートキーピングプロシジャとの比較において、閉検定手順ではプライマリエンドポイントの検出力が条件 A で 5-7%、条件 B で 8-9%、セカンダリエンドポイントの検出力が 13-14%低下した。提案した手順ではプライマリエンドポイントの検出力が条件 A で 4-5%、条件 B で 6-7%、セカンダリエンドポイントの検出力が 14-16%低下した。また、提案した手順においてボンフェローニ法、並べ替え法を適用した場合を比較すると、エンドポイント、相関、条件によらず並べ替え法の検出力が大きく、相関 0.8 では 5%-6%大きかった。提案した手順では、相関が低いときに一様分布を用いたブートストラップ法がボンフェローニ法と挙動が等しく、並べ替え分布を用いたブートストラップ法が並べ替え法と常に挙動が等しかった。

表 12、表 14、図 13、図 15 の多重性を考慮した検出力に注目する。ボンフェローニ法を適用したパラレルゲートキーピングプロシジャを閉検定手順、提案した手順と比較すると、条件 A では全エンドポイントを棄却する確率の低下が小さく大きな差はなかったが、全プライマリエンドポイントを棄却する確率は 2-3%提案した手順で大きかった。条件 B では全エンドポイントを棄却する確率は 3%、全プライマリエンドポイントを棄却する確率は 4-5%提案した手順で大きかった。並べ替え法を適用した場合には、条件 B では全エンドポイントを棄却する確率が 4%、全プライマリエンドポイントを棄却する確率が 5-6%提案した手順で大きく、任意のプライマリエンドポ

イントを棄却する確率は提案した手順でわずかながら大きくなった。提案した手順において、ボンフェローニ法と並べ替え法を比較すると、相関によらず、全エンドポイント、全プライマリエンドポイントを棄却する確率が 3-5%並べ替え法で上昇した。並べ替え分布を用いたブートストラップ法でも同様の上昇が見られた。検定方法、検定手順の違いによる検出力の違いは、仮説ごとの検出力と同様の傾向が見られた。

2 値変数と連続変数の組み合わせの設定の検出力を、表 15 から表 18、図 16 から図 19 に示す。

表 15 仮説ごとの検出力 (2 値変数と連続変数、条件 A)

		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
相関 0	H ₁ (Family1)	79.7	67.7	74.4	75.0	75.9	76.3	76.1	76.3
	H ₂ (Family1)	80.2	70.0	75.1	76.3	76.6	77.8	77.0	77.8
	H ₃ (Family2)	82.8	51.1	69.5	71.5	67.5	69.6	68.0	69.5
	H ₄ (Family2)	80.8	49.8	67.6	69.8	65.0	66.5	65.2	66.5
相関 0.2	H ₁ (Family1)	79.7	67.7	73.9	74.6	75.3	76.1	75.8	76.1
	H ₂ (Family1)	80.5	70.8	75.8	76.8	76.7	78.0	77.0	78.1
	H ₃ (Family2)	82.1	51.5	68.2	70.4	65.7	67.8	66.1	68.0
	H ₄ (Family2)	81.2	49.6	66.8	69.1	64.1	66.5	64.5	66.6
相関 0.5	H ₁ (Family1)	79.7	67.7	74.0	75.0	75.2	76.2	75.8	76.2
	H ₂ (Family1)	81.0	70.7	75.1	76.5	76.2	77.8	76.9	78.0
	H ₃ (Family2)	81.1	51.2	68.2	70.3	66.3	68.3	67.3	68.3
	H ₄ (Family2)	81.8	50.1	68.8	71.4	66.4	68.7	67.5	68.7

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 16 多重性を考慮した検出力 (2 値変数と連続変数、条件 A)

		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
相関 0	All	41.9	11.7	39.5	40.4	40.2	40.8	40.5	40.6
	All Primary	63.7	47.3	57.6	59.0	60.7	61.5	61.0	61.5
	Any (Any Primary)	99.9 (96.2)	98.2 (90.3)	91.9	92.5	91.9	92.6	92.1	92.6
	Mean	80.9	59.7	71.7	73.2	71.3	72.5	71.6	72.5
			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
相関 0.2	All	47.6	18.4	45.4	46.1	45.7	46.9	45.9	46.9
	All Primary	65.5	50.6	60.1	60.9	62.4	63.4	62.7	63.4
	Any (Any Primary)	99.2 (94.7)	95.3 (87.9)	89.6	90.6	89.6	90.8	90.0	90.8
	Mean	80.9	59.9	71.2	72.7	70.5	72.1	70.9	72.2
			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
相関 0.5	All	56.4	27.5	54.2	55.1	54.5	55.5	54.9	55.4
	All Primary	68.7	54.6	62.9	64.3	65.3	66.6	65.9	66.6
	Any (Any Primary)	97.5 (92.0)	89.0 (83.8)	86.1	87.2	86.1	87.5	86.8	87.6
	Mean	80.9	59.9	71.5	73.3	71.0	72.8	71.9	72.8

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

Mean: 平均検出力

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 17 仮説ごとの検出力 (2 値変数と連続変数、条件 B)

		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
相関 0	H ₁ (Family1)	79.7	67.7	70.7	70.5	73.4	74.1	73.4	74.0
	H ₂ (Family1)	80.2	70.0	72.5	72.8	75.0	76.4	75.2	76.3
	H ₃ (Family2)	82.8	51.1	69.5	71.5	67.5	69.6	68.0	69.5
	H ₄ (Family2)	80.8	49.8	67.6	69.8	65.0	66.5	65.2	66.5
相関 0.2	H ₁ (Family1)	79.7	67.7	70.7	70.8	73.3	74.2	73.9	74.2
	H ₂ (Family1)	80.5	70.8	73.3	73.7	75.1	76.6	75.4	76.6
	H ₃ (Family2)	82.1	51.5	68.2	70.4	65.7	67.8	66.1	68.0
	H ₄ (Family2)	81.2	49.6	66.8	69.1	64.1	66.5	64.5	66.6
相関 0.5	H ₁ (Family1)	79.7	67.7	70.7	72.0	73.2	75.0	74.3	75.1
	H ₂ (Family1)	81.0	70.7	73.3	74.2	75.2	77.3	76.2	77.4
	H ₃ (Family2)	81.1	51.2	68.2	70.3	66.3	68.3	67.3	68.3
	H ₄ (Family2)	81.8	50.1	68.8	71.4	66.4	68.7	67.5	68.7

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 18 多重性を考慮した検出力 (2 値変数と連続変数、条件 B)

		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
相関 0	All	41.9	11.7	33.1	33.5	36.5	37.8	36.6	37.5
	All Primary	63.7	47.3	51.2	51.5	56.5	57.9	56.5	57.6
	Any (Any Primary)	99.9 (96.2)	98.2 (90.3)	91.9	92.5	91.9	92.6	92.1	92.6
	Mean	80.9	59.7	70.1	71.2	70.2	71.6	70.5	71.5
			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
相関 0.2	All	47.6	18.4	39.7	40.2	42.5	44.0	42.9	44.0
	All Primary	65.5	50.6	54.4	54.8	58.8	60.1	59.2	60.0
	Any (Any Primary)	99.2 (94.7)	95.3 (87.9)	89.6	90.6	89.6	90.8	90.0	90.8
	Mean	80.9	59.9	69.8	71.0	69.5	71.3	70.0	71.4
			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
相関 0.5	All	56.4	27.5	49.2	50.4	51.9	53.9	53.0	53.9
	All Primary	68.7	54.6	57.9	59.5	62.3	64.8	63.7	65.0
	Any (Any Primary)	97.5 (92.0)	89.0 (83.8)	86.1	87.2	86.1	87.5	86.8	87.6
	Mean	80.9	59.9	70.2	72.0	70.3	72.3	71.3	72.4

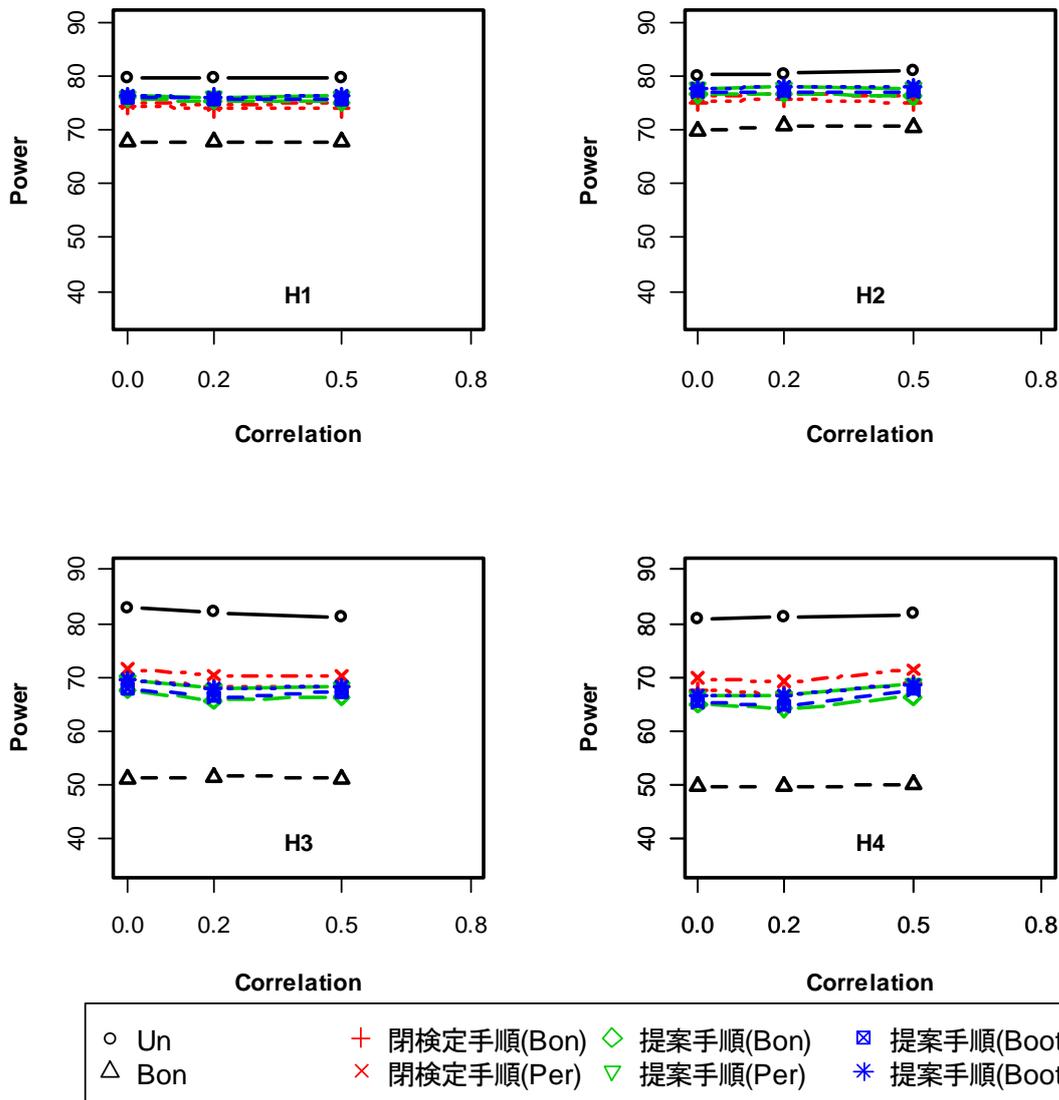
All: 全エンドポイントの棄却確率, All Primary: 全プライマリエンドポイントの棄却確率

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

Mean: 平均検出力

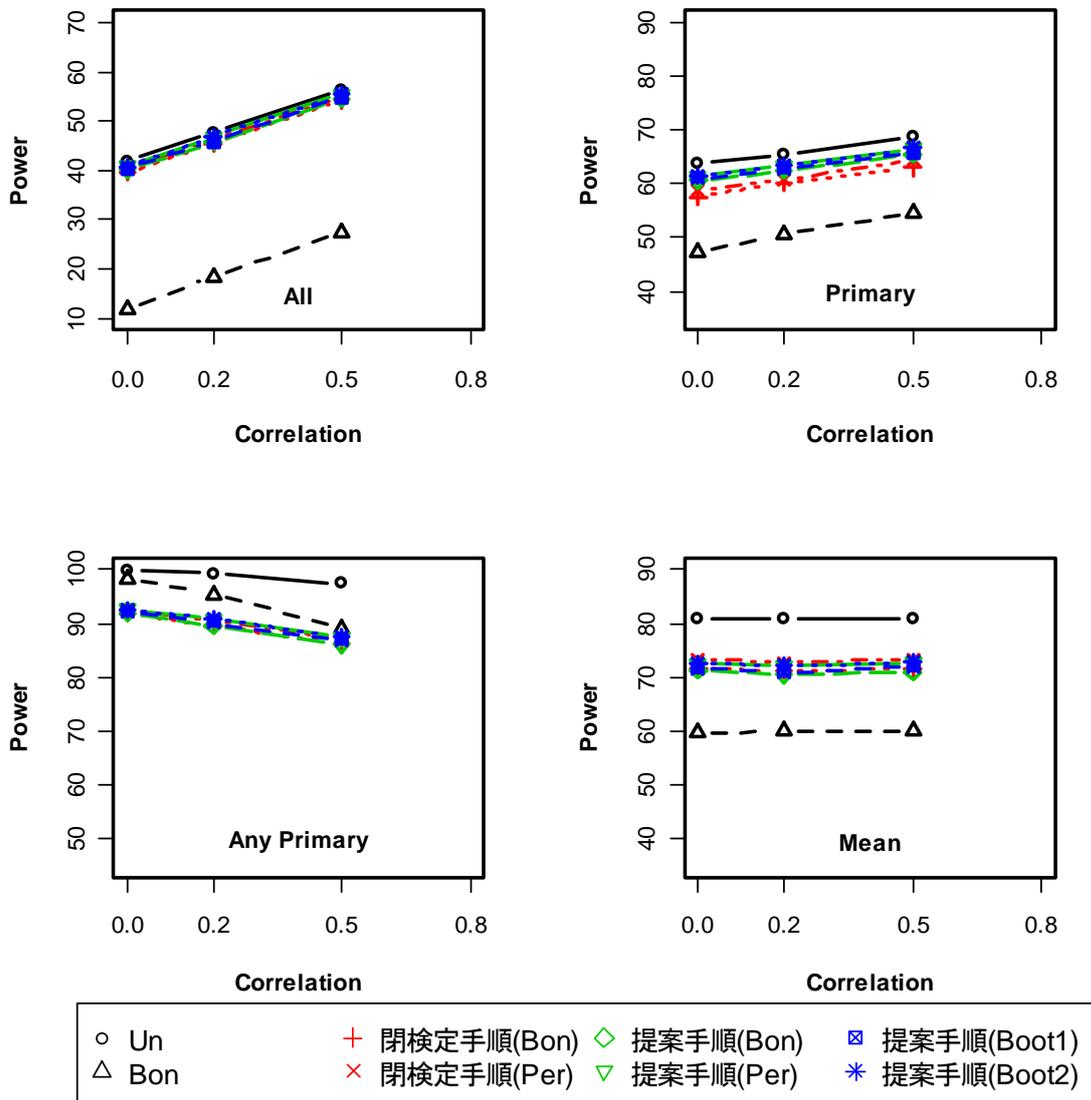
Un: 多重性を調整しない方法, Bon: ボンフェローニ法

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)



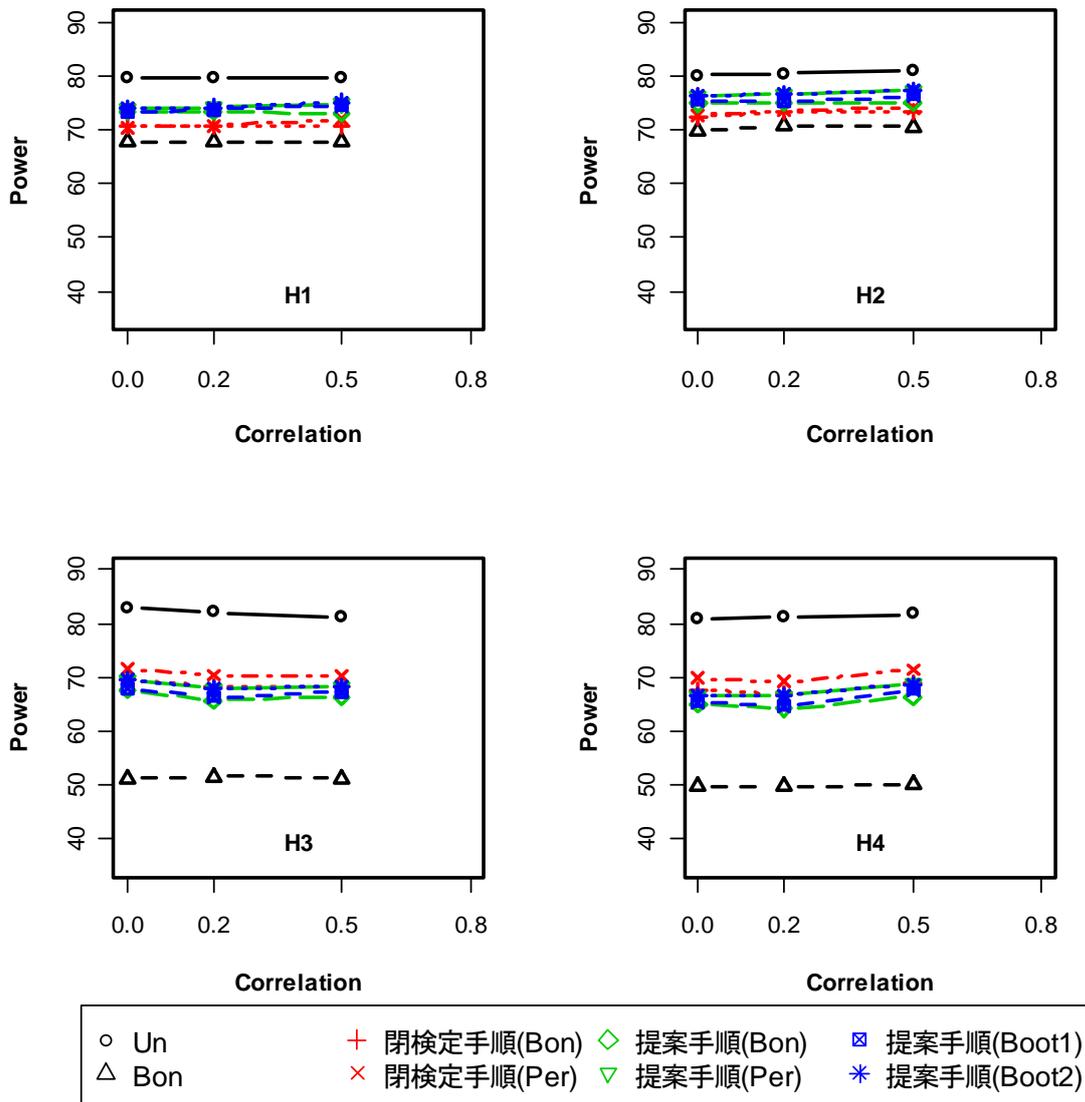
Un : 多重性を調整しない方法, Bon : ボンフェローニ法, Per : 並べ替え法
 Boot1 : ブートストラップ法 (一様分布)
 Boot2 : ブートストラップ法 (並べ替え分布)

図 16 仮説族ごとの検出力 (2 値変数と連続変数、条件 A)



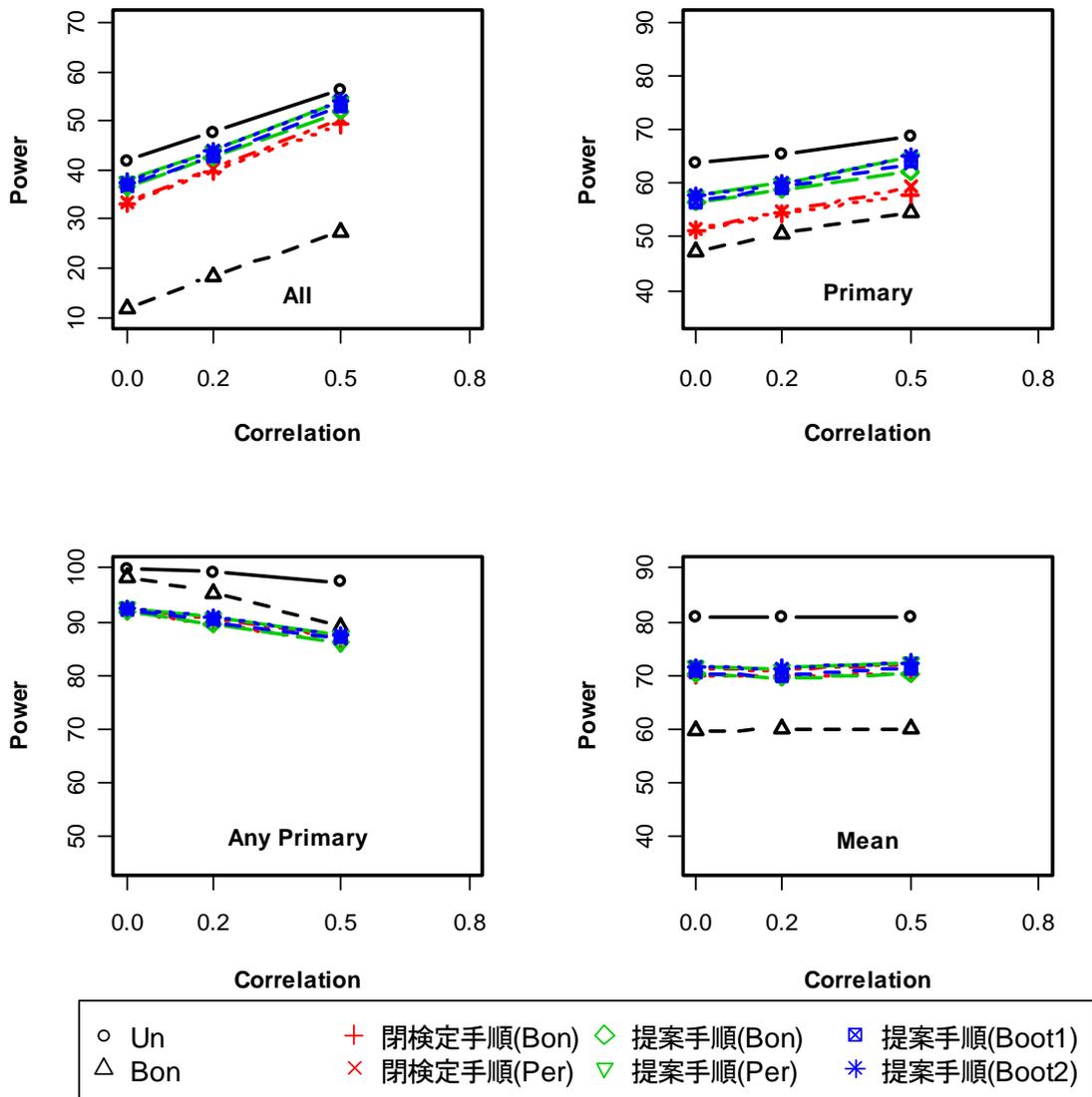
Un : 多重性を調整しない方法, Bon : ボンフェローニ法, Per : 並べ替え法
 Boot1 : ブートストラップ法 (一様分布)
 Boot2 : ブートストラップ法 (並べ替え分布)
 All : 全エンドポイントの棄却確率
 Primary : 全プライマリエンドポイントの棄却確率
 Any Primary : 任意のプライマリエンドポイントの棄却確率
 Mean : 平均検出力

図 17 多重性を考慮した検出力 (2 値変数と連続変数、条件 A)



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)

図 18 仮説族ごとの検出力 (2 値変数と連続変数、条件 B)



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)
 All : 全エンドポイントの棄却確率
 Primary : 全プライマリ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
 Mean : 平均検出力

図 19 多重性を考慮した検出力 (2 値変数と連続変数、条件 B)

表 15、表 17、図 16、図 18 の仮説ごとの検出力に注目する。多重性を調整しない方法と比較してボンフェローニ法は著しく低い値となった。同様に、ボンフェローニ法を適用したパラレルゲートキーピングプロシジャとの比較において、閉検定手順ではプライマリエンドポイントの検出力が条件 A で 5-6%、条件 B で 7-9%、セカンダリエンドポイントの検出力が 13-14%低下した。提案した手順ではプライマリエンドポイントの検出力が条件 A で 4-5%、条件 B で 5-6%、セカンダリエンドポイントの検出力が 15-17%低下した。また、提案した手順においてボンフェローニ法、並べ替え法を適用した場合を比較すると、相関 0.5 ではエンドポイント、条件によらず並べ替え法が 2%大きくなった。提案した手順において、並べ替え法と並べ替え分布を用いたブートストラップ法は常に挙動がほぼ等しく、一様分布を用いたブートストラップ法の挙動はボンフェローニ法と並べ替え法の間となった。

表 16、表 18、図 17、図 19 の多重性を考慮した検出力に注目する。ボンフェローニ法を適用したパラレルゲートキーピングプロシジャを閉検定手順、提案した手順と比較すると、条件 A では全プライマリエンドポイントを棄却する確率が 2-3%、条件 B では全エンドポイントを棄却する確率が 2-3%、全プライマリエンドポイントを棄却する確率が 4-5%提案した手順で大きかった。並べ替え法を適用した場合は、条件 B では全エンドポイントを棄却する確率が 4%、全プライマリエンドポイントを棄却する確率が 5-6%提案した手順で大きく、任意のプライマリエンドポイントを棄却する確率は提案した手順でわずかながら大きくなった。提案した手順において、ボン

フェローニ法と並べ替え法を比べると、相関 0.5 のとき、全エンドポイント、全プライマリエンドポイントを棄却する確率が 2%上昇していた。並べ替え分布を用いたブートストラップ法でも同様の上昇が見られた。検定方法、検定手順の違いによる検出力の違いは、仮説ごとの検出力と同様の傾向が見られた。

4. 考察

本論文では、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し、シミュレーション実験により提案法が FWER を名義水準以下に保ち、既存の方法に比べて検出力が高い方法であることを確認した。

ARDS に対する治療効果が死亡率、人工呼吸器の使用期間などにあると考えられるように、疾患の複数の側面、複数の疾患に対して治療効果があることは多い。そのため、既存の臨床試験のように、1つのプライマリエンドポイントのみで検証的に治療効果を評価することは効率が悪い。コストを掛けずにより多くの検証的な結果を得るため、複数のプライマリエンドポイント、複数のセカンダリエンドポイントを用いて治療効果を評価することが重要である。

その際、治療効果の推定値や信頼区間に基づいて治療効果の有無を判断すれば多重性を調整しなくてもよいという考え方が^{26,27}。しかし、複数のエンドポイントを用いる場合、治療効果が大きく見積られる可能性がある⁸⁹ため信頼区間を広くとる

べきだが、妥当な方法は提案されていない。得られる結果に対する誤った判断や言い過ぎは抑える必要がある。その方法が検定による多重性の調整である。有意性の判断や調整 p 値では治療効果の大きさがわからないが、現実的には調整 p 値に加え、治療効果の推定値と信頼区間、その他の治療方法の有無、治療費用などを用いて総合的に治療効果の評価は行われる¹⁵。この評価体系では、少ない情報から過度に多くの結果を得ることに対する罰則を与える。例えば、治療効果は大きい有意とならない結果は言い過ぎの可能性があるので追加の研究を求める。このように、多重性を調整することには意義がある。しかし、多重性を調整する際はエンドポイントの重要度による仮説の順序性を考慮することが大切である。セカンダリエンドポイントを治療効果の検証的な評価に用いる場合では、単独で治療効果を示す変数であるプライマリエンドポイントの検定から行うべきである。このような仮説の順序性を考慮した多重性の調整方法としてゲートキーピングプロシジャが注目されている。ゲートキーピングプロシジャを臨床試験に利用する際には手順の煩雑さを解消し、検出力を向上させることが重要である。

本研究で提案したゲートキーピングプロシジャは、3つの点で有用な方法である。1つは提案した検定手順が簡便で、実用性が高い点である。閉検定手順を用いる既存の方法は、全ての共通帰無仮説に関する検定を考えなければいけない。 M 個のエンドポイントを設定した場合、 $2^M - M - 1$ 個の共通帰無仮説ごとに未調整 p 値に対する重みを考え、それぞれの共通帰無仮説の検定を行う必要があるため煩雑な手順となること

が避けられない。このように複雑な方法では、エンドポイント数が違う場合でも対応できる汎用的なマクロプログラムを作ることが難しい。また、エンドポイントの設定、重みの設定、プロシジャの選択など臨床家の意見を必要とする場面での負担が大きい。提案した手順は、重みの設定は1つの組み合わせを考えるだけでよく、 M 回の検定を行うだけでよい。また、全てのプライマリエンドポイントを棄却したらセカンダリエンドポイントの検定を行う、プライマリエンドポイントを1つでも棄却したらセカンダリエンドポイントの検定を行うという手順が明確である。これらの特徴は検定手順の理解に繋がり、臨床試験をデザインする際の臨床家とのコミュニケーションにおける重要な利点となる。このことから、提案した手順は臨床家の負担が少ない実用的な方法であると言える。リサンプリング法を適用する簡潔な検定手順はシリアルゲートキーピングプロシジャのみ提案されている手順が⁷⁵適切になるが、本研究でパラレルゲートキーピングプロシジャ、特に仮説の順序性をより考慮した条件 **B** を満たす検定手順を提案したことは意義がある。

2つ目は閉検定手順を用いる場合に比べ提案した手順はプライマリエンドポイントを棄却する確率が大きくなる点である。理由の1つに未調整 p 値に対する重みの設定がある。提案した手順では、プライマリエンドポイントと比較してセカンダリエンドポイントに大きな重みを与えることはない。一方、**Dmitrienko** らの重みの与え方⁵¹はセカンダリエンドポイントに過度に大きな重みを与え、プライマリエンドポイントが棄却されにくいことがある。この問題を防ぐには共通帰無仮説ごとの重みを細かく

設定しなければいけない^{71,72}ため、既存の方法は実用的でない。もう1つの理由は、不必要な多重性の調整を除いたことである。パラレルゲートキーピングプロシジャは、少なくとも1つのプライマリエンドポイントの仮説が棄却されなければセカンダリエンドポイントは多重性の調整に寄与しない。閉検定手順はこの条件を満たさないの
で過度に多重性の調整してしまう。一方、提案した手順は条件を満たす適切な多重性の調整を行うため検出力が向上すると考えられる。臨床試験の第一目的がプライマリエンドポイントの検定の有意性であることを考えれば、プライマリエンドポイントを棄却する確率を向上する提案した手順は有用であると考えられる。

3つ目はブートストラップ法の適用による検出力の向上とその柔軟性である。ボンフェローニ法の適用に比べ、同時帰無分布の相関構造や離散性を利用するブートストラップ法は検出力が向上する。また、並べ替え法も同時帰無分布を利用するものの、2群間で分散が等しくない場合は妥当な検定でなく^{78,84}、多群比較の場合は各手順で同時帰無分布の推定を行う必要がある。そのため、コンピュータ負荷や時間的制約を考えると実用的な方法とは言えない。一方、周辺分布が特定されているときのブートストラップ法は、2群間で分散が等しくない、多群比較を行うなどのサブセットピボタリティが成立しない場合にも一度推定した同時帰無分布を利用して全ての検定が行えるため、応用範囲が広い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、検出力を向上するかをシミュレーション実験により評価した。**FWER** はいずれの設定でも名義水準に保たれた。条件 A、B

と違う手順を用いた場合でも等しい結果が得られた理由は、全てのエンドポイントで治療効果がないという設定が影響していると考えられる。どちらの条件もプライマリエンドポイントが棄却されないとセカンダリエンドポイントは棄却されないので、**FWER** はプライマリエンドポイントを誤って棄却することのみで決定される。そのため、いずれの条件でも **FWER** が等しくなった。また、2 群間の相関構造が違う設定で、仮定が成立しない並べ替え法の **FWER** が名義水準に保たれたのは、エンドポイント数が少なく相関構造の影響が少ないことが理由として考えられる。**Huang** らの指摘⁸¹や **Chen** らのシミュレーション⁸⁴は、周辺分布の平均は等しいが分散が違う場合を主な問題として挙げていた。本研究のように周辺分布は等しいが相関構造のみ違う場合はさらなる検討が必要であると考えられる。

検出力を手順の違いに関して比較すると、仮説ごとの検出力、全プライマリエンドポイントを棄却する確率が、閉検定手順より提案した手順で高かった。特に条件 **B** においてプライマリエンドポイントに関する検出力が高いことは、提案した手順の有用性を示す結果であると言える。このことは、3 種類のパラレルゲートキーピングプロシジャの検出力を比較した研究⁸⁶において、3-5%程度の検出力の上昇をもって手順の有用性を示していることから推測される。また、任意のエンドポイントが棄却される確率が向上したことから、提案した手順が仮説の順序性を適切に考慮していることが示唆される。本研究のシミュレーション実験の設定では差がわずかであったが、エンドポイント数の増加、エンドポイント間の相関構造の相違など、より複雑な構造

をしたデータにおいて提案した手順の良さが示されると考えられる。これは今後の検討課題である。

提案した手順において、各手順に適用する検定方法に関して検出力を比較すると、全ての検出力がボンフェローニ法に比べ、リサンプリング法で高かった。特に2値変数における違いが顕著に表れた。このことは、提案した手順と同様、リサンプリング法を適用することの有用性を示唆していると考えられる。一方、同時分布の相関構造の違いによる検出力の違いは相関が0.8と強くない限り影響は大きくなかった。各手順の検定で同時に考慮するエンドポイント数は2、3であり、この程度の数では相関の影響がさほど表れなかったと考えられる。また、並べ替え法とブートストラップ法を比較すると、並べ替え分布を用いたブートストラップ法は全ての設定で並べ替え法と等しく、一様分布を用いた場合は全て連続変数の設定で一様に高かった。本研究でのシミュレーション実験の設定はサブセットピボタリティと同時交換可能性が成立しており、並べ替え法は適切な検定方法である。このことから、周辺分布が適切に特定されたブートストラップ法を適用することが有効であることが示唆される。

ブートストラップ法を適用する際の周辺分布の違いは、特定した分布の情報の違いが検出力に影響する結果となった。全て連続変数の設定では、一様分布を用いた場合が最も検出力が高かった。また、相関が低いときに並べ替え分布を用いた場合の検出力がボンフェローニ法よりもわずかながら低下した。一様分布は正確なパラメトリック分布、並べ替え分布は正確なノンパラメトリック分布である。t検定に対する多重

性の調整では、連続変数が独立な場合、ボンフェローニ法はパラメトリックな方法として考えられる。以上のことから、より正確に周辺分布を特定することによる情報量の増加が検出力の上昇に寄与することが示唆される。これは、全て2値変数の設定で並べ替え分布を用いた場合の検出力が高かったことから示される。一様分布は漸近的に正確なパラメトリック分布、並べ替え分布は有限集団における正確なノンパラメトリック分布である。連続変数の場合と同様、対象集団における周辺分布の特定の重要性を示していると言える。

シミュレーション実験の結果から、提案した手順にブートストラップ法を適用することは検出力が向上する有用な方法であることがわかった。このことは、閉検定手順にボンフェローニ法を適用した場合と比べ、全てのプライマリエンドポイントを棄却する確率が約10%向上することから示唆される。臨床試験をデザインする際、検出力は80%、90%など10%区切りが慣例として利用されるため、10%の差には意味があると言える。また、ブートストラップ法の適用は並べ替え法より検出力が向上することがあり、同時帰無分布の推定に利用する周辺分布の特定が検出力に影響することがわかった。

本研究の限界として、ブートストラップ法を有限集団の同時分布推定に用いることの妥当性が十分に検証されていない点がある。van der Laanらは小標本においても利用可能であると述べている⁸⁰が厳密な証明はない。ブートストラップ法を適用することが妥当であるかは、様々な設定のシミュレーション実験により確認していく必要

がある。本研究のシミュレーション実験は人数、エンドポイント数、エンドポイントの型、相関構造が限定されていた。現実のデータは群間の人数の違い、生存時間などのエンドポイントの利用、エンドポイント間の相関の相違があると考えられる。現実の状況に合わせたシミュレーション実験により提案法の性能を評価することが課題である。また、2 値変数の設定において、周辺分布をパラメトリックに特定する方法は示していない。2 値変数の検定の調整 p 値、重み付き p 値の周辺分布は超幾何分布から導出可能であるので、今後検討していきたい。

本研究で提案した方法の発展として、ツリーゲートキーピングプロシジャをシングルステップ法の繰り返しとして検定手順を導くことがある。ツリーゲートキーピングプロシジャはシリアルゲートキーピングプロシジャ、パラレルゲートキーピングプロシジャの組み合わせであり、提案した手順を拡張することで導くことが可能であると考えられる。また、提案法を用いる際はサンプルサイズ設計を考えなければいけない。シミュレーション実験によるサンプルサイズ設計は時間がかかるため、効率よく行う方法を考える必要がある。これらは今後の課題としたい。

5. 結論

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。シミュレーション実験により提案法が **FWER** を名義水準以下に保ち、既存の方法に比べて検出力が高い方法で

あることを示した。シミュレーション実験の結果から、エンドポイントが2値変数で相関が高いときに有用な方法であることが分かった。

6. 謝辞

本研究を進めていく上で大変有用な助言を下された東京大学大学院医学系研究科疫学・予防保健学分野の大橋靖雄教授、研究の方向性についての指導、ならびに研究内容についての議論をして下さった指導教員の松山裕准教授、及び研究室のスタッフ、大学院生の皆様に深く感謝申し上げます。

文献

1. Dmitrienko A, Offen WW, Westfall PH. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Statistics in Medicine*, vol.22, pp.2387–400, 2003.
2. 厚生労働省. 重篤副作用疾患別対応マニュアル. <http://www.info.pmda.go.jp/juutoku/file/jfm0611004.pdf>
3. CHMP. Guideline on medicinal products for the treatment of Alzheimer's disease and other dementias. <http://www.emea.europa.eu/pdfs/human/ewp/055395en.pdf>
4. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. The American College of Rheumatology preliminary definition of improvement in

- rheumatoid arthritis. *Arthritis and Rheumatism*, vol.38, pp.727–35, 1995.
5. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis: consensus development at OMERACT III. *the Journal of Rheumatology*, vol.24, pp.799–802, 1997.
 6. Cruccu G, Anand P, Attal N, Garcia-Larrea L, Haanpaa M, Jorum E, et al. EFNS guidelines for neuropathic pain assessment. *European Journal of Neurology*, vol.11, pp.153–62, 2004.
 7. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*, vol.113, pp. 9–19, 2005.
 8. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *Journal of Pain*, vol.9, pp.105–21, 2008.
 9. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, et al. Core outcome domains for chronic pain clinical trials: IMMPACT recommendations. *Pain*, vol.106, pp.337–45, 2003.
 10. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, et al. Developing outcome measures for pain clinical trials: IMMPACT recommendations. *Pain*, vol.125,

pp.208–15 2006.

11. Turk DC, Dworkin RH, McDermott MP, Bellamy N, Burke LB, Chandler JM, et al. Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. *Pain*, Vol.139, pp.485–493, 2008.
12. Bretz F, Branson M. Multiple endpoint. in Encyclopedia of Clinical Trials. D’Agostino RB, Sullivan LM, Massaro JM, editors, Volume 3, New York: Wiley. 534-4, 2005.
13. メマリー錠 5 mg CTD 第 2 部 CTD の概要 2.7 臨床概要 2.7.3 臨床的有効性.
<http://www.info.pmda.go.jp/shinyaku/P201100018/>
14. Sakamoto T, Ogawa H. "Just Make It Lower" Is an Alternative Strategy of Lipid-Lowering Therapy With Statins in Japanese Patients – LDL-Cholesterol: The Lower, the Better; Is It True for Asians? (Con) –. *Circulation Journal*, vol.74, pp.1731-41, 2010.
15. Moye LA. *Multiple analyses in clinical trials*. New York: Springer 2003.
16. 厚生省医薬安全局審査管理課長. 「臨床試験のための統計的原則」について(医薬審第 1047 号, 平成 10 年 11 月 30 日). http://www.pmda.go.jp/ich/e/e9_98_11_30.pdf
17. D’Agostino RB. Controlling alphas in a clinical trial: the case for secondary endpoints. *Statistics in Medicine*, vol.19, pp.763–6, 2000.
18. Hughes MD. Multiplicity in clinical trials. in Encyclopedia of Biostatistics 2nd ed.

- Armitage P, Colton T, editors, New York: Wiley. 3446-51, 2005.
19. Davis CE. Secondary endpoints can be validly analyzed, even if the primary endpoint does not provide clear statistical significance. *Controlled Clinical Trials*, vol.18, pp.557–60, 1997.
 20. Moye LA. End-point interpretation in clinical trials: the case for discipline. *Controlled Clinical Trials*, vol.20, pp.40–9, 1999.
 21. O’Neill RT. Secondary endpoints cannot be validity analyzed if the primary endpoint does not demonstrate clear statistical significance. *Controlled Clinical Trials*, vol.18, pp.550–6, 1997.
 22. Pocock SJ. Clinical trials with multiple outcomes: a statistical perspective on their design, analysis, and interpretation. *Controlled Clinical Trials*, vol.18, pp.530–45, 1997.
 23. Windhover Information Inc. Windhover’s *in vivo*: The business and medicine report. Brain drug economics model, 2003.
 24. Committee for Proprietary Medicinal Products (CPMP). Points to consider on multiplicity issues in clinical trials. London: European Agency for the Evaluation of Medicinal Products. <http://www.emea.europa.eu/pdfs/human/ewp/090899en.pdf>.
 25. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Controlled Clinical Trials*, vol.25, pp.535–52, 2004.

26. Chang M. Discussion of “Some controversial multiple testing problems in regulatory applications”. *Journal of Biopharmaceutical Statistics*, vol.19, pp.3–41, 2009.
27. Chang M. *Modern Issues and Methods in Biostatistics*. New York: Springer 2011.
28. Fisher LD. Carvedilol and the Food and Drug Administration (FDA) approval process: the FDA paradigm and reflections on hypothesis testing. *Controlled Clinical Trials*, vol.20, pp.16–39, 1999.
29. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of International Medicine*, vol.134, pp.663–94, 2001.
30. Bauer P, Chi G, Geller N, Gould AL, Jordan D, Mohanty S, et al. Industry, government, and academic panel discussion on multiple comparisons in a "real" phase three clinical trial. *Journal of Biopharmaceutical Statistics*, vol.13, pp.691–701, 2003.
31. Bender R, Lange S. Adjusting for multiple testing - when and how? *Journal of Clinical Epidemiology*, vol.54, pp.343-49, 2001.
32. Chi GYH. Multiple testing: multiple comparisons and multiple endpoints. *Drug Information Journal*, vol.32, pp.1347–62, 1998.
33. FDA Guidance for Industry. Clinical studies section of labeling for prescription drugs and biologics. www.fda.gov/cder/guidance/5534f1.htm.
34. Cutter GR, Baier ML, Rudick RA, Cookfair DL, Fischer JS, Petkau J, et al. Development

- of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*, vol.122, pp.871–82, 1999.
35. Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *the Journal of the American Medical Association*, vol.289, pp.2554–9, 2003.
 36. Huque MF, Alosch M, Bhore R. Addressing Multiplicity Issues of a Composite Endpoint and Its Components in Clinical Trials. *Journal of Biopharmaceutical Statistics*, vol.21, pp.611-34, 2011.
 37. Ferreira-González I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *British Medical Journal*, vol.334, pp.786, 2007.
 38. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics*, vol.40, pp.1079–87, 1984.
 39. Tang DI, Geller NL, Pocock SJ. On the design and analysis of randomized clinical trials with multiple endpoints. *Biometrics*, vol.49, pp.23–30, 1993.
 40. Lefkopoulou M, Ryan L. Global tests for multiple binary outcomes. *Biometrics*, vol.49, pp.975–88, 1993.
 41. Legler JM, Lefkopoulou M, Ryan LM. Efficiency and power of tests for multiple binary outcomes. *Journal of the American Statistical Association*, vol.90, pp.680–93, 1995.

42. Pocock S, Geller N, Tsiatis A. Analysis of multiple endpoints in clinical trials. *Biometrics*, vol.43, pp.487–98, 1987.
43. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. *Statistics in Medicine*, vol.16, pp.2529–42, 1997.
44. Hochberg Y, Tamhane AC. *Multiple Comparison Procedures*. New York Wiley: 1987.
45. Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, vol.67, pp.655–60, 1976.
46. Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, vol.62, pp.626–633, 1967.
47. Westfall PH, Young SS. *Resampling-based multiple testing: examples and methods for p-value adjustment*. New York: Wiley and Sons 1993.
48. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, vol.6, pp.65–70, 1979.
49. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, vol.75, pp.800–2, 1988.
50. Hommel G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, vol.75, pp.383–6, 1988.
51. Dmitrienko A, Molenberghs G, Chuang-Stein C, Offen W. *Analysis of clinical trials using SAS: a practical guide*. Cary NC: SAS Press 2005.

52. Troendle JF, Westfall PH. Permutational Multiple Testing Adjustments with Multivariate Multiple Group Data. *Journal of Statistical Planning and Inference*, vol.141, pp.2021–9, 2011.
53. Dmitrienko A, Tamhane AC, Bretz F. *Multiple testing in pharmaceutical statistics*. Boca Raton: Chapman and Hall/CRC 2009.
54. Huang X, Hsu J. Hochberg's step-up method: Cutting corners off Holm's stepdown method. *Biometrika*, vol.94, pp.965–75, 2007.
55. Sarkar SK, Chang CK. Simes' method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association*, vol.92, pp.1601–8, 1997.
56. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Statistics in Medicine*, vol.9, pp.811–8, 1990.
57. Troendle JF, Legler JM. A comparison of one-sided methods to identify significant individual outcomes in a multiple outcome setting: stepwise tests or global tests with closed testing. *Statistics in Medicine*, vol.17, pp.1245–60, 1998.
58. Lehman W, Wassmer G, Reitmeir P. Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics*, vol.47, pp.511–21, 1991.
59. Simes RJ. An improved Bonferroni procedure for multiple tests of significance.

- Biometrika*, vol.63, pp.655–60, 1986.
60. Bretz F, Maurer W, Gallo P. Discussion of “Some controversial multiple testing problems in regulatory applications.” *Journal of Biopharmaceutical Statistics*, vol.19, pp.25–34, 2009.
 61. Hsu J, Liu Y. Discussion of “Some controversial multiple testing problems in regulatory applications.” *Journal of Biopharmaceutical Statistics*, vol.19, pp.22–4, 2009.
 62. Hung HMJ, Wang SJ Challenges to multiple testing in clinical trials. *Biometrical Journal*, vol.52, pp.747–56, 2010.
 63. Hung HMJ, Wang SJ. Some controversial multiple testing problems in regulatory applications. *Journal of Biopharmaceutical Statistics*, vol.19, pp.1–11, 2009.
 64. Chuang-Stein C, Dmitrienko A., Offen W. Discussion of “Some controversial multiple testing problems in regulatory applications.” *Journal of Biopharmaceutical Statistics*, vol.19, pp.14–21, 2009.
 65. Wiens BL. A fixed-sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics*, vol.2, pp.211–15, 2003.
 66. Bauer P, Rohmel J, Maurer W, Hothorn, L. Testing strategies in multi-dose experiments including active control. *Statistics in Medicine*, vol.17, pp.2133-46, 1998.
 67. Westfall PH, Krishen A. Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference*, vol.99, pp.25-40, 2001.

68. Dmitrienko A., Tamhane AC. Gatekeeping procedures with clinical trial applications. *Pharmaceutical Statistics*, vol.6, pp.171–80, 2007.
69. Dmitrienko A, Tamhane AC, Liu L, Wiens BL. A note on tree-gatekeeping procedures in clinical trials. *Statistics in Medicine*, vol.27, pp.3446–51, 2008.
70. Dmitrienko A, Wiens BL, Tamhane AC, Wang X. Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Statistics in Medicine*, vol.26, pp.2465–78, 2007.
71. Dmitrienko A., Tamhane AC. Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. *Statistics in Medicine*, vol.30, pp.1473–88, 2011.
72. Dmitrienko A, Tamhane AC, Wiens BL. General multistage gatekeeping procedures. *Biometrical Journal*, vol.50, pp.667–77, 2008.
73. Grechanovsky E, Hochberg Y. Closed procedures are better and often admit a shortcut. *Journal of Statistical Planning and Inference*, vol.76, pp.79–91, 1999.
74. Dmitrienko A, Tamhane AC, Wang X, Chen X. Stepwise gatekeeping procedures in clinical trial applications. *Biometrical Journal*, vol.48, pp.984–91, 2006.
75. Hommel, G, Bretz F, Maurer W. Powerful short-cuts for multiple testing procedures with special reference to gatekeeping strategies. *Statistics in Medicine*, vol.26, pp.4063-73, 2007.
76. Westfall PH, Troendle JF. Multiple testing with minimal assumptions. *Biometrical*

- Journal*, vol.50, pp.745–55, 2008.
77. Dudoit S, van der Laan MJ, Pollard KS. Multiple testing. Part I. Single-step procedures for control of general Type I error rates. *Statistical Applications in Genetics and Molecular Biology*, vol.3, no.1, Article 13, 2004. (www.bepress.com/sagmb/vol3/iss1/art13.)
 78. Dudoit S, van der Laan MJ. *Multiple Testing Procedures and Applications to Genomics*. New York: Springer 2008.
 79. Romano JP, Wolf M. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, vol.100, pp.94-108, 2005.
 80. van der Laan MJ, Hubbard AE. Quantile-function based null distribution in resampling based multiple testing. *Statistical Applications in Genetics and Molecular Biology*, vol.5, no.1, Article 14, 2006. (www.bepress.com/sagmb/vol5/iss1/art14.)
 81. Huang Y, Xu H, Calian V Hsu JC. To permute or not to permute. *Bioinformatics*, vol22, pp.2244-48, 2006.
 82. Hsu JC. Multiplicity adjustment big and small in clinical studies. *Clinical Pharmacology & Therapeutics*, vol.88, pp.251–4, 2010.
 83. Angelis DD, Young GA. Bootstrap method. in *Encyclopedia of Biostatistics* 2nd ed. Armitage P, Colton T, editors, New York: Wiley, pp.534-41, 2005.
 84. Chen J, van der Laan MJ, Smith MT, Hubbard AE. A Comparison of Methods to Control

- Type I Errors in Microarray Studies. *Statistical Applications in Genetics and Molecular Biology*, vol.6, no.1, Article 28, 2007. (www.bepress.com/sagmb/vol6/iss1/art28.)
85. Lehmann EL, Romano JP, Shaffer JP. On optimality of stepdown and stepup multiple test procedures. *Annals of Statistics*, vol.33, pp.1084-108, 2005.
86. Dmitrienko A, Soulakova JN, Millen BA. Three methods for constructing parallel gatekeeping procedures in clinical trials, *Journal of Biopharmaceutical Statistics*, vol.21, pp.768-86, 2011.
87. Emrich LJ, Piedmonte MR. A method for generating high-dimensional multivariate binary variates. *the American Statistician*, vol.45, pp.302-4, 1991.
88. Bretz F, Maurera W, Hommel G. Test and power considerations for multiple endpoint analyses using sequentially rejective graphical procedures. *Statistics in Medicine*, vol.30, pp.1489-501, 2011.
89. Bretz F, Hothorn T, Westfall P. *Multiple comparisons using R*. Boca Raton, Chapman and Hall, 2011.

附録

附録 1-1 提案した検定手順 (パラレルゲートキーピングプロシジャ、条件 A)

#raw.p : 未調整 p 値、w : 重み、fn : ファミリー、B : リサンプリング数

#検定手順

```
o.p.1 <- order(iffelse(fn==1,raw.p/w,raw.p/0))[1]
```

```
adjp.qp[o.p.1] <- sum(iffelse(apply(qperms.w.p[,fn==1], 1, min)
```

```
<= min(w.p[fn==1]), 1, 0))/B;
```

```
fn <- iffelse(fn==1, 2, fn); fn[o.p.1] <- 1;
```

```
o.p.2 <- order(iffelse(fn==2,raw.p/w,raw.p/0))
```

```
adjp.qp[o.p.2[1]] <- sum(iffelse(apply(qperms.w.p[,fn==2], 1, min)
```

```
<= min(w.p[fn==2]), 1, 0))/B;
```

```
n2 <- length(fn[fn==2]); n2.1 <- length(fn[fn==2])-1; n2.2 <- length(fn[fn==2])-2;
```

```
for(i in 1:n2.2){
```

```
  j <- i + 1
```

```
  adjp.qp[o.p.2[j]] <- sum(iffelse(apply(qperms.w.p[,o.p.2[1:i]][,fn[o.p.2[1:i]]==2], 1,
```

```
  min) <= min(w.p[o.p.2[1:i]][fn[o.p.2[1:i]]==2]), 1, 0))/B;
```

```
}
```

```
adjp.qp[o.p.2[n2]] <- sum(iffelse(qperms.w.p[,o.p.2[1:n2.1]][,fn[o.p.2[1:n2.1]]==2]
```

```
<= min(w.p[o.p.2[1:n2.1]][fn[o.p.2[1:n2.1]]==2]), 1, 0))/B;
```

```
#調整 p 値
```

```
or <- c(o.p.1,o.p.2[1:n2])
```

```
for(i in 1:np.1){
```

```
  j <- i + 1
```

```
  if(adjp.qp[or[i]] > adjp.qp[or[j]]) {
```

```
    adjp.qp[or[j]] <- adjp.qp[or[i]]
```

```
  }
```

```
}
```

附録 1-2 提案した検定手順 (パラレルゲートキーピングプロシジャ、条件 B)

#raw.p : 未調整 p 値、w : 重み、 f,fn : ファミリー、 B : リサンプリング数

```
#検定手順
```

```
o.p.1 <- order(ifelse(fn==1,raw.p/w,raw.p/0))[1]
```

```
adjp.qp.1[o.p.1] <- sum(ifelse(apply(qperms.w.p[fn==1], 1, min)
```

```
  <= min(w.p[fn==1]), 1, 0))/B;
```

```
adjp.qp.2[o.p.1] <- adjp.qp.1[o.p.1]
```

```
fn <- ifelse(fn==1, 2, fn); fn[o.p.1] <- 1;
```

```
o.p.2 <- order(ifelse(fn==2,raw.p/w,raw.p/0))
```

```
o.p.2.1 <- order(ifelse(f==1&fn==2,raw.p/w,raw.p/0))
```

```

o.p.2.2 <- order(ifelse(f==2,raw.p/w,raw.p/0))

adjp.qp.1[o.p.2.1[1]] <- sum(ifelse(apply(qperms.w.p[,fn==2], 1, min)
  <= min(w.p[f==1&fn==2]), 1, 0))/B;

adjp.qp.1[o.p.2.2[1]] <- sum(ifelse(apply(qperms.w.p[,f==2], 1, min)
  <= min(w.p[f==2]), 1, 0))/B;

adjp.qp.1[o.p.2.2[2]] <- sum(ifelse(qperms.w.p[,o.p.2.2[2]]
  <= min(w.p[o.p.2.2[2]]), 1, 0))/B;

adjp.qp.2[o.p.2[1]] <- sum(ifelse(apply(qperms.w.p[,fn==2], 1, min)
  <= min(w.p[fn==2]), 1, 0))/B;

n2 <- length(fn[fn==2]); n2.1 <- length(fn[fn==2])-1; n2.2 <- length(fn[fn==2])-2;

for(i in 1:n2.2){

  j <- i + 1

  adjp.qp.2[o.p.2[j]] <- sum(ifelse(apply(qperms.w.p[,o.p.2[1:i]][,fn[o.p.2[1:i]]==2],
  1, min) <= min(w.p[o.p.2[1:i]][fn[o.p.2[1:i]]==2]), 1, 0))/B;

}

adjp.qp.2[o.p.2[n2]] <-

sum(ifelse(qperms.w.p[,o.p.2[1:n2.1]][,fn[o.p.2[1:n2.1]]==2]
  <= min(w.p[o.p.2[1:n2.1]][fn[o.p.2[1:n2.1]]==2]), 1, 0))/B;

#調整 p 値

```

```

or1 <- c(o.p.1,o.p.2.1[1],o.p.2.2[1:2])

or2 <- c(o.p.1,o.p.2[1:n2])

for(i in 1:np.1){

  j <- i + 1

  if(adjp.qp.1[or1[i]] > adjp.qp.1[or1[j]]) {

    adjp.qp.1[or1[j]] <- adjp.qp.1[or1[i]]

  }

  if(adjp.qp.2[or2[i]] > adjp.qp.2[or2[j]]) {

    adjp.qp.2[or2[j]] <- adjp.qp.2[or2[i]]

  }

}

if (adjp.qp.1[o.p.2.1[1]] <= 0.025) {adjp.qp <- adjp.qp.1

} else {

  adjp.qp <- adjp.qp.2

  adjp.qp[o.p.2.1[1]] <- adjp.qp.1[o.p.2.1[1]]

}

```

附録2 ブートストラップ法による同時帰無分布の推定

#total.n : 総人数、label : 治療群、np : エンドポイント数、B : サンプル数

#ブートストラップ法による同時分布の推定

```
resamp <- function(total.n,label,np,B){

  label <- as.vector(label)

  uniqlabs <- unique(label)

  num.group <- length(uniqlabs)

  groupIndex <- lapply(1:num.group, function(k) which(label == uniqlabs[k]))

  obs <- sapply(1:num.group, function(x) length(groupIndex[[x]]))

  samp <- lapply(1:num.group, function(k) matrix(NA, nrow = B, ncol = obs[k]))

  for (j in 1:B) {

    for (i in 1:num.group) {

      uniq.obs <- 1

      count <- 0

      while (uniq.obs == 1) {

        count <- count + 1

        samp[[i]][j, ] <- sample(groupIndex[[i]], obs[i], replace = TRUE)

        uniq.obs <- length(unique(samp[[i]][j, ]))

      }

    }

  }

}
```

```

boots.n <<- as.vector(t(matrix(unlist(samp), nrow = B, ncol = sum(obs))))

boots <<-lapply(1:np,function(k) matrix(X[,k][boots.n],nr=total.n,nc=B))

boots.p <<- sapply(1:np,function(k) apply(boots[[k]],2,function(x) t.test(x~label,
  alternative="greater", var.equal=T)$p.value));
}

resamp(total.n,label,np,B)

#周辺分布に一様分布を用いた場合の同時帰無分布の推定

rboots.p <- apply(boots.p,2,rank)

qboots.p <- rboots.p/B

#並べ替え法による周辺分布の推定

perm <- function(total.n,label,np,B){

  per <- matrix(NA, nrow = B, ncol = total.n)

  for (j in 1:B) {

    per[j, ] <- sample(label, replace = FALSE)

  }

perms.p <<- sapply(1:np,function(k) apply(per,1,function(x) t.test(X[,k]~x,
  alternative="greater", var.equal=T)$p.value));
}

perm(total.n,label,np,B)

```

#周辺分布に並べ替え分布を用いた場合の同時帰無分布の推定

```
qperms.p <- NULL
```

```
for(i in 1:np){
```

```
  qperms.p <- cbind(qperms.p, perms.p[,i][order(perms.p[,i])][rboots.p[,i]])
```

```
}
```