

令和 5 年 6 月 19 日現在

機関番号：22701

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K11195

研究課題名（和文）小標本臨床試験におけるカテゴリカル変数に対する正確検定の開発

研究課題名（英文）Development of exact tests where the primary endpoint is a categorical variable

研究代表者

山本 紘司（YAMAMOTO, Kouji）

横浜市立大学・医学部・准教授

研究者番号：10548176

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：臨床試験における主要評価変数として、カテゴリカル変数が選択されることも少なくない。2値変数などのカテゴリカル変数に対する解析手法の多くは大標本近似を用いており、とくにアカデミアなどの医師主導臨床試験ではあまり多くの症例数を集めることができず小標本となることが多いため、大標本近似を用いた解析手法の適用の妥当性に疑問を生じることがある。本研究では2値変数などのカテゴリカル変数を主要評価変数とする臨床研究において1. 2つのスクリーニング検査間における的中率比較に対する正確検定の開発
2. 複数の治療候補から最善の治療候補を選択する臨床試験における有効性と安全性を同時考慮したデザインの開発

研究成果の学術的意義や社会的意義

医学分野では、疾患の確定診断を行う前にスクリーニング検査をすることが多いが、十分な診断精度をもっているとは言えない領域もある。たとえば胃癌の腹膜播種（転移）は術前に診断する必要があるが、その診断性能には改善の余地が残されている。そこで新たな検査法を開発する際には、既存の検査法と診断性能を比べる必要があるが、上述の胃癌の例であれば、的中率による比較で性能を比較したい。これに対する比較手法はいくつか提案されているが、症例数が十分でない場合には、既存の手法では許容できない程度に誤って新規法がよいと結論づける可能性があったが、本研究によって提案された正確な手法であればそのようなことは起きない。

研究成果の概要（英文）：Many analysis methods for categorical variables such as binary variables use large-sample approximation. However, especially in physician-initiated clinical trials in academia, the number of patients cannot be large and the sample size is often small. Since methods using large-sample approximations are often not appropriate in such cases, the present study was undertaken to develop exact methods that do not use approximations. This study focused on the following two points:

1. to develop exact tests for the comparison of predictive values between two screening tests
2. to develop a design that simultaneously considers efficacy and safety in clinical trials to select the best treatment candidate from multiple treatment candidates

研究分野：医学統計

キーワード：正確検定 臨床試験

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

臨床試験の目的を達成する上で、主要評価変数として連続変数もしくはカテゴリカル変数を選択できるとすれば、検出力向上やそれに伴う症例数削減などのメリットから、連続変数を選択することが多い。一方で、スクリーニング検査の結果（陽性か陰性）や、固形がんに対する治療効果判定[RECIST に基づいた CR, PR, SD, PD など]などカテゴリカル変数を主要評価変数とする（またはせざるを得ない）ことも多々ある。独立な 2 群間で 2 値変数に基づく割合を比較するのであれば z 検定もしくはピアソンのカイ二乗検定がよく用いられ、3 値以上のカテゴリカル変数に対する比較であれば多項ロジットモデルや比例オッズモデルなどを用いて検定することができる。しかし、これらの手法では基本的に大標本理論をもとにした検定が用いられるため、小標本の場合には近似精度などの観点からその適用の妥当性が疑われることがある（独立な 2 群の比率の検定であれば、小標本で用いることのできるフィッシャーの正確検定はよく知られている）。とくに臨床試験においては、小標本の場合に大標本理論に基づいた手法を適用することによる第 1 種の過誤確率の増大は、最終的に患者さんへの不利益へと直結する恐れがあるため、小標本の場合にも適用可能な“適切な”解析手法を用いることが望ましい。本研究では、未だ大標本の場合の漸近的性質が明らかになっていない、もしくは小標本の場合に適用可能な適切な解析手法が確立していない（もしくは存在しない）課題に対して、新たな解析手法およびデザインの開発を行い、これらの課題の解決を図る。

2. 研究の目的

上述の課題に対して、当初は次の課題を設定した：

2 つの検査法間における陽性的中率/陰性的中率の比較に対する正確検定

1 被験者から複数の病変が得られる場合の全被験者における奏効率に対する正確検定

複数の治療候補から最善の治療候補を選択する臨床試験において、同一被験者が複数の治療候補群に含まれる場合の適切なデザインとその正確検定

しかし、研究を進めていく上で、課題 および に関して、開発の前提となる部分ですでに課題があり、その課題についても未解決であることがわかったため、まずはそこに取り組む必要があると判断し、研究目的・方針を次のように修正した（課題 は修正なし）：

2 つの検査法間における陽性的中率/陰性的中率の比較に対する正確検定の開発

複数の治療候補から最善の治療候補を選択する臨床試験において、正確な確率評価に基づく適切なデザインの開発

3. 研究の方法

課題 に対して、2 つの検査法のいずれが性能がよいか比較する上で、1 人の被験者にどちらの検査法も試すという方法が多くの場合合理的であり実際よく用いられるため、このような試験デザイン（以下、paired design）のもとで考える。このとき、陽性的中率/陰性的中率の比較に対する検定手法は、Leisenring et al. (2000)、Moskowitz and Pepe (2006)、Kosinski (2013) など数編の学術論文しかなく、解析手法が確立しているとは言い難い。さらにこれらの手法はすべて大標本理論に基づくものである。Kosinski (2013) では、その論文中で提案された新たな手法がそれまでの既存手法と比べても性能がよく、第 1 種の過誤確率を名目有意水準以下に抑えることができると述べられている（シミュレーションに基づいており理論的証明の記載はない）。ところが、小標本の状況下で、実質的な第 1 種の過誤確率が名目有意水準を超える場合もあることが自身のシミュレーション結果からわかった。治験の第 3 相試験のような検証的な臨床試験においては、第 1 種の過誤確率の増大は本来治療効果のないものを患者さんへ還元してしまう危険性が増すことに直結するため、適切な解析手法の選択が重要となる。したがって、課題 に対して Kosinski (2013) を含む既存の手法では、小標本の場合に第 1 種の過誤確率が名目有意水準を上回る可能性があるため、大標本理論を用いない正確な検定の構築が新たに必要となる。そこで、並べ替え検定の考え方を用いることで正確検定を開発していく。シミュレーションによって性能評価を行う。

課題 に関しては、がんの第 2 相試験などで用いられることがある「選択問題」に主眼を置いている。これは複数の治療候補の中から最善と考えられる治療法を選択し、その治療法を第 3 相試験での被験治療とする場合などに用いられる。このとき、最善の治療として選択する際の指標として、奏効率などの 2 値変数や生存時間が用いられることが多く、たとえば Chen and Ng (1998) や Sargent and Goldberg (2001) など古くから議論されている。とくに 2 値変数に対してのデザインおよび解析手法に関する Sargent and Goldberg (2001) では、小標本の場合にも適用可能な正確な方法について述べられている。Sargent らの方法では、1 つの主要評価項目（多くの場合は奏効割合などの有効性指標）についてのみ検討するデザインであり、副次的に安全性指標を用いた意思決定を行っていた。しかし、安全性指標はとくに早期の臨床試験においては重要であり、副次的にはなく、主要評価項目として考慮すべき状況も多く、そのような状況に対応したデザイン開発を行う必要がある。そこで、Sargent らの方法におい

て、有効性のみならず、安全性指標も主要評価項目として設定し、これら 2 つの側面から最適な治療候補を選択できる正確デザインを開発する。2 つの離散確率変数の同時分布を考えることでこの目的は達成できると考えられる。

4. 研究成果

課題 に関しては、並べ替え検定の考え方を採用し、paired design のもとで、陽性的中率/陰性的中率を比較するための正確検定を提案した。通常、並べ替え検定にはコンピューターを用いた膨大な計算時間が必要であり、サンプルサイズが増えれば増えるほどその傾向は顕著になり、たとえばサンプリングに基づいた p 値の計算を行ったりするが、本研究では単純に全パターンへの並べ替えを行うのではなく、得られたデータを集計した分割表と元の素データとの関係に着目し、そこからそれらの間の関係式を見出し、この考え方に基づいた正確 p 値の導出を行った。さらにシミュレーションによりその性能を評価し、正確検定であるがゆえ、既存の方法に比べて保守的ではあったが、名目有意水準を超えることはなかった。検証的な臨床試験などでは有意水準の制御は重要であるため、そのような厳密な評価が求められる際には有用な方法であると考えられる。この成果についてはすでに論文として出版された (Takahashi and Yamamoto, 2020)。

課題 に関しては、まずは 2 つの治療法のうちどちらを選択するかにおける選択問題について、有効性および安全性の 2 つの指標を同時考慮した正確な確率計算を行うための式を導出した。ただし、この式に基づいた確率計算には膨大なメモリの使用と計算時間の問題という 2 つの課題があり、 $n=50$ を超えるサンプルサイズにおいては普通の PC ではほぼ計算不能であることもわかった。そこで、上記で求められた確率計算において、2 つのカテゴリカル変数の同時分布を正規分布に近似することで近似的な確率を求め、その精度について検証した。その結果、 $n=50$ ほどで、正確に計算を行った場合の確率と、近似式を用いた場合の確率にはほぼ差がないことがわかり、実際の臨床研究での実用上は近似式での評価で十分であることもわかった。あわせて、実地場で提案手法を利用してもらうためには、上記の確率計算はもちろんのこと、デザイン段階でのサンプルサイズ計算も必要となるため、これらが容易に行えるように web アプリケーションの開発も行った。現時点では提案手法の性能評価およびアプリケーションの開発は終了し、論文の投稿準備中である。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 1件）

| | |
|--|---------------------|
| 1. 著者名 Takahashi, K. and Yamamoto, K. | 4. 巻 19 |
| 2. 論文標題 An exact test for comparing two predictive values in small size clinical trials | 5. 発行年 2020年 |
| 3. 雑誌名 Pharmaceutical Statistics | 6. 最初と最後の頁 31-43 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/pst.1968 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |

| | |
|---|-----------------------|
| 1. 著者名 Takahashi, K. and Yamamoto, K. | 4. 巻 6 |
| 2. 論文標題 A non-inferiority test and sample size determination for comparing two predictive values of diagnostic tests | 5. 発行年 2020年 |
| 3. 雑誌名 Communications in Statistics: Case Studies, Data Analysis and Applications | 6. 最初と最後の頁 109-122 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/23737484.2020.1719559 | 査読の有無 有 |
| オープンアクセス オープンアクセスではない、又はオープンアクセスが困難 | 国際共著 - |

〔学会発表〕 計4件（うち招待講演 0件/うち国際学会 2件）

| |
|--|
| 1. 発表者名 郷洋文, 山本紘司 |
| 2. 発表標題 2つの評価変数を考慮したflexibilityを有するランダム化第 相選択デザイン |
| 3. 学会等名 日本計算機統計学会第35回大会 |
| 4. 発表年 2021年 |

| |
|-----------------------------------|
| 1. 発表者名 高橋佳苗, 山本紘司 |
| 2. 発表標題 陽性的中率・陰性的中率の比較における正確検定 |
| 3. 学会等名 応用統計学会 |
| 4. 発表年 2019年 |

| |
|--|
| 1. 発表者名 Yamamoto, K. and Takahashi, K. |
| 2. 発表標題 4. Test statistic for comparing single summary measures for multi-class classifiers |
| 3. 学会等名 WNAR2019 (国際学会) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Kanae Takahashi and Kouji Yamamoto |
| 2. 発表標題 A non-inferiority test for comparing two predictive values of diagnostic tests |
| 3. 学会等名 ENAR2019 (国際学会) |
| 4. 発表年 2019年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|--|---------------------------|-----------------------|----|
| | | | |

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
| | |