

令和 6 年 6 月 25 日現在

機関番号：22701

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K11790

研究課題名（和文）カテゴリ変数を伴う臨床研究における新たな解析法およびデザインの開発

研究課題名（英文）Developing new statistical methods and designs for clinical research involving categorical variables

研究代表者

山本 紘司（YAMAMOTO, Kouji）

横浜市立大学・医学研究科・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）：臨床研究において、カテゴリカル変数が主要評価変数として選択されることはままある。本研究では実課題から得られた次の2つの問題に対して研究を行った：(1) 3カテゴリ以上の結果をもつ検査法比較に対するF1スコアによる解析法の提案，(2) 複数の2値変数を主要評価変数としたフレキシブルな治療候補選択デザインの開発。

(1)については機械学習分野でよく用いられる性能評価指標であるF1スコアに対して、3カテゴリ以上の場合に用いられるF1スコア指標を整理し、その統計的性質を明らかにした。また、(2)については有効性と安全性という2つの2値変数を同時に評価し、最善の治療法を選択するデザインを新たに提案した。

研究成果の学術的意義や社会的意義

本研究は実際の医学研究における課題から着想を得たものであり、これらに対して新たな解析手法等を提案した。これは新たな統計的手法開発にとどまらない。課題(1)に対してはより科学的に新たに開発された検査法や診断法の有効性を述べることができ、課題(2)に対してはより多面的な角度から最善の治療法を選択できる可能性が高まる。本研究手法を今後の医学分野へ応用することにより、より効率的な研究遂行が可能となり、最終的には疾患で苦しむ患者さんへのよりよい医療の提供につながるものと期待される。

研究成果の概要（英文）：In clinical research, categorical variables are often selected as the primary endpoint. In this study, we studied two problems that were derived from real-world problems: (1) the proposal of an F1 score-based analysis method for comparisons of methods with three or more categories of results, and (2) the development of a flexible treatment candidate selection design with multiple binary variables as the co-primary endpoints. For (1), we proposed the F1 score, which is a commonly used performance measure in the machine learning field, and clarified the statistical properties of F1 score measures used in the case of three or more categories. For the second, we proposed a new design that simultaneously evaluates two binary variables, efficacy and safety, and selects the best treatment.

研究分野：医学統計

キーワード：臨床試験 F1スコア 選択デザイン

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

臨床研究においては、主要評価項目としてカテゴリカル変数が選択されることもしばしばある。たとえば、皮膚がんの種類(悪性2種類, 良性4種類)を診断する場合や、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)などは典型的なカテゴリカル変数である。上述の皮膚がんの種類診断においては、機械学習を用いた場合と熟練医の判断とで診断性能に差があるかが研究されたが、その際には上述の6種の結果を2値化した上で、感度・特異度等を用いて評価がなされた。よく知られた診断性能評価のための指標である感度・特異度などは当然2値アウトカムに対して定義されているものである。しかしこの例のように6カテゴリのアウトカムに対して感度・特異度、さらにはデータがインバランスなときに有用とされているF1スコアを用いることができればより情報を有用に活用できると期待される。

がんに対する第2相臨床試験には、いくつかの治療法の中から次の第3相試験に進めるべき治療法を選択する目的で行われるものがある。このとき、有効性指標として奏効割合を用いることが多い。もしこの奏効割合をもとに最善の治療法を選択できないようであれば、さらに他の有効性もしくは安全性指標に基づき1つの治療法を選択できるフレキシブルデザインも提案されている。しかし、この他の指標に基づき選択するというのはとくに統計的に妥当な評価がなされるわけではなく、最初からその第2の指標まで加味してデザインを組み、最善の治療法を選択する方がより科学的に妥当な方法であると考えられる。

2. 研究の目的

本研究では、上述の課題に対して、次の2つの研究テーマを設定した：

(1). 3カテゴリ以上の結果をもつ検査法比較に対するF1スコアによる解析法の提案

(2). 複数の2値変数を主要評価変数としたフレキシブルな治療候補選択デザインの開発

(1)については、多値カテゴリに対するF1スコア自体は3種類提案がなされているが、応用上の使用法はその点推定値のみを用いて性能比較がなされているため、当然ばらつきが考慮されていない。したがって、多値カテゴリ版のF1スコアの推定値に対する漸近分布を導出することで統計的推測が可能となると考えた。また、(2)については、第2相試験開始時点ではまだ十分な安全性情報が得られていないことが多いことから、安全性についても主要評価変数として取り入れたデザインを用いて最善の治療法を選択することが妥当と考え、この考えに基づくデザイン・解析法・症例数設計法を提案する。

3. 研究の方法

課題(1)に対して(令和3~4年度に取り組む)

1つの判別法に対しては、いわゆる confusion matrix (真の結果と判別法による判別結果をまとめた分割表)の形でデータが得られるため、データに多項分布を想定することで1標本におけるF1スコアの推定値に対する漸近分布の導出が可能と考えられる。2つの判別法比較に対しても同様の考えが適用可能である。ただし対応のあるデザインの場合には2標本が独立ではないため、相関を考慮することになる。ある程度サンプルサイズが大きい場合には、多変量中心極限定理による漸近分布や信頼区間・検定の構築が可能だが、どの程度のサンプルサイズがあれば近似が十分かについてもシミュレーションを通して明らかにする。

課題(2)に対して(令和4~5年度に取り組む)

デザインとしてまずは主要評価変数が2つの場合におけるデザイン開発を行う。具体的には主要評価変数とともに2値変数の場合のフレキシブル選択デザインを開発する。これは、相関をもつ多項分布の積からなる同時確率分布を求めることで達成可能と考えられる。さらに余裕があれば次年度に、主要評価変数が2値変数と連続変数の場合を想定した場合における同デザインを考える。現実的な状況設定において、このデザインによる必要サンプルサイズの計算を行い、実用的なデザインかをシミュレーションを通して評価する。また、提案デザインを使用してもらうためにもR等によるアプリケーションの開発も同時に行う。

4. 研究成果

課題(1)：3カテゴリ以上の結果をもつ検査法比較に対するF1スコアによる解析法の提案について

多値カテゴリ版のF1スコアを3種類に整理し、各versionのF1スコアの推定値に対する漸近分布を導出し、シミュレーションによりその性能評価を行った。多くのシナリオ下でサンプルサイズが200~300程度あれば信頼区間の信頼度はほぼ名目水準に達することが示された。さらに、対応のある2標本問題における多値カテゴリに対するF1スコアを用いた実データ解析を行い、既存の2値化した場合の解析法との比較も行った。また、本研究での成果を実際の研究で広く使用してもらうため、Rによるプログラムコードも論文内に掲載する形で学術誌への投稿を行い採択された。

課題 (2) : 複数の 2 値変数を主要評価変数としたフレキシブルな治療候補選択デザインの開発について

主要評価変数として、有効性および安全性のどちらも 2 値変数である状況を想定し、どのような状況であればどんな意思決定を下すかの decision rule を定義し、それに従った定式化を行った。このとき正しく最善に治療法を選択できる確率に基づき判断を下すことになるが、この確率の計算については、まずは多項分布の積からの正確確率に基づく計算を行った。ただし、プログラムにて当該確率を求める際、この計算は非常に時間がかかり、さらにサンプルサイズ 50 程度で一般的な PC でメモリオーバーフローを起こすことがわかった。そこで、上記で計算した正確な確率計算について、近似を用いた計算を行うよう工夫し、そのもとでの近似精度の評価を行ったところ、サンプルサイズ 20~30 程度で実用上十分な近似精度が得られた。これらのことから、サンプルサイズが 20 未満の場合は正確な確率計算を行い、それ以上のサンプルサイズのもとでは近似計算により確率計算を行うようにし、これらをアプリケーションとして開発した。現在この研究成果については論文投稿中である。

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 2件/うちオープンアクセス 1件）

1. 著者名 Takahashi Kanae, Yamamoto Kouji, Kuchiba Aya, Koyama Tatsuki	4. 巻 52
2. 論文標題 Confidence interval for micro-averaged F1 and macro-averaged F1 scores	5. 発行年 2021年
3. 雑誌名 Applied Intelligence	6. 最初と最後の頁 4961 ~ 4972
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10489-021-02635-5	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 該当する

1. 著者名 Takahashi Kanae, Yamamoto Kouji, Kuchiba Aya, Shintani Ayumi, Koyama Tatsuki	4. 巻 42
2. 論文標題 Hypothesis testing procedure for binary and multi class F_1 scores in the paired design	5. 発行年 2023年
3. 雑誌名 Statistics in Medicine	6. 最初と最後の頁 4177 ~ 4192
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/sim.9853	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 該当する

〔学会発表〕 計2件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 Kanae Takahashi and Kouji Yamamoto
2. 発表標題 A superiority test for comparing sensitivity, specificity, and predictive values of two diagnostic tests
3. 学会等名 The 37th International Workshop on Statistical Modelling（国際学会）
4. 発表年 2023年

1. 発表者名 高橋 佳苗, 山本 紘司
2. 発表標題 多値分類臨床検査データにおける感度, 陽性的中率, F1 スコアの仮説検定
3. 学会等名 2024年度応用統計学会年会
4. 発表年 2024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			
米国	Vanderbilt University			