

令和元年5月21日現在

機関番号：14101

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K19249

研究課題名(和文) ベイズ法を用いたマクネマー検定の改善

研究課題名(英文) Improvement of McNemar test using Bayesian method

研究代表者

小椋 透 (OGURA, Toru)

三重大学・医学部附属病院・講師

研究者番号：00580060

交付決定額(研究期間全体)：(直接経費) 1,200,000円

研究成果の概要(和文)：マクネマー検定は実質水準が極端に低い場合があることが知られており、ベイズ型の検定として表せることを用いてマクネマー検定の改善を行った。次に、マクネマー検定が行われている複数試験を併合して検定する方法として、各試験の事後確率の積を検定統計量に用いることを提案した。さらに、マクネマー検定の改善の研究を生かして、二項割合の信用区間についてベイズ法を用いて改善した。二項確率のロジット変換値による最高事後密度区間を算出し、その区間を逆ロジット変換した区間を二項確率の信用区間とする方法を提案した。

研究成果の学術的意義や社会的意義

マクネマー検定は有意水準5%で検定しても実質水準が5%よりはるかに小さいことがあった。本研究の提案法を用いると有意水準を厳守しながら実質水準の高い検定を行うことができる。すなわち、検定が有意になりにくいことが改善された。また、複数試験の併合では、従来の各試験の分割表における各セルを足し合わせて併合の分割表によるマクネマー検定では、2試験の結果が極端であっても、併合後は良い結果の試験に引っ張られる可能性があったが、提案法では2試験の結果が安定していないと併合後の結果は良ならず、偶然を排除することができる検定であった。

研究成果の概要(英文)：It was known that the McNemar test was limited by its unexpectedly small power, and the McNemar test was improved using what can be expressed as a Bayesian test. As the determination of the critical region within the framework of "T" ( ), ' was set so that the actual level becomes as high as possible while strictly observing the significance level. Next, we extended the proposed test to combinations of binary matched-pairs data from multiple strata. The test statistic was the product of the posterior probabilities of the multiple strata. In numerical experiments, the proposed test for multiple strata was superior when the test statistics of the individual strata were close. In addition, a novel credible interval of the binomial proportion was proposed by improving the Highest Posterior Density interval using the logit transformation. By the inverse logit transformation of the interval, the novel credible interval for the binomial proportion was obtained.

研究分野：生物統計学

キーワード：分割表 マクネマー検定 メタアナリシス 信用区間

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

(1) 対応のある  $2 \times 2$  の分割表における検定には、マクネマー検定がよく用いられていた。しかしながら、マクネマー検定は、実質水準が有意水準より低くて、検定が有意になりにくいことが知られていた。

(2) 対応のある  $2 \times 2$  の分割表の複数試験の併合 (メタアナリシス) の方法として、各試験の対応のある  $2 \times 2$  の分割表における各セルを足し合わせて併合の  $2 \times 2$  の分割表を作成し、その分割表に対してマクネマー検定が行われていた。この方法は、併合前の各試験の検定結果は、併合後の検定に与える影響が少ないという欠点があった。

### 2. 研究の目的

(1) マクネマー検定は実質水準が極端に低い場合があることが知られている。マクネマー検定がベイズ法で表せることを用いて、有意水準を厳守しながら実質水準の高い検定となるようにマクネマー検定を改善する。

(2) 対応のある  $2 \times 2$  の分割表のメタアナリシスに、ベイズ法を用いて新たな検定法を確立させる。本研究では各試験の検定統計量の積をメタアナリシスの検定統計量として用いる。(1) の研究成果であるマクネマー検定の改善を用いて、メタアナリシスにおける検定法へ拡張させる。

### 3. 研究の方法

(1) 有意水準を厳守しながら実質水準の高い検定法にマクネマー検定を改善する研究の方法は、Altham (1971) がマクネマー検定をベイズ型の検定として表していることを用いて、マクネマー検定における検定統計量  $T$  は事後確率を用いては次のように表せる。

$$T = pp(x_{12}, x_{21}) = 1 - F(1/2; x_{12} + a, x_{21} + b)$$

上式において、最初に  $a=b=0.5, 1$  と設定して研究を行い、有効性が検証された後に最適な  $a, b$  について研究を行う。また、通常は「 $T > 1 - \alpha$ 」( $\alpha$ : 有意水準) の枠組みで棄却域を決めることが多いが、本研究は「 $T > 1 - \alpha'$ 」( $\alpha'$ : 有意水準) の枠組みで棄却域を決める。

(2) 対応のある  $2 \times 2$  の分割表のメタアナリシスに、ベイズ法を用いた新たな検定法に関する研究の方法は、Fisher (1973) のアイデアである各試験の  $P$  値の積を用いた方法を参考に、各試験の検定統計量  $T$  の積をメタアナリシスの検定統計量として用いる。

### 4. 研究成果

マクネマー検定は実質水準が極端に低い場合があることが知られていた。本研究は、マクネマー検定がベイズ型の検定として表せることを用いて改善を行った。その方法は、一般に「 $T > 1 - \alpha$ 」( $\alpha$ : 有意水準) の枠組みで棄却域を決められることが多いが、本研究では「 $T > 1 - \alpha'$ 」( $\alpha'$ : 有意水準) の枠組みで棄却域を決めることとした。このとき、 $\alpha'$  は有意水準を厳守しながら実質水準がなるべく高くなるように定めた。さらに、ベイズ型の検定であることを生かして、現実的な事前密度を用いることとした。数値計算と実例を用いた有効性の検証より、有意水準を厳守しながら実質水準の高い検定法であることが示された。その結果の一つとして、図 1 は  $n=30, 50, 100$  における  $\alpha=0.05$  とした場合の実質水準のグラフであり、濃点線が  $a=b=0.25$ 、濃破線が  $a=b=0.5$ 、濃実線が  $a=b=1$ 、薄点線が  $a=b=1.5$ 、薄破線が  $a=b=2$ 、薄実線がマクネマー検定を示す。提案法の実質水準は  $\alpha=0.05$  以下と有意水準を厳守しながら、 $0.05$  近くの高い値であった。一方、マクネマー検定は常に実質水準が低いことが示された。他の数値計算や実例の結果を併せて検討すると、最適な  $a, b$  は  $n$  により変わることが分かった。目標症例数  $n$  は臨床試験の開始前に決定されることから、どの  $a, b$  を用いるかを試験開始前に作成されるプロトコールに事前明記が可能であり、後付け解析と疑われることを防ぐことができる。

先の研究結果を生かして、マクネマー検定が行われているメタアナリシスの検定に、各試験の事後確率の積をメタアナリシスの検定統計量として用いた。実例として、2 試験のメタアナリシスの結果を図 2 に示す。灰色の領域は棄却域、黒丸が検定統計量を表す。左の (i) は、各試験で有意差が認められないが似た結果である場合、右の (ii) は、各試験で有意差が認められなくて、2 試験の結果が大きく異なる場合である。(i) は併合すると有意差が認められるようになり、(ii) は併合しても有意差が認められなかった。臨床試験を併合するための条件は同一方法の複数試験であり、その時には似た結果が得られることが妥当であることから、似た結果の臨床試験の併合に有利となる提案法の検定は実用的であった。一方、同一方法の試験で結果が大きく異なる場合とは、試験実施上の問題がある可能性が考えられて、従来の各試験の対応のある  $2 \times 2$  の分割表における各セルを足し合わせて併合の  $2 \times 2$  の分割表によるマクネマー検定の場合は、試験間にあった結果の差は併合後には影響されにくい、提案法の検定では試験間にあった結果の差は併合後にも考慮されて有意差が認められにくいという特徴があった。以上より、提案法の検定は実用的であると考えられた。

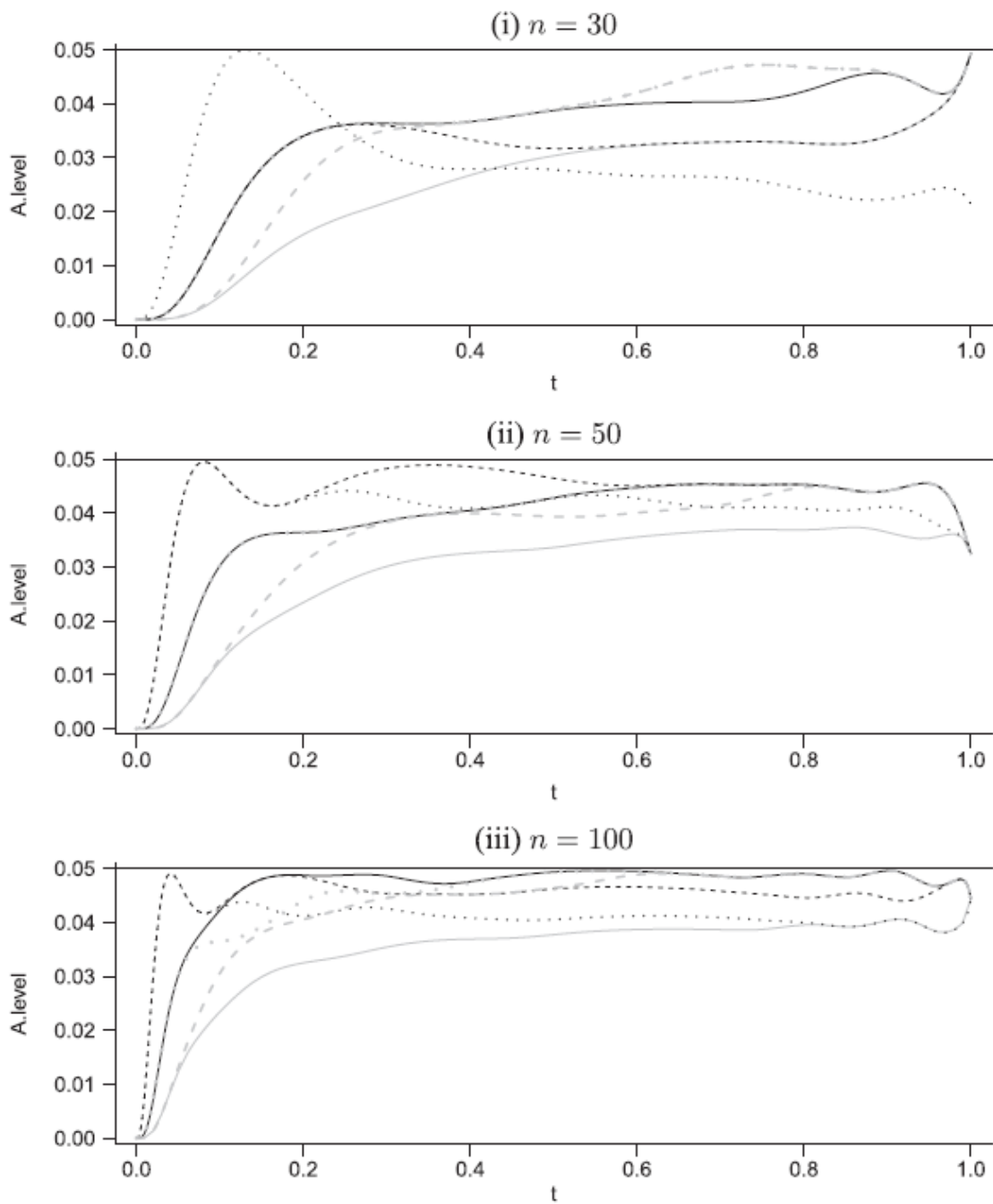


図1 実質水準比較 ( $n=30, 50, 100$ )

当初の計画にはなかったが、マクネマー検定は群間比較の相対評価であるため、その検定のみでは各群の絶対評価は分からないことから、各群の区間推定について研究を行った。従来のいくつかの区間推定方法では、二項確率  $p > 0$  が小さく、試験結果  $x=0$  が観測された場合に区間推定の下限值は 0 であった。しかしながら、二項確率が小さくても  $p=0$  と考えられない場合に、区間推定の下限值が 0 となることは適切でないと考えられた。例えば、A 地域で報告されている疾病が、B 地域で 1 例も報告されていない場合、B 地域の住民における疾病発生確率は小さいけれど 0% とは考えられないことである。これまでのマクネマー検定の改善の研究を生かして、二項割合の信用区間についてベイズ法を用いて改善した。その方法は、二項確率  $p$  のロジット変換値による最高事後密度で区間を算出して、その区間を逆変換することにより、二項確率  $p$  の信用区間として提案した。数値例の一部として、 $n=8$ 、 $x=0$  における 95% 信用区間を表 1 に示す。提案信用区間、ロジット変換しない既存の最高事後密度による信用区間、両裾を  $1/2$  とした信用区間、Clopper-Pearson 法による信頼区間を比較した。 $x=0$  の場合に、提案信用区間と両裾を  $1/2$  とした信用区間の下限値は 0 を超えているのに対して、既存の最高事後密度による信用区間と Clopper-Pearson 法による信頼区間の下限値は 0 であった。また、 $n=20$ 、 $a=b=0.5$  における 95% 信用区間のエラー確率を図 3 に示す。提案信用区間のエラー確率は、 $p=0$  付近で高いが、信用区間の下限値は 0 を超えると考えられるため、 $p=0$  付近で高いことの影響は少なく、 $p$

が 0 付近以外で  $\approx 0.05$  付近で安定していることの利点大きい。一方、既存の最高事後密度による信用区間と両裾を  $\sqrt{2}$  とした信用区間は、 $p=0.1$  付近でエラー確率が 0.1 を超えるなど、エラー確率が大きいときがあった。すなわち、提案信用区間は、群ごとに信用区間を算出しても、どの群のエラー確率も大きくなりすぎることはなく、既存の最高事後密度による信用区間と両裾を  $\sqrt{2}$  とした信用区間は、群ごとに信用区間を算出すると、その群の中に  $p=0.1$  が含まれているとエラー確率が大きい可能性があることを意味していた。以上より、提案信用区間は実用的であった。

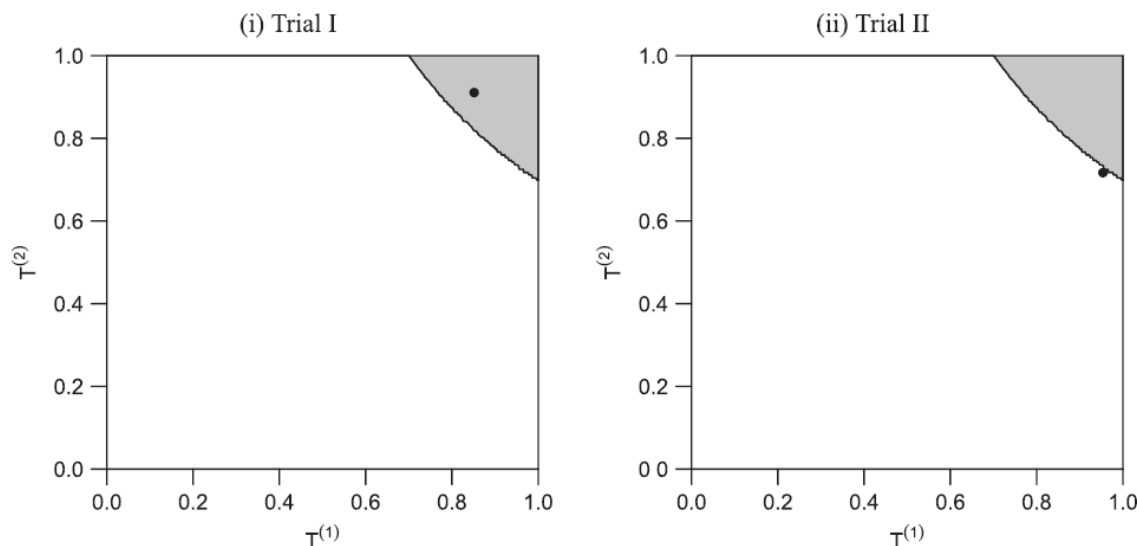


図 2 2 試験の併合

表 1 95%信用区間の比較

n	x	Clopper-Pearson	(a,b)	提案信用区間	既存最高事後密度	両裾 $\sqrt{2}$
8	0	0.0	36.9417 (0.5,0.5)	0.0180 36.0747	0.0 20.7508	0.0059 26.2217
			(1,1)	0.4436 39.5517	0.0 28.3129	0.2809 33.6267
			(0.5,1.5)	0.0162 33.0665	0.0 18.7366	0.0053 23.7610

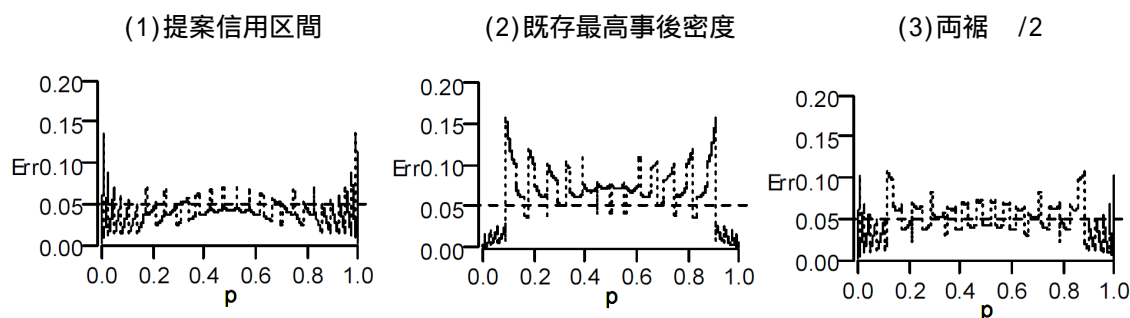


図 3 95%信用区間におけるエラー確率(n=20, a=b=0.5)

## 5. 主な発表論文等

### 〔雑誌論文〕(計 1 件)

Toru Ogura, Takemi Yanagimoto. Improvement of Bayesian Credible Interval for a Small Binomial Proportion Using Logit Transformation, Current Research in Biostatistics, 査読有, Vol.8, 2018, 1-8.  
DOI:10.3844/amjbsp.2018.1.8

### 〔学会発表〕(計 5 件)

Toru Ogura, Takemi Yanagimoto, Performance of Bayesian Credible Interval for Binomial Proportion using Logit Transformation, IASC-ARS/NZSA, 2017.  
小椋透, 柳本武美, 二項確率のロジット変換を用いた最高事後密度による信用区間, 日本

計量生物学会年会, 2017.

小椋透, 柳本武美, 対応のある  $3 \times 3$  分割表の検定方法に対するベイズ法を用いた改善, 統計関連学会連合大会, 2017.

Toru Ogura, Takemi Yanagimoto, Combining Evidences from Multiple Binary matched-Pairs Data using a Bayesian Approach, Annual Conference of the international society of clinical biostatistics, 2016.

小椋透, 柳本武美, 二項確率のロジット変換値における信用区間の構成とその適用例, 統計関連学会連合大会, 2016.

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。