

〈教育報告〉

平成18年度専門課程Ⅱ
生物統計分野

シグナル検出における統計的方法

土居正明

A scan statistic for signal detection

Masaaki DOI

Abstract

In post marketing surveillance of drugs, it is important to find out their unknown undesirable effects. These are called signals. Some different methods have been used in Europe and U.S. for detecting signals. However these methods tend to detect many false positive signals because of multiple testing without multiplicity adjustments. In this paper, a scan statistic is proposed as a method for signal detection in pharmacovigilance, applying the Kulldorff's spatial scan statistic (1997). The proposed scan statistic is designed for detecting only the highly significant signals. The property of the proposed scan statistic is examined in comparison with those of PRR (Proportional Reporting Ratios), BCPNN (Bayesian Confidence Propagation Neural Network), and GPS (Gamma-Poisson Shrinker) using the data from the FDA. It is shown by simulations that the proposed method has low sensitivity but high predictive value of true positive.

Keywords: signal detection, scan statistics, multiple testing

Thesis Advisors: Kunihiko TAKAHASHI, Masako NISHIKAWA, Toshiro TANGO

I. 目的

現在市販されている薬剤の未知の副作用を早期発見することは、製薬企業においても行政においても非常に重要な課題である。一般に、それまで知られていなかったか不完全にしか証拠付けられていなかった薬剤と有害事象との因果関係の可能性に関する情報をシグナルと呼び、このとき「詳細な調査が必要な自発報告の発見と調査の必要性の優先順位付け」を行うことをシグナル検出という¹⁾。

欧米諸国においては、いくつかのシグナル検出のシステムが運用されているが、その検出にはそれぞれ異なる統計手法が用いられており、Gold Standardな手法は存在していない。また、日本においては、自発報告のデータベースは存在するが、システムティックな解析は現時点ではまだ行われていない状況である。

欧米で用いられている手法では、シグナルを多く検出しすぎる傾向にあるが、これは検定の多重性を考慮していないことが原因のひとつと考えられる。そこで本研究では、疾病の集積性の検出のために Kulldorff (1977) によって

開発された spatial scan statistic を本問題に応用し検定の多重性を調整した scan statistic によるシグナル検出法を提案し、欧米の手法との比較を行う。Kulldorff ら²⁾は、spatial scan statistic の考え方を疾病サーベイランスのデータに適用して、これまで知られていなかった疾病リスクファクターの同定方法を提案しているが、本研究は Kulldorff ら²⁾の研究にヒントを得たものである。

II. 手法

欧米の手法、PRR 法、BCPNN 法、GPS 法と、以下の提案手法、scan 法の比較を行う。

薬剤 i ($i = 1, 2, \dots, I$) における有害事象 j ($j = 1, 2, \dots, J$) の報告数を c_{ij} とおき、薬剤 i の有害事象 j における期待報告数を

$$E_{ij} = \frac{c_{i+}c_{+j}}{C} \quad \dots (1)$$

とおく。このとき、 c_{ij} は、それぞれ独立にポアソン分布 $Po(\lambda_{ij}E_{ij})$ に従うものとする。ここで、薬剤 G_1 と有害

指導教官：高橋邦彦，西川正子，丹後俊郎（技術評価部）

事象 G_2 を考え、直積集合 $G_1 \times G_2$ を G とし、
 $c_G = c_{ij}, E_G = E_{ij} ((i, j) = G)$ とおく。ここで尤度関数は

$$L(\lambda_{ij}, c) = \frac{C!}{\prod_{i,j} (c_{ij})!} \prod_{i,j} \left(\frac{\lambda_{ij} E_{ij}}{\sum_{i,j} \lambda_{ij} E_{ij}} \right)^{c_{ij}}$$

である。

ここで

帰無仮説: $\lambda_{ij} = \lambda_0 \quad (\forall i, j)$

対立仮説: $\lambda_{ij} = \begin{cases} \lambda_G & ((i, j) = G) \\ \lambda_R (< \lambda_G) & ((i, j) \neq G) \end{cases}$
 $(\exists G = (i, j))$

の検定問題を考える。

尤度比統計量を考え、対数をとると

$$T = \max_G \left(c_G \log \frac{c_G}{E_G} + (C - c_G) \log \frac{C - c_G}{C - E_G} \right) \cdot I(c_G > E_G)$$

となる。これを検定統計量とする。帰無仮説の検定には、モンテカルロ検定を利用する。

III. FDA の AERS (Adverse Event Reporting System) データの解析

2006年4月～6月の間にFDAに報告されたデータのうち、薬剤名・有害事象名・性別がすべて記入されているものの中から、薬剤は第一被疑薬のみ用いた。このデータ(以下AERSデータと呼ぶ)に対して、scan法、PRR法、BCPNN法、GPS法の四手法でそれぞれ解析を行った。

本データに含まれる自発報告は、薬剤4066種類、有害事象7730種類であり、報告された組み合わせは133,798組、総報告数は380,915件であった。この4066×7730=31,430,180個のセルのうち、報告数が0のセルが31,296,382 (99.6%)であった。

表1より、結果としてscan法で検出されるシグナルは他の手法でもほぼ検出されることが分かった。

表1. AERSデータの各手法での検出数

	検出シグナル数	scan法と共通の検出シグナル数
scan法	2,500	—
PRR法	13,781	2,410
BCPNN法	12,261	2,379
GPS法	9,750	2,443

IV. シミュレーションによる手法の比較

AERSデータをもとにして、真のシグナルをいくつか設定し、 $\lambda > 1$ とした。それ以外においては、 $\lambda = 1$ として、薬剤*i*における有害事象*j*の報告数 c_{ij} が $Po(\lambda E_{ij}) (E_{ij}$ はAERSデータから(1)式で求めた値を用いた)に従うものとして、モンテカルロシミュレーションによりscan法とPRR法、BCPNN法、GPS法の性能を比較した。

V. 結果と考察

シミュレーションの結果、scan法は偽陽性の予測値がほぼ0であることが分かった。これより、scan法で検出されたシグナルは真のシグナルである可能性が非常に高いと思われる。欧米の手法は真のシグナルでないものも大量に検出してしまふという問題が指摘されているが、本手法はその欠点を補っていると思われる。一方、感度が低いことを考えると、この手法だけでは十分に真のシグナルが検出できないものと思われる。そこで、他の手法(たとえば感度の高いPRR法)と組み合わせて利用する、つまり、まずscan法で検出を行い、検出シグナルを優先的に調査することで、効率よく因果関係のある薬剤と有害事象の組み合わせが発見されると考えられる。

参考文献

- 1) 久保田潔. 自発報告からのシグナル検出—英国MCA, 米国FDA, WHOの新しい方法—. 薬剤疫学 2001 Dec;6(2):101-8.
- 2) Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods 1997;26(6):1481-96.

〈教育報告〉

平成18年度専門課程 II
生物統計分野

ケース・コントロール研究における頻度マッチングの効率の定量的評価

都田桂子

Quantitative Assessment of Efficiency of Frequency Matching in Case-Control Studies

Keiko MIYAKODA

Abstract

Objective: I propose a method to quantitatively assess the efficiency of frequency matching for estimating an odds ratio when planning a case-control study.

Methods: In a previously reported simulation study, the efficiency of frequency matching in a case-control study was assessed by continuously changing the degree of matching and calculating the statistical power and variance of estimated odds ratio using a logistic regression model, where both the exposure variable of interest and the matching variable were binary. However, the settings of the simulation were somewhat different from those in an actual case-control study, especially in the procedure of sampling matched controls. I therefore improved the settings of previous study by considering the concept of actual sampling procedure, performed a simulation study using both methods, and compared the results between the two methods.

Results: The results obtained by the methods of present and previous studies were essentially comparable. My study showed that the statistical power could be improved by frequency matching even when the odds ratio of a sample was accidentally smaller than the true odds ratio.

Conclusions: The methods would be useful in planning a case-control study.

Keywords: case-control study, frequency matching, odds ratio, logistic regression

Thesis Advisors: Tetsuji YOKOYAMA, Kazue YAMAOKA, Toshiro TANGO

キーワード： ケース・コントロール研究, 頻度マッチング, オッズ比, ロジスティック回帰

I. 目的

関心のある曝露の有無と疾病の有無の関連をロジスティック回帰によるオッズ比の推定により検討するケース・コントロール研究という状況設定のもとで、頻度マッチングの効率をオッズ比の推定精度および検出力に基づき定量的に評価することを考える。先行研究(1)では、母集団の部分集合である母集団「症例」と母集団「マッチド対照」においてマッチング因子の比率を一致させてシミュレーションを行っていたが、実際の調査では標本で一致させるという点で異なっているという問題点がある。本研究

では、実際の手順を反映させたシミュレーションによって頻度マッチングの効率の定量的評価を行い、また先行研究のシミュレーション結果との比較を行うことを目的とする。

II. 先行研究のシミュレーション方法

1. ケース・コントロール研究の設定

ケース・コントロール研究にて、強い2値リスク因子Mの存在が既知である罹患率の低い疾病Dと、関心のある2値曝露因子Eとの関連を、ケース・コントロール研究にて検討することを想定する。リスク因子M（以降、マッチング因子と呼ぶ）に関する頻度マッチングを行った上で、ロジット関数を、

指導教官： 横山徹爾, 山岡和枝, 丹後俊郎 (技術評価部)

$$\text{logit } p(x) = \alpha + \beta_M M + \beta_E E \quad (1)$$

とする条件なしロジスティック回帰に当てはめ、これにより推定される曝露因子 E と疾病 D の関連のオッズ比 e^{β_E} を指標として推定効率を検討する。

2. シミュレーションの手順

2.1 母集団「症例」、母集団「マッチド対照」

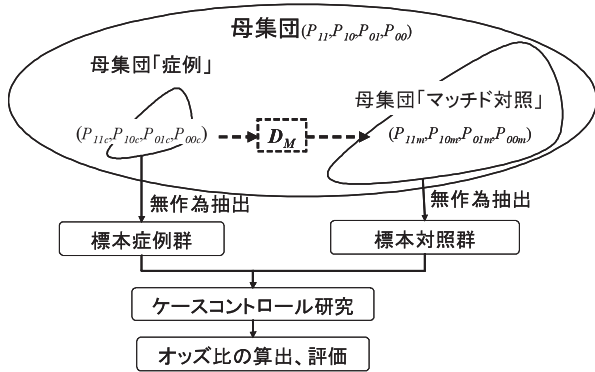


図1 D_M によるシミュレーションの手順

2.2 母集団「症例」と母集団「マッチド対照」との間の一致の程度の指標 D_M

母集団、母集団「症例」、母集団「マッチド対照」におけるマッチング因子 M の比率 (M+ の割合) をそれぞれ P_M, P_{Mc}, P_{Mm} とする。母集団「症例」、母集団「マッチド対照」のマッチング因子 M に関する分布の一致の程度の指標として degree of matching (D_M) を次のように定める。

$$D_M = \frac{P_{Mm} - P_M}{P_{Mc} - P_M} \quad (2)$$

III. 方法

1. ケース・コントロール研究の設定

先行研究との比較のため、同じ設定とする。

2. 標本比率を一致させるシミュレーションの手順

2.1 母集団「症例」と母集団「マッチド対照」

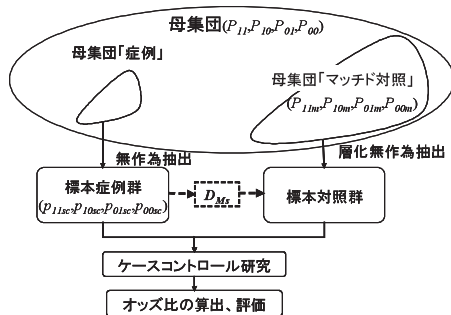


図2 D_{Ms} によるシミュレーションの手順

2.2 両標本群間の一致の程度の指標の D_{Ms}

標本症例群とマッチング因子 M の比率が一致する標本対照群を集めたい。その一致の程度の指標として Sampled

cases' degree of matching (D_{Ms}) を次のように定める。

$$D_{Ms} = \frac{P_{Msm} - P_M}{P_{Msc} - P_M} \quad (3)$$

ここで、 P_M, P_{Msc}, P_{Msm} は、母集団、標本症例群、および標本対照群におけるマッチング因子 M の比率 (M+ の割合) である。

3. D_M と D_{Ms} の比較

D_M による方法では、分布に関するパラメータ、例数に関するパラメータ、および D_M ごとに10,000回のケース・コントロール研究を発生させ、 D_{Ms} による方法では、同様に D_{Ms} ごとに合計100,000回のケース・コントロール研究を発生させる。

それぞれのシミュレーション方法、分布に関するパラメータ、例数に関するパラメータおよび D_M もしくは D_{Ms} ごとに、得られた β_E の推定値の分散、および条件なしロジスティック回帰モデルにより β_E の推定値が Wald 検定で 5% 有意となった割合を検出力として算出し、頻度マッチングの効率の評価指標とする。この評価指標を D_M を用いた場合と D_{Ms} を用いた場合とで比較する。

IV. 結果

いずれのパラメータを変化させた場合も、 D_M もしくは D_{Ms} によるオッズ比の推定効率の推移は似たパターンを示した。曝露 E とマッチング因子 M の関連のオッズ比 OR_{EM} を変化させた場合を、図3に示す。 OR_{EM} が0.25から10へ増加するに伴い、検出力がプラトーもしくは最大になる D_M もしくは D_{Ms} が 0 から1.5へと単調に増加している。 $OR_{EM} = 1$ のとき、すなわち曝露因子 E とマッチング因子 M が独立である場合には、 $D_M = 1$ のとき検出力最大になる。

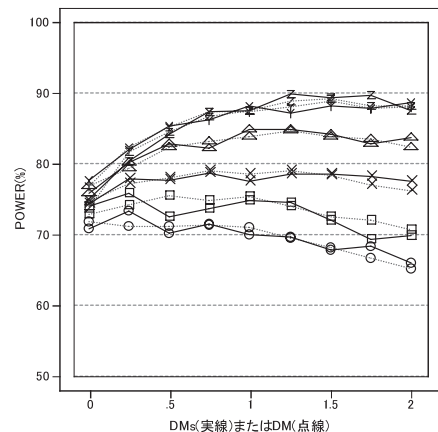


図3 曝露因子 E とマッチング因子 M の関連のオッズ比 OR_{EM} (\circ :0.25, \square :0.5, \times :1, \triangle :2, \diamond :5, \ast :10) と、 D_M または D_{Ms} と検出力の関係

V. 考察

両法の結果は推移としては似たパターンであり、強い 2 値リスク因子の存在が既知である罹患率の低い疾病と関心

のある2値曝露との関連をケース・コントロール研究にて検討する際、頻度マッチングは推定効率をよくすることが、実際的な手順を考慮した D_{Ms} 用いた場合でも確認された。ただし、その程度はパラメータの値により様々に変化する。

VI. まとめ

ケース・コントロール研究における頻度マッチングの効

率の定量的評価を、先行研究による方法と実際の手順を考慮した方法で行い比較したが、両法による結果はほとんどよく一致した。

参考文献

- 1) Stümer T, Brenner H. Degree of matching and gain in power and efficiency in case control studies. *Epidemiology* 2001; 12: 101-8.

〈教育報告〉

平成18年度専門課程Ⅱ
生物統計分野

連続変量に適用した折れ線 Cox モデルと従来の比例ハザードモデルの比較

井上永介

The comparison of the piecewise linear Cox model and the conventional proportional hazards model applied to the continuous covariates

Eisuke INOUE

Abstract

Objective: To examine the piecewise-linear Cox model (PL-Cox) is more applicable model than the conventional proportional hazards model conducting categorization when the assumption of log-linearity would not be accepted for analyzing continuous variables. We proposed the objective method to set the knots for PL-Cox analyzing continuous variables.

Methods: We performed the simulation study. In this simulation, we used two variables. The one was the binary variable indicating treatment effect, and the other was continuous variable following the normal distribution. We set the non-linear effect against the log-relative hazards into the continuous variable, and we evaluated the conventional model, PL-Cox, and spline model as the analysis model of the continuous variable. We used two types of knot selecting methods for PL-Cox, which were the ordinary method and our proposed method using the Martingale residuals. Evaluations were done by the power, bias, and mean square error for the treatment effect.

Results: We got the decreased power and the biased coefficient in the conventional model. PL-Cox showed equivalent results as spline regardless of the methods to set knot position.

Conclusion: The conventional model categorizing continuous variable could not adjust the effects of a covariate. PL-Cox was practical model regardless of the knot position.

Keywords: proportional hazards model, log-linearity, piecewise linear Cox model, categorization, spline.

Thesis Advisors: Masako NISHIKAWA, Kunihiko TAKAHASHI, Toshiro TANGO

1 目的

Cox の比例ハザードモデル¹⁾

$$\log(\lambda(t | \mathbf{x}_i) / \lambda_0(t)) = \sum_{j=1}^p \beta_j x_{ij} \quad (1)$$

には、比例ハザード性と対数線形性という仮定がある。対数線形性とは共変量の効果が式(1)の左辺の対数相対ハザードに対して線形であるという仮定で、連続変量に対してこの仮定が成立しない場合は、カテゴリ化を行ったり(従来モデル)、spline 関数によるモデル^{2,3)}が用いられたりする。非線形な関数を直線で近似する折れ線 Cox モデル(PL-Cox)⁴⁾もその一つであるが、その評価は連続変量に

対して行われておらず、かつその時は屈曲点を適当に設定しなければならない。よって本研究は、PL-Cox を連続変量に適用する時の屈曲点の設定方法を提案し、PL-Cox と従来モデル、および spline 関数によるモデルの比較を目的とする。比較はシミュレーションにより行う。

2 方法

2.1 折れ線 Cox モデル

式(1)の右辺を共変量部と呼ぶ。PL-Cox をカテゴリ変量に適用する方法はすでに示されている⁴⁾。連続変量への適用は、事前に与えた m 個の屈曲点 r_l に対し、共変量部を $\sum_{l=0}^m \beta_l (x_i - r_l)_+$ として変数選択を行う。 $(x)_+$ は x の正の部分を示し、 $r_0 = 0$ である。変数選択を行うため、最終

指導教官：高橋邦彦，西川正子，丹後俊郎（技術評価部）

的なモデルで用いられる変数の数は m 以下となる。

2.2 屈曲点の設定方法

共変量部が $\sum_{j=1}^p h_j(x_{ij})$ と表される時、 h_j を共変量 j の関数型と呼ぶ。非線形な関数を折れ線で効率よく表現するためには、その関数が大きく変化する部分に屈曲点が存在すればよい。そこで、関数型 h に対して m 個の屈曲点を置く時、 $g(x) = dh(x)/dx$ として、

$$d(x_{(k)}) = C \sum_{j=2}^k \left| \frac{g(x_{(j)}) - g(x_{(j-1)})}{x_{(j)} - x_{(j-1)}} \right| \quad (2)$$

とし、 $d^{-1}(k/(m+1))$ に屈曲点を置くと h が大きく変化する部分に効率よく置くことができる。ここで、 C は h の値域が $[0,1]$ となるようにする定数で、 $x_{(i)}$ は x_i の順序統計量である。 h には Martingale 残差を滑らかな関数で smoothing した曲線を用いる。この方法を Martingale 法と呼ぶことにする。また、すでに提案されている方法である、各屈曲点に入る症例数が同じになるように区切る方法を Quantile 法と呼ぶことにする。

2.3 シミュレーションの設定

ある疾患の従来治療と新治療の比較試験を考慮したシミュレーションにおいて、変数として新治療の効果を示す二値変量と正規分布に従う連続変量を用いる。連続変量の真の関数型を線形、二次関数、Logistic 関数、Bathtub 型の関数で与え、解析モデルとして連続変量をカテゴリ化する従来モデル、PL-Cox、spline 関数をとした。カテゴリ化の水準は3および4とし、PL-Cox の屈曲点は Martingale 法と Quantile 法により5および10点設定した。評価項目は新治療効果の検出力、偏り、平均二乗誤差とした。新治療効果の検出力が80%となるように症例数を設計した。また、連続変量の数を4まで増加させて同様に評価した。

3 結果

連続変量が一個のとき、4水準にカテゴリ化した従来モデルの検出力は関数型が線形、二次関数、Logistic 関数、Bathtub 型の順に77.3、73.5、71.3、67.6と低下し、それに従って偏りと平均二乗誤差も大きくなった。PL-Cox の結果は屈曲点選択方法によらず、同等であった。また、PL-Cox と spline も同等の結果であった。共変量の数を増加させると、全ての解析モデルで検出力が低下していき、偏りの絶対値も大きくなっていった。また、共変量が複数の時に、Bathtub を除いて、PL-Cox の検出力が spline よりも高くなった。

4 考察

本研究のシミュレーションの状況においては、連続変量のカテゴリ化を行う従来モデルは、真の関数型が線形であっても PL-Cox や spline に比べ検出力が低下した。偏りは100倍程度となり、共変量の効果を適切に調整できていなかった。

共変量が一個の場合、PL-Cox と spline はどの関数型に対しても80%程度の検出力で、偏りも同等であった。Spline は非線形の関数型を適切に表現できることが知られているため、それと同等な PL-Cox は実用的な解析モデルであると言える。

複数の共変量のシミュレーションにおいて、本研究で提案した PL-Cox の屈曲点設定方法は、症例数が等しくなるように設定した方法に比べ、若干検出力が低下した。これは、本研究で提案した方法に改良の余地があることを示している。

更なる PL-Cox の利点を見出すためには、より現実的なシミュレーションによる評価が必要であろう。共変量の分布や関数型を公開されているデータを参考にして設定することにより、現実的な評価が可能となる。

5 まとめ

連続変量をカテゴリ化する従来解析モデルは、共変量の効果を適切に調節できなかった。対数線形性が成立しない状況において、連続変量に適用した PL-Cox は、spline 関数を用いたモデルと同等な解析モデルであった。

参考文献

- 1) Cox, D.R. Regression methods and life tables. Journal of the Royal Statistical Society 1972; 34:187-202.
- 2) Hastie, T., and Tibshirani, R. Exploring the nature of covariate effects in the proportional hazards model. Biometrics 1990;46:1005-16.
- 3) Gray, R. J., Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. Journal of the American Statistical Association 1992;87:942-51.
- 4) Akazawa, K., Nakamura, T., and Palesch, Y. Power of logrank test and Cox regression model in clinical trials with heterogeneous samples. Statistics in Medicine 1997;16:583-97.
- 5) 井元清哉, 小西貞則. B-スプラインによる非線形回帰モデルと情報量基準. 統計数理1999;47-2:359-73.

〈教育報告〉

平成18年度専門課程Ⅱ
生物統計分野

randomization model における平均値の差の検定の 検出力および信頼区間に関する研究

白岩健

Permutation test and confidence interval for the difference in means under the randomization model

Takeru SHIROIWA

Abstract

Objective: There are two problems in the statistical analysis of clinical trials. (1) First, though the assumption of random sampling is not met, the data of clinical trials are analyzed based on the population model or invoked population model. (2) Allocation methods to analyze clinical trial data are not taken into account. To solve these problems, we used the randomization model and compared it with invoked population model

Method: Based on the randomization model, we calculated p-value of permutation test for the difference in means by Monte Carlo simulation and confidence interval by Robins-Monro search process. The p-value, coverage probability and the length of confidence interval were used to compare the performance of three statistical methods; (a) permutation test (b) ordinary t-test (c) analysis of covariance (ANCOVA). We considered the three allocation methods; (a) complete randomization (b) stratified randomization (c) minimization method.

Results: The analysis based on the randomization model, it was shown that both tests and confidence interval were the nearly equal performance with ANCOVA though randomization test has a little lower power and the longer confidence interval. These results are consistent regardless of the distribution of error terms. The minimization method was superior to complete randomization and stratified randomization especially when the marginal distribution of covariates was not balanced.

Conclusion: The analysis based on the randomization model can be used for the primary statistical analysis of clinical trials without model assumption and the adjustment of covariates.

Keywords: clinical trial, randomization model, MonteCarlo, permutation test, nonparametric

Thesis Advisors: Masako NISHIKAWA, Toshiro TANGO

I. 目的

現在の臨床試験における統計解析には以下の二つの課題が存在している。

(1) 臨床試験では試験参加施設に来院した患者のうち同意のとれたもののみを試験に組み込むため、被験者は母集団からの無作為抽出とはみなせない。にもかかわらず無作為抽出を仮定した population model や被験者と母集団

に同一のパラメータ分布を仮定した invoked population model などが用いられている。

(2) 通常の臨床試験では様々な割り付け方法が用いられているにもかかわらず割り付け方法を考慮した解析がなされていない。

以上の課題を検討するために無作為抽出やパラメータ分布の仮定を置かない randomization model に基づいたデータ解析を行い invoked population model における解析とその性能を比較する。

指導教官： 西川正子, 丹後俊郎 (技術評価部)

II. 研究デザインと方法

以下のシミュレーションにおいて、ある試験薬とプラセボを比較する臨床試験を想定する。この臨床試験では被験者96人がいずれかの治療に割り付けられる。被験者は治療の効果に対して2つの共変量「性別 (α)」「重症度 (β)」を持っており、それらは以下のような分布を仮定している。

table1: sample1: distribution of covariate

		Severity (β)		
		minor ($\beta 1$)	moderate ($\beta 2$)	serious ($\beta 3$)
sex (α)	male ($\alpha 1$)	16	16	16
	female ($\alpha 2$)	16	16	16

table2: sample2: distribution of covariate

		severity (β)		
		minor ($\beta 1$)	moderate ($\beta 2$)	serious ($\beta 3$)
sex (α)	male ($\alpha 1$)	15	23	10
	female ($\alpha 2$)	15	23	10

このとき、試験結果として得られる反応値 $y_j (j=1, \dots, 96)$ は以下の式で表されるとする。

$$y_j = \delta T_j + 3\alpha_j + \beta_j + \varepsilon_j$$

δ は effect size であり、 T_j は治療群なら1、プラセボ群なら0をとる。共変量の大きさは「 $\alpha 1=1, \alpha 2=0$ 」「 $\beta 1=0, \beta 2=1, \beta 3=2$ 」とし、 ε は誤差項である。誤差項には標準正規分布を含め8種類の分布を用いた。

この2群の反応値の差を統計的に検出する手法として、randomization model にもとづく平均値の差の検定 (R-test) を用いて p 値と信頼区間を求めた。R-test の性能を比較する対照として、invoked population model にもとづく通常の t 検定 (t-test) と共変量を調整した共分散分析 (ANCOVA) も同時に計算した。

割り付け方法としては完全無作為化法 (CR)、層別無作為化法 (SR)、最小化法 (MIN) の3種類を用い、割り付け方法ごとに3種類の分析方法の比較を行った。

R-test の計算方法は以下の通りである。

1. 並び替え検定

Randomization model における2群の平均値の差の p 値は以下の式で求める。

$$p\text{-value} = \sum_{a=1}^{\Omega} I(|S_a - \bar{S}| \geq |S_{obs} - \bar{S}|) \Pr(A = a)$$

$I(\cdot)$ は indication function とよばれ括弧内の条件が成立するときは1、それ以外では0を示す。S は平均値の差の検定における統計量であり、

$$S = \sum_{j=1}^n (Y_j - \bar{Y}_j) T_j$$

で計算される。 Y_j は被験者 j の試験で得られた値である。

実際の p 値の算出にあたっては1000回のモンテカルロシミュレーションを行った。

2. 信頼区間の構成

Robbins-Monro search process にもとづく Garthwaite の方法に従って randomization model における平均値の差の信頼区間を構成した。steps 数 (繰り返し数) は2000回で行い、信頼区間の評価方法として coverage probability と信頼区間の幅の長さを求めた。

III. 結果

検定・信頼区間ともに randomization model にもとづく解析はやや性能が低く、信頼区間では coverage probability はほぼ同様だが幅が広がる傾向があったもののほぼ ANCOVA と同程度の性能を示した。この結果は誤差項の分布形には依存しなかった。割り付け方法としては、特に共変量の周辺分布が不揃いの時に最小化法が優れていた。

図1は sample2をSRで割り付けたときの検出力、図2は sample2をMINで割り付けたときの検出力、table3は sample2の coverage probability と信頼区間の幅をそれぞれ示す。誤差項はすべて正規分布を仮定している。

IV. 考察

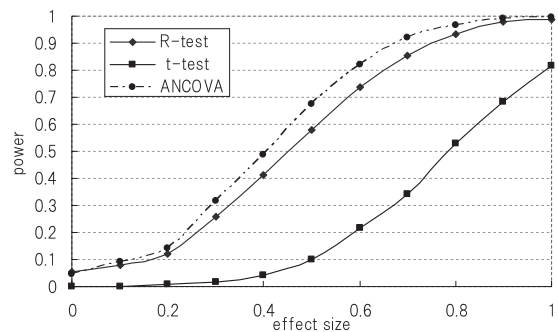


図1: SRで割り付けたときの検出力

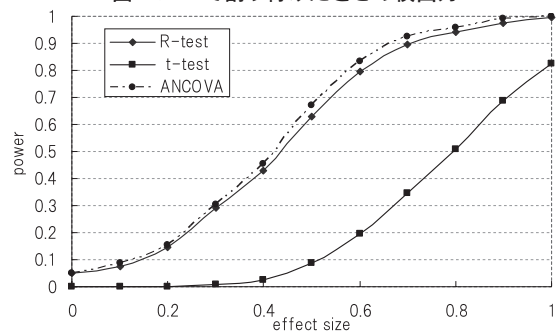


図2: MINで割り付けたときの検出力

Table3: coverage probability and length of sample2

(under $\varepsilon_i \sim N(0,1)$)

	coverage probability (mean)			the length of CI (mean)		
	R test	ANCOVA	t-test	R test	ANCOVA	t-test
CR	0.948	0.952	0.948	1.602	0.821	1.584
SR	0.951	0.951	1.000	0.895	0.807	1.582
MIN	0.950	0.954	1.000	0.859	0.808	1.582

randomization model に基づいた解析は、モデルの仮定

や解析時の共変量調整なしにほぼ ANCOVA と同程度の性能を示した。

一般に臨床試験の解析においては、primary な解析として線型モデルを仮定しない t 検定や Wilcoxon 検定が、secondary な解析として線型モデルを仮定して共変量を調整した ANCOVA が用いられることが多い。しかしこのような手順では、両者の結果が食い違うときその解釈は困難

となる。randomization model にもとづく解析を使用すれば、モデルを仮定することなしに共変量の調整が行えるので、2 段階の解析が不要であり一度ですむので有用である。

ただし、p 値や信頼区間の算出はモンテカルロシミュレーションを用いた方法であり、シミュレーションにおける最適な反復回数の設定等は今後の課題である。