数理解析研究所講究録
第1804巻 2012年 90-106

Covariate-Adaptive designを用いた時の並べ替え検定の性能評価

北里大学大学院　薬学研究科　臨床薬学研究・教育センター　臨床統計学
Department of Clinical Medicine (Biostatistics), Graduate School of Pharmaceutical Sciences, Kitasato University
高橋　政樹(Masaki Takahashi)
高橋　史朗(Fumiaki Takahashi)
竹内　正弘(Masahiro Takeuchi)

1.序論

無作為化試験を実施する目的のひとつには試験結果に影響を与える変数(共変量)を考慮しない下での治療効果(周辺治療効果)を推定することがある。共変量に不均衡が生じた場合には、周辺治療効果の推定にバイアスが生じてしまう。そのため、事前に共変量が特定できている場合には、被験者の割り付けを行う際に意図的に共変量の均衡を保ち、主要評価項目への共変量の不均衡を抑える方法が採られている。その主なものにStratified designおよびCovariate-Adaptive Design(以下CA)と呼ばれる方法がある。また、試験終了後に共変量の存在が明らかになった場合には、事後的に共変量を解析で考慮する。例えば、結果変数が二値変数で共変量を考慮した検定にはロジステック回帰分析、Zhangにより提案された検定(Zhangの検定)[1]、Kochにより提案された検定(Kochの検定)[2] がある。

ロジステック回帰分析は結果変数が二値変数で共変量を調整する方法として良く知られている。このロジステック回帰分析では、正しいモデルを特定しない限りバイアスが生じることが知られている[3]-[5]。しかし、データに基づいてモデルに当てはまりの良い共変量を特定することはできるものの、その共変量が正しいモデルに含まれる共変量であるか否かを判断することはできない。加えて、ロジステック回帰分析による解析結果は周辺治療効果における結果ではなく、ロジステック回帰分析に用いられた共変量の均衡が保たれたという条件のもとでの解析結果となる。その為に、共変量を考慮した検定で周辺治療効果を求めるためには別の検定を用いることになる。その方法がZhangの検定およびKochの検定である。Zhangの検定はセミパラメトリックな方法であり、Kochの検定はノンパラメトリックな方法である。この2つの検定には、Zhangの検定がKochの検定を包括する方法であるという関係が存在する。

今日の臨床試験の多くはStratified designおよびCAで割り付けを行い、周辺治療効果に対して共変量を考慮しない仮説検定が実施されている。CAを用いた状況下で周辺治療効果に対してこのような仮説検定を行うと、Type I errorが名義的な有意水準を大きく下回ることが報告されている[6][7]。このような場合に症例数が多ければ、Zhangの検定およびKochの検定を用いて共変量を調整することができる。なぜなら、Zhangの検定やKochの検定が必要とする共変量と治療群の独立性を、CAでは漸近的に示すことができるからである[7]。症例数が小さい場合には共変量と治療群の独立性が成り立たなくなり、Type I errorや検出力に何らかの影響を及ぶと考えられる。そこで、CAを用いた状況で症例数の大小に係わらず、Type I errorをより名義的な有意水準付近の値

で保持する検定が提案された。それは、2009 年に Hasegawa らにより提案された並べ替え検定[8]や、2010 年に Shao らにより提案された Bootstrap を用いた検定(Bootstrap 検定) [9]である。

2 つの方法は提案こそされているが、性能を比較検討したという報告はされていない。また、それぞれの検定は結果変数が連続値変数の場合に機能することが示されているが、二値変数へ適用した際にどのような性能を示すかに関しては報告されていない。そこで、併合分散を用いる割合の差の検定が保守的になる理論的な根拠を与えるとともに提案法の性能を評価する。

次項より、2 章として本研究で用いた検定に関する説明を行い、次いでシミュレーションの設定条件を説明する。その後、3 章として併合分散を用いる割合の差の検定が保守的になる理論的な根拠を示し、シミュレーションの結果を提示する。そして、4 章として結果に対する考察を行う。

2.方法

2.1. 二値変数の無作為化比較試験における周辺治療効果の推定

2.1.1 無作為化比較試験

Y を結果変数、X を共変量、Z を各群への割り付けの指示変数とした二群間比較試験を想定した。それぞれの表記は以下に示す。なお本研究では、結果変数が二値変数である場合に着目した。

$$Y\begin{cases}0：生存\\1：死亡\end{cases}\quad Z\begin{cases}0：control\\1：treatmet\end{cases}\quad X：共変量$$

被験者の各群への割り付けは、CA の 1 つである最小化法を想定した。最小化法とは各群における共変量の均衡を保つように被験者を割り付ける方法で、無作為化を多少犠牲にしてでも共変量の均衡を保ちにいく方法である。最小化法により均衡をとる共変量は以下の 3 通りを想定した。

Situation 1 :発生させた全ての共変量を考慮させた場合
Situation 2:共変量を誤特定した場合
Situation 3:共変量の調整後に、交互作用項が明らかになった場合

2.1.2 並べ替え検定(Permutation test)

並べ替え検定とは観測されたデータに基づいて算出される検定統計量よりも極端な値を取る確率を直接的に得る方法である。

基準となる検定統計量(S_{obs})は以下の式に従って算出される。

$$S=\sum_{i=1}^{n}(Y_i-\bar{Y}_n)T_Z$$

ここで、i は症例数を表す（ $i=1,2,\cdots,N$ ）。g は各群への割り付けを表す指示変数である（ $g=1\ or\ 0$ ）。ただし、$\bar{Y}_n$は$\bar{Y}_n=\sum Y_i/N$から得られる。T_Zは treatment であれば 1 を、control であ

れば 0 を取るものとする。
その後、基準よりも極端な値を取る確率を算出するために、新たな並べ替え列(M)を作成する。その並べ替え列に基づいて検定統計量(S_m)を計算する。

$$S_m = \sum_{i=1}^{n} (Y_i - \bar{Y}_n) T_Z$$

その後、作成した並べ替え列に基づいて算出される検定統計量が基準となる検定統計量以上の値を取った際に 1 を、それ以外は 0 である指示変数と並べ替え列の発現確率の積から、基準検定統計量より極端な値を取る確率を直接計算する。

$$p = \sum_{m=1}^{\Omega} I(|S_m - \bar{S}| \geq |S_{obs} - \bar{S}|) Pr(M = m)$$

ここで、$\bar{S} = \sum s/M$、Ωは並び替え列数を表す。Ω_Uは各群の比をまったく考慮しないで並び替え列を表し、例えばすべての被験者が Control または Treatment に属することを許容する。Ω_Cは観測された各群の比を保持して並べ替え列を作成することを表す。
本研究においては、並べ替え検定を行う際に被験者の登録順を固定して、Ω_C条件下で並び替え検定を実施した。

2.1.3 Bootstrap 検定

Bootstrap 検定は Bootstrap 法を用いて標準誤差を推定し、検定に用いる方法である。Bootstrap 法とは、標本の繰り返し復元抽出データから推定量の経験分布を求め、その分布を興味のある推定量の近似分布として、興味のある母数を推測する方法である。本研究では、観測されたデータから同じ大きさのデータを繰り返し復元抽出し、復元抽出されたデータよりパラメータを算出し、そのパラメータ間における標準偏差を推定し、その値を標本の標準誤差とした。

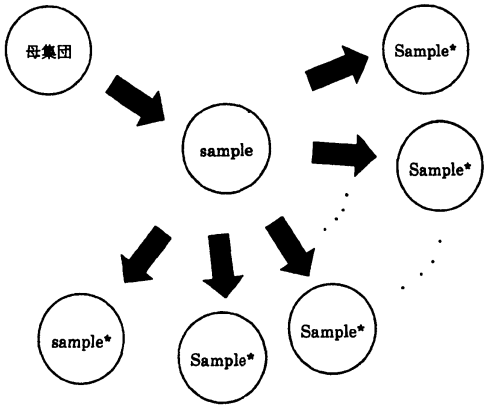

Fig.1 The image of Bootstrap

各群のイベントの発現割合の差を算出する。それを Bootstrap 法により推定された標準誤差 $\widehat{Var_B}(\hat{\rho}^*)$で除し、検定統計量を求める。ただし、$\hat{\rho}^*$は復元抽出されたデータより算出されるパラメータを表す。

$$Z = \left| \widehat{P_1 - P_0} \right| \Big/ \sqrt{Var_B(\hat{\rho}^*)}$$

ただし、$P_1 = \sum Y_1/N_1$、N_1は treatment 群に属する症例数、$P_0 = \sum Y_0/N_0$、N_0は control 群に属する症例数である。

求めた検定統計量が正規分布に従うことから、検定統計量によりも極端になる確率を算出する。

2.1.4 割合の差の検定

割合の差の検定とは共変量は考慮していない周辺治療効果を評価するのに用いる仮説検定である。

各群のイベントの発現確率の差を算出する。そして、併合分散を推定し、その併合分散を検定に用いる。併合分散とは、各群への割り付けを考慮していない被験者全体のバラツキである。

$$Z = \left| \widehat{P_1 - P_0} \right| \Big/ \sqrt{\bar{P}((1-\bar{P}))/\left(1/N_1 + 1/N_0\right)}$$

ただし、$\bar{P}$は$\bar{P} = \left(\sum Y_1 + \sum Y_0\right) \Big/ (N_1 + N_0)$により算出することができる。

2.1.5 Fisher's exact test

Fisher's exact test は割合の差の検定と同様、周辺治療効果を評価するために用いる仮説検定である。

Fisher's exact test は、超幾何分布に基づいた検定である。両側検定を実施する際には、観測されたデータy_{obs}より極端な値を示す確率をP値として算出する方法と、$Pr(Y \leq y_{obs})$または$Pr(Y \geq y_{obs})$のいずれか小さい値の2倍でP値を算出する方法がある。今回の研究では、前者の方法を用いた。

観測されたデータ(a)に対する確率は以下のように表現できる。

$$Pr(Y = a) = \frac{\binom{n_1}{a}\binom{N-n_1}{m_1-a}}{\binom{N}{m_1}}$$

そして、P値は以下のように表現できる。

$$P - value = \sum_{\{a|Pr(Y=a) \leq Pr(Y=y_{obs})\}} Pr(Y = a)$$

2.1.6 logistic 回帰分析

logistic 回帰分析とは、結果変数を二値変数とし、各群の指示変数や共変量を説明変数とした際に一般的に用いられる解析法である。

$$\text{logit}\{E(Y|Z)\} = \text{logit}\{Pr(Y = 1|Z)\} = \log\left\{\frac{Pr(Y = 1|Z)}{1 - Pr(Y = 1|Z)}\right\} = \alpha + bX + \beta Z$$

この logistic 回帰分析において、周辺治療効果を求めるためには、各群へ割り付けを表す指示変数のみを説明変数としてモデル化し、その回帰係数βを検定する。

$$\mathrm{logit}\{E(Y|Z)\} = \mathrm{logit}\{\Pr(Y=1|Z)\} = \log\left\{\frac{\Pr(Y=1|Z)}{1-\Pr(Y=1|Z)}\right\} = \alpha + \beta Z$$

また、logistic 回帰分析における回帰係数 β の推定値は対照群と治療群の対数オッズ比として表現することができる。

$$\beta = \mathrm{logit}\{\Pr(Y=1|Z=1)\} - \mathrm{logit}\{\Pr(Y=1|Z=0)\}$$

2.1.7 Zhang の検定

Zhang らにより提案された方法は、セミパラメトリック共変量調整である。具体的には、結果変数と共変量の関係をモデル化し、それを用いて共変量の偏りを調整する。

共変量を考慮せずに logistic 回帰分析を用いて周辺治療効果の回帰係数を求めるとすると以下の式を解くことで求められる。

$$m(Y,Z;\Theta) = \binom{1}{Z}\left\{Y - \frac{\exp(\alpha+\beta Z)}{1+\exp(\alpha+\beta Z)}\right\}$$

$$\sum_{i=1}^{N} m(Y_i, Z_i; \Theta) = 0$$

Zhang らはこの β の推定に結果変数と共変量の関係をモデル化した補正項 $q_g(X)$ を用いることにより、共変量を考慮した周辺治療効果の回帰係数を求める方法を提案した。

$$m(Y,Z,X;\Theta) = m(Y,Z;\Theta) - \sum_{g=0}^{1}\{I(Z=g) - \pi_g\}q_g(X)$$

ただし、$\pi_g = \frac{1}{N}\sum_{i=1}^{N} I(Z_i = g)$、$q_g(X_i) = E\{m(Y_i, Z_i; \Theta)|X_i, Z_i = g\}$、$\sum_{i=1}^{N} m(Y_i, Z_i, X_i; \Theta) = 0$ とする。

Zhang らにより提案された方程式より以下の式を求めることができる。

$$\hat{P}_0 = \bar{Y}_o - \frac{N_1}{N}\{\bar{q}_0^*(X_i, \delta)_{cont} - \bar{q}_0^*(X_i, \delta)_{treat}\}$$

$$\hat{P}_1 = \bar{Y}_1 - \frac{N_0}{N}\{\bar{q}_1^*(X_i, \varepsilon)_{treat} - \bar{q}_1^*(X_i, \varepsilon)_{cont}\}$$

$$\beta = \log\left(\frac{\hat{P}_1}{1-\hat{P}_1}\right) - \log\left(\frac{\hat{P}_0}{1-\hat{P}_0}\right)$$

$$\text{notation}: \bar{q}_1^*(X_i, \delta) = E(Y_i|X_i, Z_i = 1) \quad \bar{q}_0^*(X_i, \varepsilon) = E(Y_i|X_i, Z_i = 0)$$

$\bar{q}_1^*(X_i, \varepsilon)_{treat}$:治療群の結果変数と共変量をモデル化、治療群のデータを代入し平均を算出

$\bar{q}_1^*(X_i, \varepsilon)_{cont}$:治療群の結果変数と共変量をモデル化、対照群のデータを代入し平均を算出

$\bar{q}_0^*(X_i, \delta)_{cont}$:対照群の結果変数と共変量をモデル化、対照群のデータを代入し平均を算出
$\bar{q}_0^*(X_i, \delta)_{treat}$:対照群の結果変数と共変量をモデル化、治療群のデータを代入し平均を算出

$q_g^*(X)$ のモデル化には一般的に logistic 回帰分析が用いられることが多い。

2.1.8Koch の検定

Koch らにより提案された方法は、ノンパラメトリック共変量調整法である。具体的には、群間における共変量の不均衡を結果変数と共変量の共分散行列を用いて調整する。

Koch らにより提案された方法を用いるとイベントの発現割合の差は

$$\hat{P}_1 - \hat{P}_0 = (\bar{Y}_1 - \bar{Y}_0) - \mathbf{Var}'_{\mathbf{YX}}\mathbf{Var}_{\mathbf{XX}}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_0)$$

と表現することができる。但し、$\mathbf{Var}_{\mathbf{YX}}$は結果変数と共変量の分散共分散行列を表し、$\mathbf{Var}'_{\mathbf{YX}}$は$\mathbf{Var}_{\mathbf{YX}}$転置行列を、$\mathbf{Var}_{\mathbf{XX}}^{-1}$は共変量の分散共分散行列の逆行列を表す。$\bar{Y}_1$および$\bar{Y}_0$は結果変数の平均を表し、$\bar{\mathbf{X}}_1$および$\bar{\mathbf{X}}_0$は共変量の平均値ベクトルを表すものとする。

この仮説検定に用いられる分散(Var_K)はイベントの発現割合の差と同様に分散共分散行列により補正され

$$Var_K = Var_{YY} - \mathbf{Var}'_{\mathbf{YX}}\mathbf{Var}_{\mathbf{XX}}^{-1}\mathbf{Var}_{\mathbf{YX}}$$

である。ただし、Var_{YY}は結果変数の分散とする。

以上の結果を用いると検定統計量は

$$Q_g = \frac{(\hat{P}_1 - \hat{P}_0)^2}{Var_K}$$

と表現することができ、この検定統計量が自由度 1 のχ二乗分布に従うことを利用して、求めた検定統計量よりも極端になる確率を算出する。

2.2.シミュレーションによる検討

本研究では、結果変数を二値変数とし、最小化法を用いて各群への割り付けを行った場合の並び替え検定、Bootstrap 検定の性能を評価した。評価指標としては検出力および Type I error を用い、比較対象は割合の差の検定、Zhang の検定および Koch の検定を用いた。

シミュレーションは、症例数 192 例の二群比較を目的とする無作為化比較試験を想定した。各症例は 4 つの共変量を有し、共変量をX_1, X_2, X_3, X_4と表現する。X_1は 0 または 1 の二値変数を想定し、X_1が 1 または 0 である確率をそれぞれ 0.5 とした。X_2は 0 または 3 の二値変数を想定し、X_2が

3 である確率を 0.33 とし、0 である確率を 0.67 とした。X_3は 0、1、2 または 5 を取ることを想定した。X_3が 0 である確率を 0.5、1 である確率を 0.3、2 である確率を 0.1、5 である確率を 0.1 とした。X_4は 0、1 または 1.5 を取ることを想定した。X_4が 0 である確率を 0.45 とし、1 である確率を 0.3 とし、1.5 である確率を 0.25 とした。

交互作用項に関しては各症例の共変量のX_3およびX_4を掛け合わせることにより得た。

2.2.1 結果変数発生のメカニズム

最小化法により割り付けを行い、その後 logistic モデルに基づき結果変数を発生させる。

・最小化法による割り付け

最小化法は、Pocock-Simon により提案された方法を用いた。Pocock-Simon により提案された方法では群間における共変量の均衡を図る確率 P を設定することができる。本研究では P に 1、4/5、2/3 の 3 通りの場合を想定した。

無作為化試験の項に記載した通り、最小化法により調整された共変量は 3 通りを想定した。

	最小化法で調整した共変量	検定に考慮した共変量
発生させた全ての共変量を考慮させた場合	X_1, X_2, X_3, X_4	X_1, X_2, X_3, X_4
共変量を誤特定した場合	X_2, X_3	X_2, X_3
交互作用項の存在が確認された場合	X_2, X_3	$X_1, X_2, X_3, X_4, X_3 \cdot X_4$

不均衡を改善する確率と最小化法に考慮する共変量、それぞれを組み合わせ全部で 9 通りの場合を想定して割り付けを行った。

・結果変数の発生

イベントの発現確率を以下の式で推測した。

$$\hat{P}_g = \frac{\exp(q \cdot z + X_1 - X_2 + X_3 - X_4)}{1 + \exp(q \cdot z + X_1 - X_2 + X_3 - X_4)} \quad (1)$$

ただし、q は検出したい差を表している。

推測した発現確率をパラメータとし Bernoulli 分布より、0 または 1 の結果変数を発生させた。

2.2.2 シミュレーションで用いた検定

・割合の差の検定

・Fisher's exact test

・並び替え検定

並び替え検定をシミュレーションで実施する上では Monte Carlo Simulation を用いた。検定を実施

するために 999 回の並べ替えを行い、確率を算出した。

・Bootstrap 検定

Bootstrap 検定においても Monte Carlo Simulation を用いてシミュレーションを実施した。復元無作為抽出回数は 500 回とした。

・Zhang の検定

$q_g^*(X)$のモデル化は logistic モデルを用いた。

結果変数と共変量のモデル化を行う際に、共変量X_3, X_4をダミー変数化する必要があった。その際は、Situation 1 および Situation 2 においてはそれぞれに対応するダミー変数(Dx)を用いた。しかし、交互作用項が存在する場合は、モデルへの収束が悪かったために、ダミー変数を変更した。Z3 が 2 または 5 であるときに対応するダミー変数を 1 とした。加えて、交互作用項が 0 以外の時に対応するダミー変数を 1 とした。以外のダミー変数は Situation 1 および Situation 2 と同じものを用いた。

共変量	ダミー変数
X3=2 X3=5	Dx31=1
X3・X4=0 以外	Dx3x4=1

・Koch の検定

各シミュレーションは有意水準(α)を 0.05 とし、5000 回、繰り返した。

3.結果

3.1CA を用いた際に割合の差の検定で用いられる分散が増大していることの証明

Y_1を Group 1 の結果変数、Y_2を Group 2 の結果変数とし、結果変数は 0 または 1 の 2 値変数とする。Group 1 の症例数をN_1、Group 2 の症例数をN_2で表し、全症例数 N は$N = N_1 + N_2$とする。Y_1の有効割合をP_1、Y_2の有効割合をP_2とすると、Y_1およびY_2はそれぞれ二項分布に従い、$Y_1 \sim bin(N_1, P_1)$、$Y_2 \sim bin(N_2, P_2)$のように表わすことができる。また、Z_1に着目した際の結果変数をY_{1Z1}、Y_{2Z1}、有効割合をP_{1Z1}、P_{2Z1}、症例数をN_{1Z1}、N_{2Z1}とする。同様にZ_2の結果変数をY_{1Z2}、Y_{2Z2}、有効割合をP_{1Z2}、P_{2Z2}、症例数をN_{1Z2}、N_{2Z2}とする。さらに、Z_1およびZ_2は互いに独立であるために共分散は 0 となる。よって$Y_{1Z1} \sim bin(N_{1Z1}, P_{1Z1})$となることから分散$var(Y_{1Z1})$は、

$$var(Y_{1Z1}) = \frac{P_{1Z1}(1 - P_{1Z1})}{N_{1Z1}}$$

となる。同様に$Y_{1Z2} \sim bin(N_{1Z2}, P_{1Z2})$、$Y_{2Z1} \sim bin(N_{2Z1}, P_{2Z1})$、$Y_{2Z2} \sim bin(N_{2Z2}, P_{2Z2})$となることから、

$$var(Y_{1Z2}) = \frac{P_{1Z2}(1 - P_{1Z2})}{N_{1Z2}}$$

$$\mathrm{var}(Y_{2Z1}) = \frac{P_{2Z1}(1-P_{2Z1})}{N_{2Z1}}$$

$$\mathrm{var}(Y_{2Z2}) = \frac{P_{2Z2}(1-P_{2Z2})}{N_{2Z2}}$$

ここで、Z_1およびZ_2を共変量とし、互いに独立とする。結果変数と共変量の関係は、

$$\mathrm{logit}(Y_1/N_1) = \mu + \alpha \cdot Z_1 + \beta \cdot Z_2$$

$$\mathrm{logit}(Y_2/N_2) = \mu + \gamma + \alpha \cdot Z_1 + \beta \cdot Z_2$$

のように、logistic モデルで表現することができる。logistic モデルにおいて、μを共通の薬効とし、γを検出したい薬効とする。

仮説検定の帰無仮説下では治療群間および共変量によるサブグループ間にも差はないという立場に立っていますので、共変量のサブグループの有効割合に対して各サブグループの症例を重みとして重み付き平均(P_T)を求め、その平均に対応する分散を求めることになる。その求めた分散を検定に用いることなる。

$$P_T = \frac{N_{1Z1} \cdot P_{1Z1} + N_{2Z1} \cdot P_{2Z1} + N_{1Z2} \cdot P_{1Z2} + N_{2Z2} \cdot P_{2Z2}}{N_{1Z1} + N_{2Z1} + N_{1Z2} + N_{2Z2}}$$

$$= \frac{(Y_1+Y_2)}{N}$$

$$\mathrm{Var}(P_T) = \mathrm{Var}\left(\frac{(Y_1+Y_2)}{N}\right)$$

$$= \frac{N_1\bar{P}(1-\bar{P}) + N_2\bar{P}(1-\bar{P})}{N^2}$$

$$= \bar{P}(1-\bar{P}) \Big/ \left(1/N_1 + 1/N_0\right)$$

ただし、$\bar{P} = (Y_1+Y_2)/(N_1+N_2)$とする。

しかし、Covariate Adaptive Design を用いた際の帰無仮説では、治療群間に差はないという立場は同じだが、共変量のサブグループにおける有効割合には差があるという立場なので、まずサブグループの有効割合に対応する分散を求め、その分散に対して各サブグループの症例数を重みとして、平均を求める。その分散($\mathrm{Var}_{\mathrm{Perm}}$)を検定に用いている事になる。

$$\mathrm{Var}_{\mathrm{Perm}} = \frac{N_{1Z1} \cdot \mathrm{Var}(P_{1Z1}) + N_{2Z1} \cdot \mathrm{Var}(P_{2Z1}) + N_{1Z2} \cdot \mathrm{Var}(P_{1Z2}) + N_{2Z2} \cdot \mathrm{Var}(P_{2Z2})}{N_{1Z1} + N_{2Z1} + N_{1Z2} + N_{2Z2}}$$

$$= \frac{P_1 z_1(1-P_1 z_1) + P_2 z_1(1-P_2 z_1) + P_1 z_2(1-P_1 z_2) + P_2 z_2(1-P_2 z_2)}{2N}$$

以上の事から、上に凸な二次関数の性質より、$Var(P_T)$よりもVar_{Perm}の方が小さな値を示すことが明らかになった。

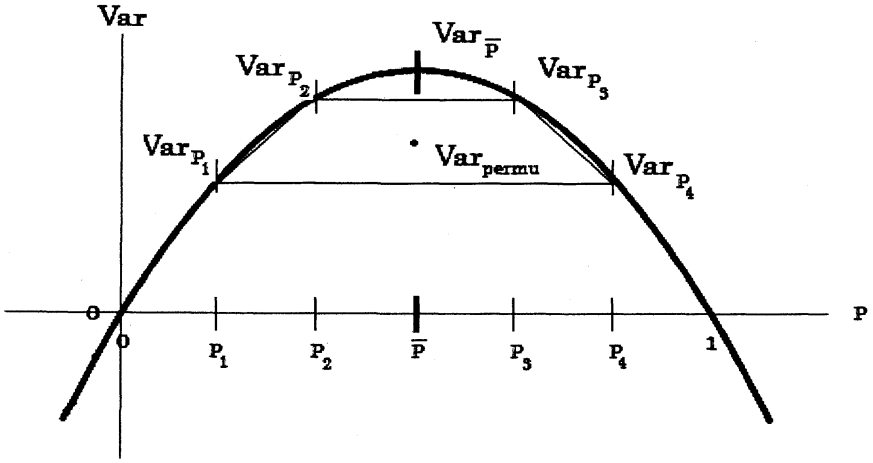

Fig.2 The image of variance under the null hypothesis

3.2.1.発生させた共変量すべてを考慮した場合

3.2.1.1. 最小化法における共変量の均衡を図る確率 P=1

発生させた共変量すべてを考慮させ、最小化法における共変量の均衡を保持する確率を P=1 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0164、Fisher's exact test で 0.0108、並べ替え検定で 0.0484、Bootstrap 検定で 0.0540、Koch の検定で 0.0526、Zhang の検定で 0.0478 であった。

検出力は Zhang の検定、Bootstrap 検定、Koch の検定、並べ替え検定、割合の差の検定、Fisher's exact test の順であった。

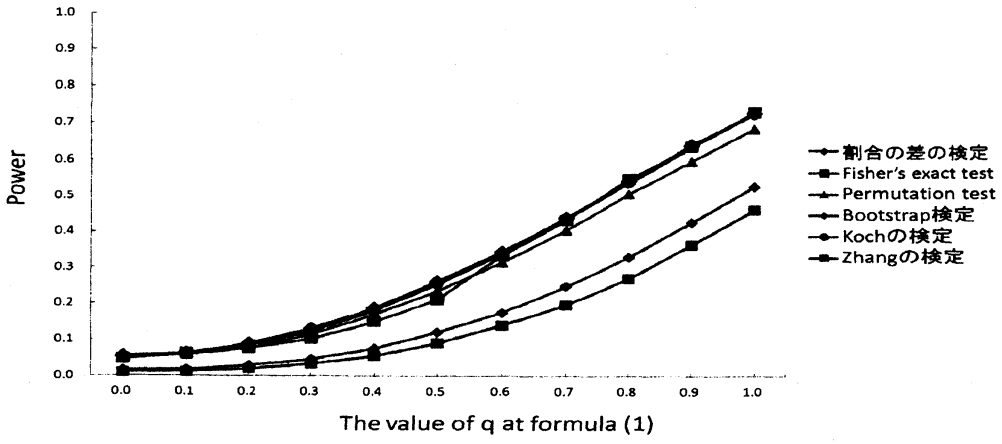

Fig.3 Simulation results power for 3.2.1.1 (α=0.05 5000runs N=192 K=4)

3.2.1.2. 最小化法における共変量の均衡を図る確率 P=4/5

発生させた共変量すべてを考慮させ、最小化法における共変量の均衡を保持する確率を P=4/5 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0188、Fisher's exact test で 0.00134、並べ替え検定で 0.0488、Bootstrap 検定で 0.0474、Koch の検定で 0.0536、Zhang の検定で 0.0481 であった。

検出力は Koch の検定、Zhang の検定、Bootstrap 検定、並べ替え検定、割合の差の検定、Fisher's exact test の順で有った。

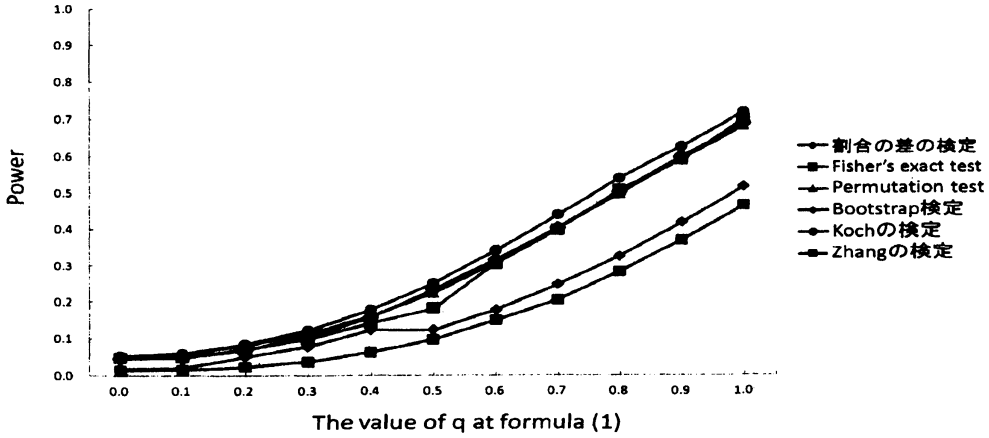

Fig.4 Simulation results power for 3.2.1.2 (α=0.05 5000runs N=192 K=4)

3.2.1.3. 最小化法における共変量の均衡を図る確率 P=2/3

発生させた共変量すべてを考慮させ、最小化法における共変量の均衡を保持する確率を P=2/3 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0226、Fisher's exact test で 0.0156、並べ替え検定で 0.0528、Bootstrap 検定で 0.0466、Koch の検定で 0.0496、Zhang の検定で 0.0538 であった。

検出力は Koch の検定、Zhang の検定、並べ替え検定、Bootstrap 検定、割合の差の検定、Fisher's exact test の順で有った。

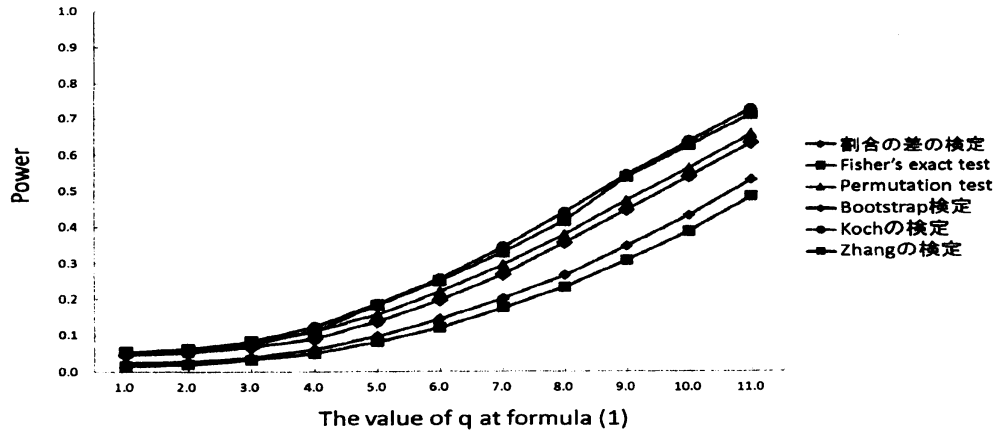

Fig.5 Simulation results power for 3.2.1.3 (α=0.05 5000runs N=192 K=4)

3.2.2.共変量を誤特定した場合

3.2.2.1. 最小化法における共変量の均衡を図る確率 P=1

共変量を誤特定し、最小化法における共変量の均衡を保持する確率を P=1 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0204、Fisher's exact test で 0.0114、並べ替え検定で 0.0506、Bootstrap 検定で 0.0488、Koch の検定で 0.0486、Zhang の検定で 0.0382 であった。

検出力は Zhang の検定、Bootstrap 検定、Koch の検定、並べ替え検定、割合の差の検定、Fisher's exact test の順で有った。

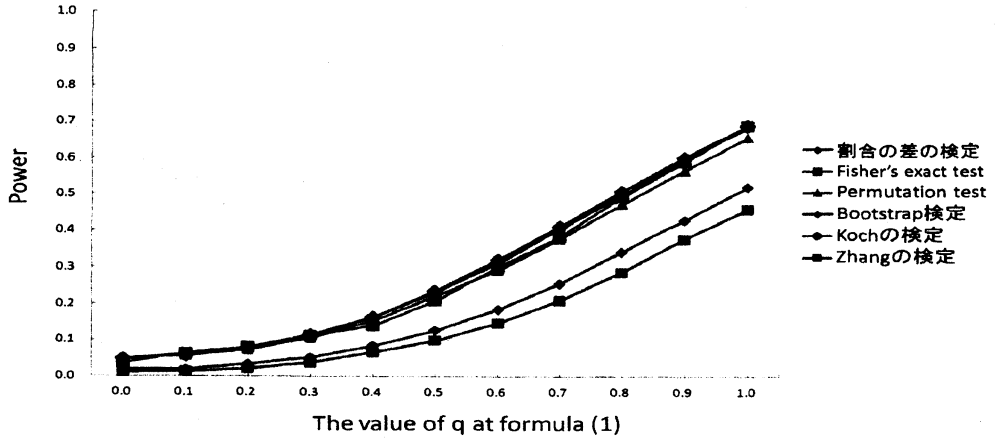

Fig.6 Simulation results power for 3.2.2.1 (α=0.05 5000runs N=192 K=4)

3.2.2.2. 最小化法における共変量の均衡を図る確率 P=4/5

共変量を誤特定し、最小化法における共変量の均衡を保持する確率を P=4/5 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0204、Fisher's exact test で 0.0140、並べ替え検定で 0.0476、Bootstrap 検定で 0.0510、Koch の検定で 0.0526、Zhang の検定で 0.0479 であった。

検出力は Koch の検定、Bootstrap 検定、並べ替え検定、Zhang の検定、割合の差の検定、Fisher's exact test の順で有った。

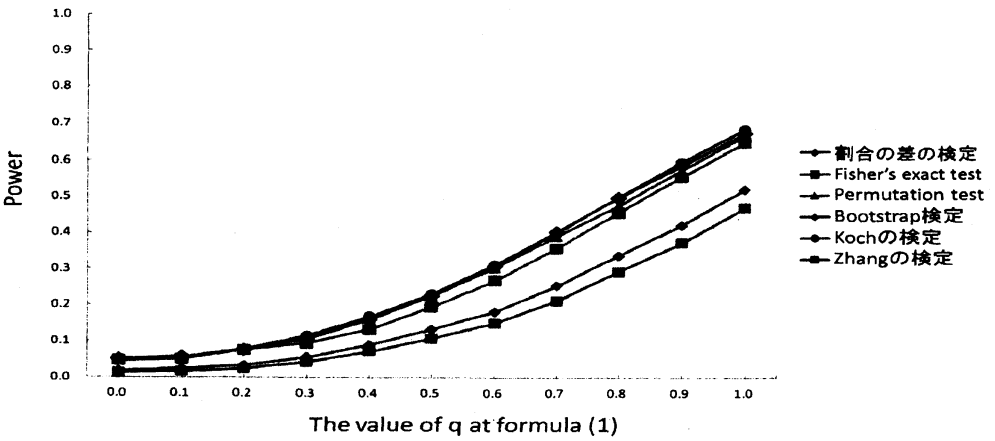

Fig.7 Simulation results power for 3.2.2.2 (α=0.05 5000runs N=192 K=4)

3.2.2.3. 最小化法における共変量の均衡を図る確率 P=2/3

共変量を誤特定し、最小化法における共変量の均衡を保持する確率を P=2/3 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0238、Fisher's exact test で 0.0190、並べ替え検定で 0.0484、Bootstrap 検定で 0.0452、Koch の検定で 0.0540、Zhang の検定で 0.0566 であった。

検出力は Koch の検定、Zhang の検定、並べ替え検定、Bootstrap 検定、割合の差の検定、Fisher's exact test の順で有った。

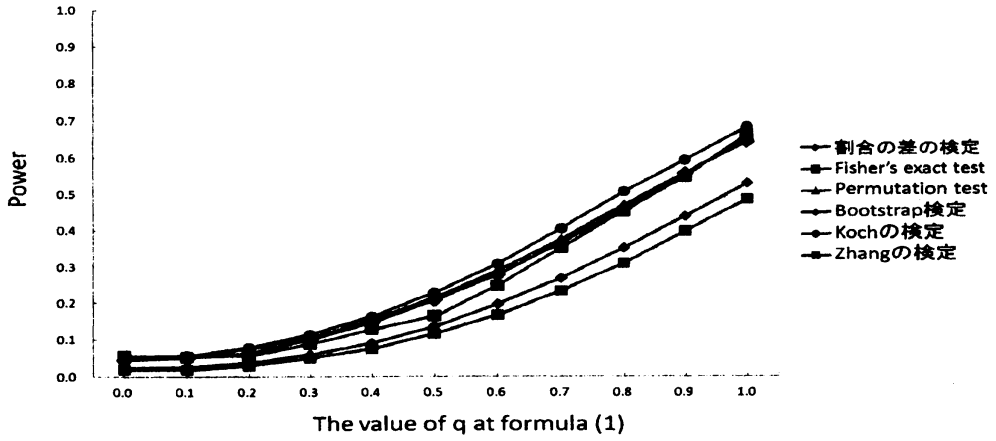

Fig.8 Simulation results power for 3.2.2.3 (α=0.05 5000runs N=192 K=4)

3.2.3.共変量に交互作用項が存在した場合

3.2.3.1. 最小化法における共変量の均衡を図る確率 P=1

共変量に交互作用項が存在し、最小化法における共変量の均衡を保持する確率を P=1 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0148、Fisher's exact test で 0.0096、並べ替え検定で 0.0448、Bootstrap 検定で 0.0462、Koch の検定で 0.0478、Zhang の検定で 0.0560 であった。

検出力は Zhang の検定、Koch の検定、Bootstrap 検定、並べ替え検定、割合の差の検定、Fisher's exact test の順で有った。

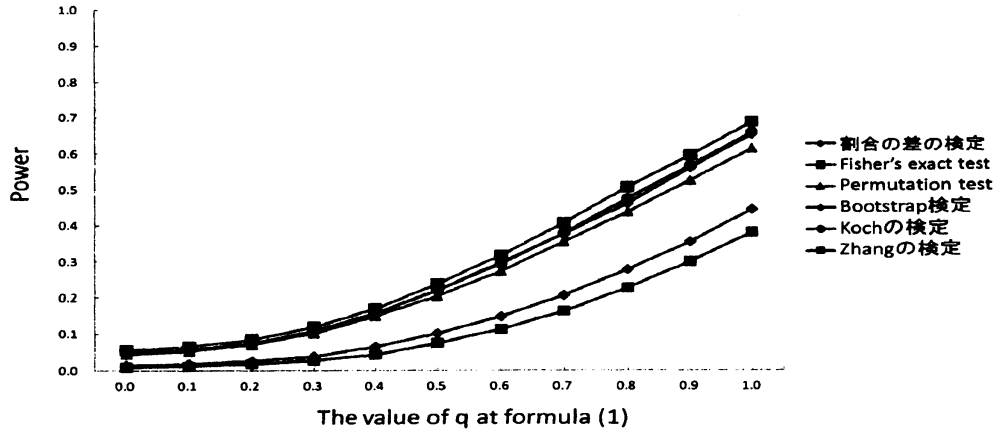

Fig.9 Simulation results power for 3.2.3.1 (α=0.05 5000runs N=192 K=4)

3.2.3.2. 最小化法における共変量の均衡を図る確率 P=4/5

共変量に交互作用項が存在し、最小化法における共変量の均衡を保持する確率を P=4/5 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0158、Fisher's exact test で 0.0120、並べ替え検定で 0.0532、Bootstrap 検定で 0.0518、Koch の検定で 0.0520、Zhang の検定で 0.0567 であった。

検出力は Zhang の検定、Koch の検定、Bootstrap 検定、並べ替え検定、割合の差の検定、Fisher's exact test の順で有った。

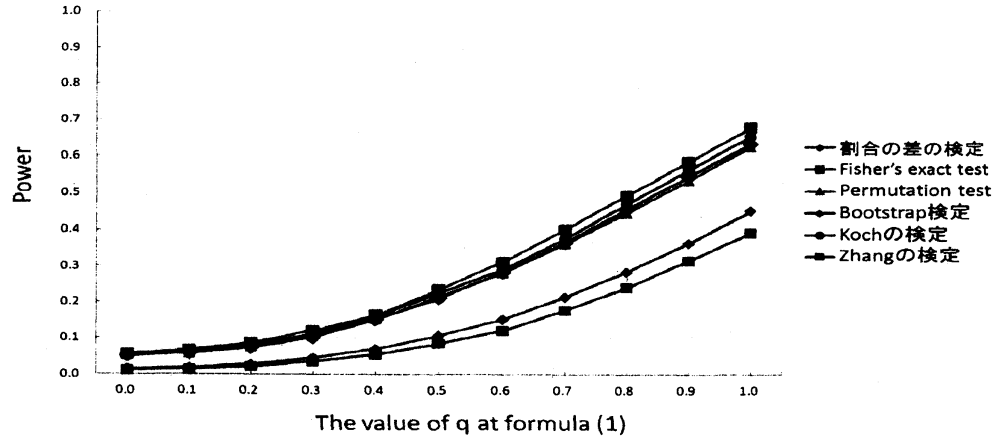

Fig.10 Simulation results power for 3.2.3.2 (α=0.05 5000runs N=192 K=4)

3.2.3.3. 最小化法における共変量の均衡を図る確率 P=2/3

共変量に交互作用項が存在し、最小化法における共変量の均衡を保持する確率を P=2/3 に設定し、シミュレーションを実施した。

その結果、Type I error は割合の差の検定で 0.0166、Fisher's exact test で 0.0128、並べ替え検定で 0.0464、Bootstrap 検定で 0.0388、Koch の検定で 0.0450、Zhang の検定で 0.0494 であった。

検出力は Zhang の検定、Koch の検定、並べ替え検定、Bootstrap 検定、割合の差の検定、Fisher's exact test の順で有った。

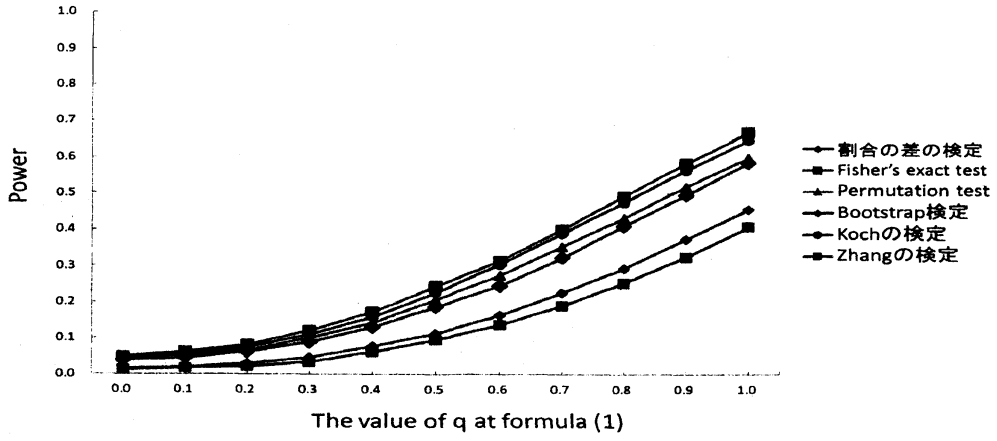

Fig.11 Simulation results power for 3.2.3.3 (α=0.05 5000runs N=192 K=4)

4.考察

本研究では、CA の一つである最小化法を用いて被験者が割り付けられた時の周辺治療効果の推定方法を評価した。検討の結果として、併合分散を用いた割合の差の検定が保守的になる理論的な根拠を与えることができた。加えて、並べ替え検定および Bootstrap 検定は結果変数が二値変数の場合でも機能することを確認した。

凸関数の性質より、並べ替え検定や Bootstrap 検定で用いられる分散は割合の差の検定で用いられる併合分散よりも小さな値であることが明らかになり、その結果、検出力は割合の差の検定よりも並べ替え検定や Bootstrap 検定の方が高くなること、Type I error は割合の差の検定で名義的な有意水準より大きく下回ることが明らかになった。

シミュレーションによる検討において、並べ替え検定と Bootstrap 検定の検出力は割合の差の検定よりも高く、本研究で証明した通りの結果を得た。並べ替え検定と Bootstrap 検定の検出力を比較すると、二つの方法は同程度であった。また、共変量を考慮した検定である Koch の検定および Zhang の検定と比較すると、検出力はほぼ同程度であった。しかし、Situation 1 および Situation 3 で最小化法における共変量の均衡を保つ確率が P=2/3 のとき、検出力は Koch の検定および Zhang の検定の方が並べ替え検定および Bootstrap 検定よりも高かった。これは、最小化法が完全無作為割り付けに近い性質を示す事が原因だと考えられる。共変量の均衡を図る確率を P=1/2 に設定した時の最小化法は完全無作為割り付けと同等な性質を示す。そのため、P=2/3 のときに、共変量と治療群の独立性が強まり、Koch の検定および Zhang の検定に対する CA の影響は弱まる。結果として、Koch の検定および Zhang の検定の検出力が向上すると考えられる。また、Situation 2 である共変量を誤特定した場合には、Koch の検定および Zhang の検定に用いる分散が最小でない。そのため、検出力が低下し、CA の影響が弱まることからくる検出力の向上は相殺された。その結果として並べ替え検定、Bootstrap 検定、Koch の検定、Zhang の検定それぞれの検出力に差が認められなかったと考えられる。また、Koch の検定および Zhang の検定に、観測されたデータに最も良く当てはまるモデルを用いることにより分散を最小にし、検出力の低下は回避することができると考えられる。

最小化法における共変量の均衡を保つ確率が並べ替え検定におよぼす影響は少なかった。一方、Bootstrap 検定では P=4/5 を基準にしてそれぞれの検出力を比較すると、Situation 1 において、必ず共変量の均衡を図る P=1 では検出力が約 6%程度向上し、CA の影響が弱まり完全無作為割り付けに近づく P=2/3 では検出力が約 8%程度低下する。同様に Situation 2 において P=1 のときには検出力が約 2%程度向上し、P=2/3 のときには約 5%程度低下する。また、Situation 3 において P=1 のときには検出力が約 3%程度向上し、P=2/3 のときには約 8%程度低下していた。このために、ノンパラメトリック検定である並べ替え検定を用いることにより、共変量の均衡を図る確率に依らない、検出力が安定した検定が実施できるだろう。

シミュレーションにおける割合の差の検定の Type I error は、本研究で示した通り、名義的な有意水準を大きく下回っていた。それに対して、並べ替え検定および Bootstrap 検定は名義的な有意水準付近の値を保持していた。しかし、Bootstrap 検定においては、Situation 3 の交互作用項が存在する場合において、Type I error が 0.0388 と名義的な有意水準を下回る状況が存在した。また結

果は示していないが予備検討において、結果変数が連続値を想定した場合のBootstrap検定のType I errorは、名義的な有意水準を超過していたという結果を得ている。また、Kochの検定およびZhangの検定は名義的な有意水準を超過していることが多かった。

以上の事から、現在実施されている割合を評価項目としたCAを用いた臨床試験の多くに並べ替え検定およびBootstrap検定を用いることで、Type I errorを名義的な有意水準付近の値を保持し、割合の差の検定やFisher's exact testよりも検出力を改善することができると考えられる。しかし、Bootstrap検定よりも並べ替え検定の方がType I errorが安定しているために、CAを用いた場合の検定には並べ替え検定を用いることが推奨される[10]-[12]。

今後の課題としては実データへの適用、並べ替え検定を生存時間解析に適用した場合にどのような性質を明らかにしていくことがあげられる。

5.参考文献

[1] Zhang,M. Tsiatis,AA. Davidian,M. (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics.* 64(3). 707-715.

[2] Koch,GG. Tangen,CM. Jung,J-W. Amara,AA. (1998). ISSUES FOR COVARIANCE ANALYSIS OF DICHOTOMOUS AND ORDERED CATEGORICAL DATA FROM RANDOMIZED CLINICAL TRIALS AND NON-PARAMETRIC STRATEGIES FOR ADDRESSING THEM. *Statist Med.* 17. 1863-1892.

[3] Gail,MH. Wieland,S. Piantadosi,S. (1984). Biased estimates of treatment effect in randomized experiments with non-linear regressions and omitted covariates. *Biometrika.* 71:431-444

[4] Kim,H-M, Yasui,Y. Burstyn,I. (2006). Attenuation in Risk Estimate in Logistic and Cox Proportional-Hazards Model due to Group-Based Exposure Assesment Strategy. *Am.Occup.Hyg.* Vol.50. No.6 pp623-635.

[5] Robinson,LD. Jewell,NP. (1991). Some Surprising Results about Covariate Adjustment in Logistics Regression Models. *Int Stat Rev.* 58. 227-240.

[6] Hagino,A., Hamada,C., Yoshimura,I., Sakamoto,J. and Nakazato,H. (2004). Statistical comparison of random allocation methods in cancer clinical trials. *Contr Clin.Trials* 25:572-584

[7] Rosenberger,WF. Sverdlov,O. (2008). Handling Covariates in the Design of Clinical Trials. *Statistical Science.* 23(3). 404-419.

[8] Hasegawa,T. and Tango,T. (2009). Permutation test following cavariate-adaptive randomization in randomized controlled trials. *Journal of Biopharmaceutical Statistics*. 19:106-119

[9] Shao, J. and Yu,X. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*. 97:347-360.

[10] Buyse,M. (2000). Centralized treatment allocation in comparative clinical trials. *Applied Clinical Trials* 9, 32-37

[11] Kalish,LA. Begg,CB. (1987). The Impact of Treatment Allocation Procedures on Nominal Significance Levels and Bias. *Controlled Clinical Trials*. 8. 121-135.

[12] Proschan,M. Brittain,E. Kammerman,L. (2011). Minimize the Use of Minimization with Unequal Allocation. *Biometrics*. 67. 1135-1141.