

特集：臨床試験・治験の最近の動向**<総説>****統計モデル：母集団モデルと無作為化モデル**

長谷川貴大

塩野義製薬株式会社

Statistical models: Population model and randomization model

Takahiro HASEGAWA

SHIONOGI & CO., LTD.

抄録

新しい薬剤や治療方法の有効性および安全性を評価するため、臨床試験が計画され実施される。その中で無作為化の実施は、科学的なエビデンスを構築する上で特に重要な要素の1つである。つまり、適格条件を満たし、かつ同意を取得できた被験者が随時、試験へ登録され各群へ無作為に割り付けられる。

無作為化を行う目的は、まず治療群間の比較可能性を担保することである。また、観測された治療群間の差が偶然変動を超えたものであるか否かにより、観測された治療効果が統計的に意味のある大きさであるかが判定される。この偶然変動を確率的なものとするのも無作為化実施のもう1つの目的である。形状が既知の分布から無作為抽出されたものが観測された結果であると考え、未知パラメータを推定する母集団モデルに基づく通常の解析と比べて、無作為化による推測はかなり異なるものである。なぜなら、臨床試験へ登録される被験者は、各治療群の母集団から無作為に選ばれたとは決して考えられないからである。しかし、実際にはこの仮定があたかも成り立っているかのようにして、解析が行われているのが実状である。したがって、無作為化比較試験は各群で無作為抽出が行われておらず、無作為化モデルに基づく並べ替え検定で解析を実施することが自然と考えられる。無作為化モデルでは、母集団モデルとは異なり、帰無仮説のもとで割り付け結果を変量、観測された応答を固定と仮定する。

少なくとも完全無作為化法または無作為割り付け規則による割り付けが実施された大規模な試験においては、これら異なるモデルの統計的推測の性質は同等であることが知られている。しかし、重要な予後因子が存在する試験では、層別無作為化法、または Pocock-Simon 法や Taves の最小化法などにより割り付けが実施される。このように、予後因子について積極的にバランスをとる割り付けが行われた場合には、統計的推測の性質が同等になるとは必ずしも言えない。したがって、無作為化に基づき、仮定を必要としない並べ替え検定を主たる解析として実施することが、特に重要な予後因子が存在する検証的な無作為化比較試験では必要と考えられる。

キーワード：無作為化比較試験，無作為化モデル，母集団モデル，疑似母集団モデル，並べ替え検定

Abstract

A clinical trial is designed to evaluate the beneficial and adverse effects of a new medical treatment or intervention. The use of randomized controlled trials (RCTs), in particular, is of utmost importance to provide reliable evidence. In RCTs, eligible and consenting subjects sequentially enter a study and are randomly assigned to one of two or more study treatments.

Randomization promotes comparability among the study groups. In addition, it provides a probabilistic basis to derive an inference from observed results with respect to all possible results. This randomization approach to inference is different from the usual testing of unknown parameters that arise from an independent and identically distributed sample of a known

連絡先：長谷川貴大

〒 553-0002 大阪府大阪市福島区鷺洲 5-12-4

12-4, Sagisu 5-chome, Fukushima-ku, Osaka 553-0002, Japan

Tel: 06-6458-5861 Fax: 06-6455-2037

E-mail: takahiro.hasegawa@shionogi.co.jp

[平成 23 年 1 月 17 日受理]

distribution based on a population model. This difference explains why subjects are never randomly sampled from a large population of patients receiving treatments under a study. Furthermore, the population model can then be invoked as the basis for data analysis. Because RCTs lack a formal sampling basis, it is logical to apply a permutation test under the randomization model, where it is assumed that the treatment assignment is random and that the set of observed responses is fixed under the null hypothesis.

The statistical inference derived from either of the 2 models is the same in very large trials that use complete randomization or the random allocation rule. However, stratified randomization or covariate-adaptive randomizations, such as the Pocock-Simon procedure and the so-called deterministic minimization method by Taves that are used in many trials with important prognostic factors, do not necessarily have this property. Therefore, a permutation test that is an assumption-free randomization-based inference could be used as the primary statistical analysis, particularly in a confirmatory RCT with important prognostic factors.

Keywords: randomized controlled trial, randomization model, population model, invoked population model, permutation test

I. はじめに

臨床試験は、患者の病気を治療する新たな薬剤や方法を評価するために実施される。したがって、新たな薬剤や治療方法の効果を科学的に評価できる臨床試験の方法を選択する必要がある。最も質の高い科学的なエビデンスを提供できる唯一の試験デザインと言われているのが無作為化比較試験である。無作為化比較試験では、試験へ参加した被験者を各試験治療群へ無作為に割り付ける。そして、試験治療間の差により治療効果の大きさを評価する。つまり、絶対的ではなく、比較している別の治療法に対する相対的な測度を用いている。このため、被験者の反応性に関する治療群間の類似性が求められ、無作為化の実施が最も重要な要素の1つとなる。予測不可能性の担保、及び選択バイアスの除去により、各群の被験者背景を期待的に均一とし、各治療群を対等に比較できる。つまり、比較可能性が保証される。ただし、注意が必要なのは、被験者背景が均一になるとするのは期待的なだけであり、必ずしも全ての背景因子が揃うことは保証されない。特に、被験者数が少ない試験では、観測結果に影響を与える予後因子の分布が群間で揃わず、推定される治療効果の大きさに無視できない偏りをもたらす確率が高くなる。そこで、試験計画時に影響の大きい予後因子の存在が予め分かっている場合には、これらの予後因子について積極的に群間で分布が揃うように無作為化を行う必要がある。予後因子に対して、割り付けで積極的にバランスをとらない無作為化法として、完全無作為化法、無作為割り付け規則 [1]、置換ブロック法などがある。そして、積極的にバランスをとる方法として、層別無作為化法、Pocock-Simon 法 [2]、Taves の最小化法 [3] などがある。

また、観測された治療群間の差が偶然変動を超えたものであるか否かにより、無作為化比較試験で観測された治療効果が統計的に意味のある大きさであるかが判定される。この偶然変動を確率的なものとするのも無作為化を行うもう1つの目的である。ここで発生する確率的な偶然変動は、用いられた無作為化の方法によって異なるものと考えられる。したがって、解析方法を無作為化の方法に対応さ

せて選択する必要があると考えるのは自然であろう。表1に、予後因子のバランスをとる割り付け方法を用いた無作為化比較試験の例を示し、それぞれで用いられた割り付け方法と解析方法を表した。層別無作為化法による割り付けが実施された3試験において、実施された解析方法が割り付け因子による調整なしの場合と共分散分析 (ANCOVA) による調整ありの場合が混在している。更に、最小化法が用いられた3試験はいずれも共分散分析が適用されているものの、そのときに用いた調整因子 (共変量) は、最小化法で積極的にバランスをとった割り付け因子の場合とそれとは別因子の場合が混在している。これらの不一致は、無作為化によって発生される偶然変動とそれに対応する解析方法を選ぶ基準が明確化されていないことを表している。この背景には、無作為化比較試験でよく用いられる t 検定や共分散分析などの多くの統計的推論は、母集団モデル [10] のもとで成り立っていることが考えられる。このモデルのもとでは、標本が母集団から無作為に抽出されたと仮定する。したがって、割り付けの方法とは無関係な統計手法となっている。つまり、実施された割り付け方法を無視して、各群で無作為抽出が行われたと仮定した解析方法

表1 予後因子のバランスをとる割り付け方法を用いた無作為化比較試験

目的	割付因子	割付方法	解析方法	
			調整	調整因子
Brudi et al.[4] 低比重リボタンパクコレステロールの低下	スタチンの使用量	層別無作為化法	なし	---
TARGET試験[5] 潰瘍合併症発現リスクの低下	アスピリンの使用、年齢	層別無作為化法	あり (ANCOVA)	割付因子
IFIGENIA試験[6] 特発性肺線維症の治療	国、肺活量	層別無作為化法	あり (ANCOVA)	割付因子
HORIZON試験[7] グルココルチコイド誘発性骨粗しょう症の予防と治療	グルココルチコイド治療の期間	最小化法	あり (ANCOVA)	性別、地域
RECORD試験[8] 骨折の2次予防	年齢、性別、骨折から時間、骨折の種類	最小化法	あり (ANCOVA)	割付因子
SMaRT試験[9] 癌患者の鬱病治療	性別、年齢、癌の種類、病気の程度	最小化法	あり (ANCOVA)	割付因子

が選択され、適用されていると考えられる。しかし、無作為化比較試験では、非無作為に選ばれた施設において、選択基準を満たし、かつインフォームド・コンセントの得られた被験者が登録される。そして、被験者は無作為に各群へ割り付けられる。この状況は、被験者それぞれが各群の母集団から無作為抽出されていることを決して表していない。そこで、割り付け方法によって異なる偶然変動を考慮した無作為化モデル [10] に基づく推論を行うことが自然と考えられる。無作為化モデルでは、母集団モデルとは異なり、帰無仮説のもとで割り付け結果を変数、観測された応答を固定と仮定する。

そこで、本論文では、多くの無作為化比較試験で用いられている解析方法で仮定されている母集団モデルについて II 節で示す。そして、III 節では無作為化比較試験で適切な仮定と考えられる無作為化モデルについて紹介する。IV 節においては、これらのモデルの比較結果をまとめる。

II. 母集団モデル

t 検定や共分散分析などの多くの統計的推論は母集団モデル (population model) のもとで成り立っている。このモデルでは、標本が母集団を代表し、それぞれ独立で同じ確率分布に従う確率変数であることを仮定している。ここでは、母集団モデルのもとでの 2 群比較を考えてみる。群 A と群 B それぞれで仮定される母集団から n_A 個と n_B 個の標本が無作為に抽出されたとする。各群で n_g ($g=A,B$) 個の応答 (Y_{g1}, \dots, Y_{gn}) が観測され、これらは各群の母集団分布を特徴づけるパラメータ θ_g をもつ $G(y|\theta_g)$ に従うとする。この関係は図 1 に示すとおりである。

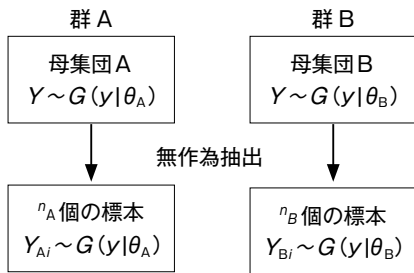


図 1 母集団モデル

このもとで群間比較を行う場合の仮説は、それぞれパラメータ θ_g に対して、

$$H_0: \theta_A = \theta_B, \quad H_1: \theta_A \neq \theta_B$$

と設定される。そして、尤度を用いて多くの統計的推測の手法が導出されている。例えば、 t 検定では、分布 G がパラメータ $\theta_g = (\mu_g, \sigma^2)$ をもつ正規分布とし、各群の平均について

$$H_0: \mu_A = \mu_B, \quad H_1: \mu_A \neq \mu_B$$

の仮説が検定される。

しかしながら、この母集団モデルに基づく統計的推測を無作為化比較試験へ適用する際には注意が必要である。なぜなら、無作為化比較試験では、被験者が各群の母集団か

ら無作為抽出されてはいないからである。すなわち、試験へ登録される被験者は、専門的知識や被験者登録の見込みなどの観点から非無作為に選択された施設から適格条件を満たし同意を得られた患者である。そして、無作為にどちらかの群へ割り付けられる。このように各群の被験者は無作為に抽出されていないにもかかわらず、通常、操作上の仮定としてこの母集団モデルを導入し、あたかもこの仮定が成り立っているかのようにして割り付け方法とは無関係な統計手法で解析が行われている。Lachin [10] は図 2 を示し、このモデルを疑似母集団モデル (invoked population model) と呼んだ。このモデルのもとでは、応答を確率変数、割り付け結果を固定と考える。

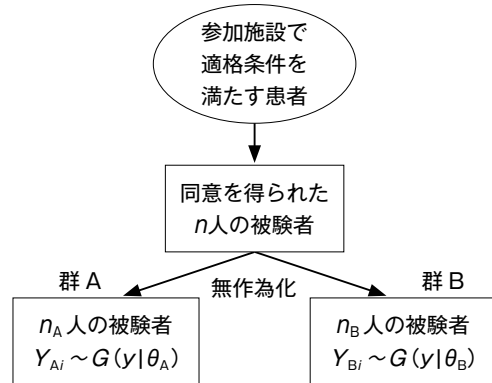


図 2 疑似母集団モデル

III. 無作為化モデル

無作為化比較試験においては、無作為抽出の仮定が成り立たない。そこで、割り付け方法を考慮した無作為化モデル (randomization model) に基づく推論を行うことを考える。無作為化が意味するところは、無作為化の方法によって生じる割り付けのデータ列 ABAAB... を確率変数と考える点である。用いられた無作為化の方法によってこの性質が異なり、これを考慮して統計的推測が行われるモデルを Lachin [10] は無作為化モデルと呼んだ。図 3 に示したこのモデルのもとでは、割り付け結果を確率変数、応答を固定と考える。すなわち、これは Fisher [11] が提案した並

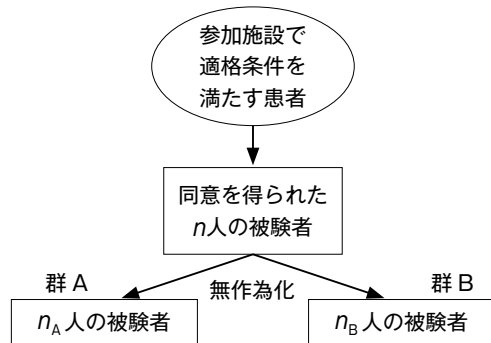


図 3 無作為化モデル

べ替え検定に対応する。用いられた無作為化の方法によって、並べ替えられる割り付けのデータ列の分布が変わり、そのもとで統計的推測が行われることになる。割り付けのデータ列 $w (w=1, \dots, \Omega)$ に対する検定統計量を S_w とする。そして、実際に観測された割り付けのデータ列 w_{obs} に対する検定統計量を s_{obs} とおく。このとき、並べ替え検定の両側 p 値 [12] は以下のように与えられる。

$$p = \sum_{w=1}^{\Omega} I(|S_w - \bar{S}| \geq |s_{obs} - \bar{S}|) \Pr(W=w)$$

ここに、 \bar{S} は検定統計量の平均値 $\sum_{w=1}^{\Omega} S_w \Pr(W=w)$ を表し、 $I(\cdot)$ はカッコ内の条件を満たす場合は 1、満たさない場合は 0 を与える指標関数(indicator function)である。ここで、無作為化モデルに基づく並べ替え検定の例を示す。4 例の被験者が無作為化比較試験へ参加し、群 A または群 B へ無作為に割り付けられるものとする。そして、各被験者の割り付け結果が(A,B,B,A)、応答が(2,1,3,4)と観測された場合を考える。群 A へ割り付けられた被験者応答の平均値からの差の合計を検定統計量 S とすると、観測された統計量は $1.0 (= (2-2.5) + (4-2.5))$ となる。割り付けが完全無作為化法または無作為割り付け規則で実施された場合、考えられる割り付けパターンは 6 通りある。また、それぞれの発生確率は $1/6$ である。各割り付けパターンに対応する統計量は表 2 に示すとおりである。したがって、検定統計量の絶対値が 1.0 以上の割り付けパターンは 4 通りあることから、これらが観測される確率すなわち両側 p 値は $4/6=0.67$ と得られる。これが、完全無作為化法または無作為割り付け規則が用いられた無作為化比較試験における無作為化モデルのもとでの検定結果である。もし、ブロックサイズが 2 の置換ブロック法により割り付けられていた場合には、観測される可能性のある割り付けパターンは 4 通りのみである。そして、それぞれの発生確率は $1/4$ である。したがって、このときの両側 p 値は $2/4=0.50$ となる。このように、無作為化モデルでは、用いられた割り付け方法に対応する割り付けパターンの種類と発生確率により、算出される p 値は異なる。すなわち、割り付け方法に対応した解析を行っていることになり、実際に実施した無作為化を考慮できている。

表 2 無作為化モデルに基づく並べ替え検定の例

割りパターン	検定統計量 S	発生確率	
		完全無作為化法 無作為割り付け規則	置換ブロック法 (ブロックサイズ 2)
AABB	<u>-2.0</u>	1/6	0
ABAB	0.0	1/6	1/4
ABBA	<u>1.0</u>	1/6	1/4
BAAB	<u>-1.0</u>	1/6	1/4
BABA	0.0	1/6	1/4
BBAA	<u>2.0</u>	1/6	0

さて、被験者が登録されるたびに、被験者の特性に応じて逐次的に割り付けが実施される Pocock-Simon 法や Taves の最小化法などの割り付け方法では、割り付けのデータ列

w の発生確率 $\Pr(W=w)$ を求めることは難しく、並べ替え検定の両側 p 値を正確に計算することは実行上、現実的でない。そこで、モンテカルロ法により p 値を推定することになる。観測された被験者の応答、特性、登録順序は固定し、適用された割り付け方法で割り付けを M 回繰り返す。 m 回目の割り付け結果から求められる検定統計量を $S_m (m=1, \dots, M)$ とおくと、モンテカルロ両側 p 値 [12] は、

$$p = \frac{1 + \sum_{m=1}^M I(|S_m - \bar{S}| \geq |s_{obs} - \bar{S}|)}{M+1}, \bar{S} = \frac{\sum_{m=1}^M S_m}{M}$$

と計算することができる。ここでモンテカルロ両側 p 値の算出例を示す。2 群(A, B)を比較する無作為化比較試験を考え、10 例の被験者が試験へ登録される。そして、性別を割り付け因子として Pocock-Simon 法で割り付けが実施されるものとする。このとき、バランスを確保するための割り付け確率を 0.8 とする。各被験者の登録順、性別、割り付け結果、応答を表 3 に示した。検定統計量として、B 群での応答の平均値から A 群の平均値を引いた値と定義し、0.7 が観測された。モンテカルロ法による繰り返しを 9 回とし、それぞれの割り付け結果と対応する検定統計量を表 3 に示した。このうち、観測された検定統計量と検定統計量の平均値との差よりも大きかった割り付け結果は 1 つのみであったことから、モンテカルロ両側 p 値は 0.2 と算出された。なお、ここでは簡単のため、モンテカルロ法による繰り返しを 9 回としたが、実際の計算ではこの繰り返し数を多くすることで、より精度の高い p 値を得ることができる。

表 3 無作為化モデルにおけるモンテカルロ p 値の計算例

登録 順序	性別	応答	割付 結果	モンテカルロ法による割付結果									
				1	2	3	4	5	6	7	8	9	
1	女	4.5	B	B	A	B	B	B	B	B	B	B	A
2	男	3.5	B	A	B	B	A	A	A	B	A	B	B
3	男	2.9	A	B	A	B	B	B	B	A	B	A	A
4	男	2.5	A	A	A	A	B	A	B	A	B	B	B
5	女	4.2	B	B	B	A	B	A	A	A	A	A	A
6	男	1.4	A	B	B	A	A	B	A	A	B	A	A
7	男	2.6	B	A	A	B	B	A	A	B	B	B	B
8	女	4.2	A	A	B	A	A	A	A	A	B	B	B
9	女	3.8	A	A	A	B	A	B	B	B	A	B	B
10	男	3.5	B	A	B	B	A	A	A	B	B	A	A
検定統計量				<u>0.7</u>	-0.1	0.1	0.4	0.1	-0.3	0.2	0.5	<u>-0.7</u>	0.0

IV. 無作為化モデルと疑似母集団モデルの比較

ここでは、2 群からなる無作為化比較試験において、無作為化モデルと割り付け方法に依存しない疑似母集団モデルによる統計的推測の関係を比較する。まず、予後因子のない無作為化比較試験を考える。Lachin [1] は、完全無作為化法または無作為割り付け規則で割り付けを行い、線形順位検定を用いて群間比較を行う場合には、無作為化モデルと疑似母集団モデルの統計的推測は、漸近的に一致する

ことを示した。その論拠は以下のとおりである。線形順位検定では、 n 例の応答を Y_1, \dots, Y_n とし、これをあるスコアへ変換する。スコアとしてそのままの応答を用いれば t 検定、順位を与えれば Wilcoxon 順位和検定、Savage スコアであれば logrank 検定に対応し、線形順位検定はこれらを一般化したものである。被験者 i のスコアを A_i ($i=1, \dots, n$)、スコアの平均値を \bar{A} とおく。また、被験者 i の割り付け結果を T_i とする。群 A へ割り付けられた場合には $T_i=1$ 、群 B の場合には $T_i=0$ である。このとき、線形順位検定の統計量 S は、

$$\bar{S} = \sum_{i=1}^n (A_i - \bar{A}) T_i$$

で与えられる。まず、疑似母集団モデルのもとでは、応答を確率変数、割り付けを固定と考えることから、統計量 $S_{pop.}$ の分散は、

$$\text{Var}(S_{pop.}) = \frac{n_A n_B}{n} \text{Var}(A_i)$$

と算出される。ここに、群 A および群 B それぞれに割り付けられた例数を n_A, n_B とする。一方、無作為化モデルのもとでは、割り付けを確率変数、応答を固定と考える。完全無作為化法と無作為割り付け規則のような割り付けでは、被験者 i または j によらず、割り付けの分散 $\text{Var}(T_i)$ と共分散 $\text{Cov}(T_i, T_j)$ は一定であることから、観測された群 A の例数で条件づけた統計量 $S_{ran.}$ の分散は、

$$\text{Var}(S_{ran.} | N_A(n) = n_A) = \frac{n_A n_B}{n} \left(\frac{\sum_{i=1}^n (a_i - \bar{a})^2}{n-1} \right)$$

であることが知られている。これは、疑似母集団モデルのもとでの分散の一致推定量であることがわかる。したがって、大規模な無作為化比較試験では、2つのモデルの統計量の分散は等しくなる。また、帰無仮説のもとで、検定統計量の期待値は両モデルにおいてゼロで等しいことから、完全無作為化法または無作為割り付け規則を用いた無作為化比較試験では、疑似母集団モデルと無作為化モデルの統計的推測は同等と考えられる。

また、割り付けで置換ブロック法が用いられた場合、Matts and Lachin [13] は無作為化モデルよりも疑似母集団モデルのもとでの統計量の分散が小さくなることを示した。特に、ブロックサイズが小さくなるにつれて、この差は大きくなる。したがって、置換ブロック法を用いた無作為化比較試験において、疑似母集団モデルのもとでの統計的推測は、分散を過小評価し、第1種の過誤確率が名目有意水準以下に保たれないため、注意が必要である。なお、ブロックサイズが4のときで、疑似母集団モデルの統計量の分散は無作為化モデルの0.75倍程度小さくなる。

次に、重要な予後因子が存在する無作為化比較試験を考える。この場合、重要な予後因子について積極的にバランスを確保するため、層別無作為化法または Pocock-Simon 法などによって割り付けが実施される。そして、疑似母集団モデルのもとでは、2つの考え方により解析方法が選択される。1つは、無作為化により偏りのない比較ができると考え、予後因子について調整を行わない t 検定や

logrank 検定を実施する場合である。もう1つは、予後因子の影響を調整する解析を選択する場合である。例えば、解析方法として共分散分析や層別 logrank 検定などが挙げられる。しかし、Hasegawa and Tango [14], Kalish and Begg [15], Hagino et al. [16] はそれぞれ応答が連続量、2値、生存時間の場合について、調整なしの解析を実施した場合には第1種の過誤確率が有意水準を下回り、保守的な解析を導くことを示している。これは、治療効果がある場合には検出力の低下に繋がる。反対に、割り付けで考慮した因子で調整する解析を実施した場合には、第1種の過誤確率は有意水準付近に保たれることが示されている。したがって、疑似母集団モデルのもとでは、割り付けで積極的にバランスを確保した予後因子について調整した解析を選択することが望まれる。また、調整を伴う疑似母集団モデルと無作為化モデルの統計的推測の比較については、Hasegawa and Tango [14] が報告している。そこでは、応答として連続量を考え、共分散分析と無作為化モデルのもとでの並べ替え検定の統計的推測が同等であることを理論的に示している。更には、応答に影響を与える重要な予後因子間に交互作用効果がない場合には、割り付け方法として層別無作為化法よりは予後因子の周辺分布を揃える Pocock-Simon 法を用い、解析方法として並べ替え検定を実施する手順を提案している。これにより、共分散分析と同等の統計的推測を行え、かつ無作為化比較試験において自然な仮定での解析を実施できることになる。

V. 考察

無作為化比較試験において、被験者は各群へ無作為に割り付けられている。それにも関わらず、被験者が各群の母集団から無作為に抽出されたかのようにあたかも仮定し、疑似母集団モデルのもとで解析がなされているのが実状である。特に、予後因子のバランスをとる割り付け方法を用いた無作為化比較試験では、用いられた割り付け方法との間の対応関係が不明確なまま、何種かの解析方法が適用されてしまっている。統計的推測において、第1種の過誤確率が有意水準付近に保たれていればよいというものではない。欧州医薬品審査庁 (EMA; 現欧州医薬品庁) は2003年に共変量調整に関する考え [17] を示した。その中で、最小化法の利用に対して強く反対している。逐次的に割り付けが行われる Taves の最小化法は統計的推測を行う上での正当な確率の変動を与えないという点が最も大きな理由と考えられる。疑似母集団モデルを仮定した解析を行う限りにおいて、これは正しい。しかし、無作為抽出が行われていない状況では、最小化法の代わりに層別無作為化法を用いたとしても、疑似母集団モデルのもとでは統計的推測を行う上での基礎が与えられず、状況は変わらない。この点で、統計的推測を行う上での正当な確率の変動を与えないという理由から最小化法の利用が否定される根拠にはならない。実際には、疑似母集団モデルを仮定した解析に対して懐疑しなければならぬと考えられる。そこで、この

議論に対する回答として、無作為化モデルに基づく解析の実施が提案される。統計的推測を行う上での確率の変動が無作為化によって与えられることから、上記の懸念を払拭できる。特に、癌領域などの小規模で重要な予後因子が複数存在するような臨床試験で、最小化法の利用を正当化できる。また、同時に割り付け方法と解析方法の不一致についても解決できることになる。

ところで、無作為化モデルで用いられる並べ替え検定の結果は、解析された集団に対してのみ結果の反映ができる。このため、結果の一般化が問題となる。実際の試験で登録される被験者は無作為抽出されていないことから、たとえ疑似母集団モデルのもとで解析が実施されたとしても、結果の一般化可能性には疑問が残る。このため、どちらのモデルを用いたとしても、この問題は解決できない。もし、一般化可能性が要求される場合には、同様の臨床試験を複数回繰り返す、メタアナリシスによってその結果を併合し、解釈していく必要があるだろう。

さて、並べ替え検定を適用する際、実行上はモンテカルロ法によりp値を推定する必要がある。この場合、同一データに対して検定の実施を繰り返すたびに、異なるp値が得られてしまうという問題がある。検証試験における主たる解析は事前に規定しておく必要があることから、統計解析計画書には、モンテカルロ法による繰り返し回数および乱数発生時に与えるシード値を記載しておくことが現実的な対応と考えられる。また、治療効果の有意性を判定する検定結果だけでなく、治療効果の大きさを推定する信頼区間が必要な場合には、Garthwaite [18] が提案している方法を用いることで手に入れられる。

このように、無作為化比較試験では、無作為化モデルのもとで解析を実施することで統計的推測の妥当な土台が与えられる。したがって、無作為化比較試験、特に重要な予後因子が存在する検証的な試験では、無作為化に基づき、仮定を必要としない並べ替え検定を主たる解析として実施することが必要と考えられる。今後、この無作為化モデルの研究、および実際の適用に向けた議論が盛んになることが望まれる。

参考文献

- [1] Lachin JM. Properties of simple randomization in clinical trials. *Controlled Clinical Trials* 1988;9:312-26.
- [2] Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975;31:103-15.
- [3] Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics* 1974;15:443-53.
- [4] Brudi P, et al. Efficacy of ezetimibe/simvastatin 10/40 mg compared to doubling the dose of low-, medium- and high-potency statin monotherapy in patients with a recent coronary event. *Cardiology* 2009;113(2):89-97.
- [5] Schnitzer TJ, et al. Comparison of lumiracoxib with naproxen and ibuprofen in the Therapeutic Arthritis Research and Gastrointestinal Event Trial (TARGET), reduction in ulcer complications: randomised controlled trial. *Lancet* 2004;364:665-74.
- [6] Demedts M, et al. High-Dose Acetylcysteine in Idiopathic Pulmonary Fibrosis. *The New England Journal of Medicine* 2005;353(24):2229-42.
- [7] Reid DM, et al. Zoledronic acid and risedronate in the prevention and treatment of glucocorticoid-induced osteoporosis (HORIZON): a multicentre, double-blind, double-dummy, randomised controlled trial. *Lancet* 2009;373:1253-63.
- [8] Grant AM, et al. Oral vitamin D3 and calcium for secondary prevention of low-trauma fractures in elderly people (Randomised Evaluation of Calcium Or vitamin D, RECORD): a randomized placebo-controlled trial. *Lancet* 2005;365:1621-28.
- [9] Strong V, et al. Management of depression for people with cancer (SMaRT oncology 1): a randomised trial. *Lancet* 2008;372:40-8.
- [10] Lachin JM. Statistical properties of randomization in clinical trials. *Controlled Clinical Trials* 1988;9:289-311.
- [11] Fisher RA. *The design of experiments*. Edinburgh: Oliver and Boyd; 1971.
- [12] Ernst MD. Permutation methods: a basis for exact inference. *Statistical Science* 2004;19(4):676-85.
- [13] Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Controlled Clinical Trials* 1988;9:345-64.
- [14] Hasegawa T, Tango T. Permutation test following covariate-adaptive randomization in randomized controlled trials. *Journal of Biopharmaceutical Statistics* 2009;19:106-19.
- [15] Kalish LA, Begg CB. The impact of treatment allocation procedures on nominal significance levels and bias. *Controlled Clinical Trials* 1987;8:121-35.
- [16] Hagino A, et al. Statistical comparison of random allocation methods in cancer clinical trials. *Controlled Clinical Trials* 2004;25:572-84.
- [17] The European Agency for the Evaluation of Medicinal Products. Points to consider on adjustment for baseline covariates; CPMP/EWP/2863/99. 2003-3-22. <http://www.emea.europa.eu/pdfs/human/ewp/286399en.pdf>. (accessed 2010-11-21)
- [18] Garthwaite PH. Confidence intervals from randomization tests. *Biometrics* 1996;52:1387-93.