

생물학적 동등성 시험의 실험설계 및 데이터 해석*

김 병 수

연세대학교 상경대학

생물학적 동등성시험의 통계적 특성

Metzler¹⁾는 생체이용율에 대한 정의를 다음과 같이 내리고 있다. "Study of factors which influence and determine the amount of active drug which gets from the administered dose to the site of pharmacologic action as well as the rate at which it gets there" 그리고 생체이용율 연구의 기본적인 가정으로서 생체이용율에 큰 차이가 없는 두 제제는 치료학적 효능에서도 큰 차이가 없다고 간주하고 있다.

대부분의 약은 신장, 체중, 그리고 체격등이 비슷한 사람들에게 같은 양을 투여하였다고 하더라도 혈중농도에는 많은 차이를 보이고 있다. 따라서 이질적인 군에서 혈중농도는 더 큰 변동을 보인다. 혈중농도연구에서 변동의 가장 큰 원인중 하나는 바로 피험자간(between subjects)변동이다. 2×2 교차실험계획은 애초에 생물학적 동등성 시험에서 제제들간의 효과를 비교하는데 피험자간 변동을 실험오차에 포함시키지 않는 방법으로 제안되었다. 3개 이상의 제제를 비교할 때는 확률화 불완비 블록 계획 (randomized incomplete block design)을 사용하고, 2개의 제제를 비교할 때는 2×2 교차실험계획을 사용한다. 그리고 이러한 계획들을 일반적으로 블록계획이라고 하는데, 생물학적 동등성 시험에서는 2개의 제제를 비교하는 경우가 대부분이므로 본 논의에서는 2×2 교차실험계획만 다루기로 하겠다.

2×2 교차실험계획은 일찌기 낙농학에서 그 예를 찾아볼 수 있다.²⁾ 두 종류의 사료 (A,B)가

Table I—두 종류의 사료에 따른 젖소의 우유양

순서	젖소	수유기	
		I	II
AB	I	y ₁₁₁	y ₁₂₁
	.	.	.
	.	.	.
	.	.	.
	n ₁	y _{11n1}	y _{12n1}
BA	n ₁ +1	y ₂₁₁	y ₂₂₁
	.	.	.
	.	.	.
	.	.	.
	n ₁ +n ₂	y _{21n2}	y _{22n2}

젖소의 우유양에 미치는 효과를 분석할 때 이 계획을 사용할 수 있다. 개개의 젖소들은 우유생산량에 많은 변이를 보이고 있으며, 각각의 젖소에 대하여 두 종류의 사료를 수유기의 전반기와 후반기에 걸쳐서 모두 실험함으로써 개개의 젖소로부터 반복 측정치를 얻어야 한다. 이는 모든 젖소는 수유기의 전반기에 비하여 후반기에 우유양이 현저하게 감소하는데 이러한 현상을 실험 계획에 반영하고자 함이다.

두 종류의 사료(A,B)가 젖소의 우유양에 미치는 효과를 분석하는 자료는 다음 Table 1과 같이 표시할 수 있다.

2×2 교차실험계획은 Table 1의 y_{ijk}에 대한 모형으로서 다음 식(1)을 구성한다.

$$y_{ijk} = \mu + S_{ik} + \pi_j + \tau_{a(i,j)} + \lambda_{[i,j-1]} + e_{ijk} \quad (1)$$

단 μ : 일반평균

S_{ik} : i 번째군의 k 번째 개체의 효과, $i=1,2$,
 $k=1,\dots,n_i$

*본 논문은 제 24회 한국약제학회 총회 및 학술대회의 심포지움 "생물학적 동등성시험의 실제와 문제"에서 발표된 것임.

Table II-ANOVA F-검정의 분산분석표

Source	df.	SS	EMS
Between-subjects : Carry-over	1	$\frac{2n_1 n_2}{(n_1+n_2)} (\bar{y}_1 \dots - \bar{y}_2 \dots)^2$	$\frac{2n_1 n_2}{(n_1+n_2)} (\lambda_1 - \lambda_2)^2 + 2\sigma_s^2 + \sigma^2$
B-S residual	(n_1+n_2-2)	$\sum_{i=1}^2 \sum_{k=1}^{n_i} \frac{y_{i \cdot k}^2}{2} - \sum_{i=1}^2 \frac{y_{i \cdot}^2}{2n_i}$	$2\sigma_s^2 + \sigma^2$
Within-subjects : Direct Treatments (adjusted for Periods)	1	$\frac{n_1 n_2}{2(n_1+n_2)} (\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} + \bar{y}_{22})^2$	$\frac{2n_1 n_2}{(n_1+n_2)} \left[(\tau_1 - \tau_2) \frac{(\lambda_1 - \lambda_2)}{2} \right] + \sigma^2$
Periods (adjusted for Treatments)	1	$\frac{n_1 n_2}{2(n_1+n_2)} (\bar{y}_{11} - \bar{y}_{12} - \bar{y}_{21} - \bar{y}_{22})^2$	$\frac{2n_1 n_2}{(n_1+n_2)} (\pi_1 - \pi_2)^2 + \sigma^2$
W-S residual	(n_1+n_2-2)	$\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_i} y_{ijk}^2 - \sum_{i=1}^2 \sum_{k=1}^{n_i} \frac{y_{i \cdot k}^2}{2} - \sum_{i=1}^2 \sum_{j=1}^2 \frac{y_{ij}^2}{n} + \sum_{i=1}^2 \frac{y_{i \cdot}^2}{2n_i}$	σ^2
Total	$2(n_1+n_2)-1$	$\sum_{i=1}^2 \sum_{j=1}^2 \sum_{k=1}^{n_i} y_{ijk}^2 - \frac{y_{\cdot \cdot \cdot}^2}{2(n_1+n_2)}$	

π_j : j 번째 시기효과, $j=1, 2$

$\tau_{d(i, j)}$: i 번째군의 j 번째 시기에 취해진 처리효과

$\lambda_{d(i, j-1)}$: i 번째군의 j 번째 시기에 취해진 처리의 이월효과

$$\lambda_{d(i, 0)} = 0$$

e_{ijk} : i 번째군의 j 번째 시기의 k 번째 개체의 확률오차

*iid

$$s_{ik} \sim N(0, \sigma^2)$$

$$e_{ijk} \sim N(0, \sigma^2)$$

{ s_{ik} }는 { e_{ijk} }와 독립

Hills와 Armitage³⁾는 식(1)의 2x2 교차실험계획의 장점으로서는 같은 피험자에게 두 제제를 투여하여 효과를 비교하므로 두 군간을 평행으로 비교하는 평행계획보다 더 효율적이라는 점을 들고 있다. 또한 Jones와 Kenward⁴⁾는 식(1)의 모형의 구조적 제약으로 인하여 이월효과와 차이, 처리와 시기의 교호작용, 그리고 두 군간의 차이가 식별되지 않는 점을 단점으로 지적하고 있다.

*iid는 독립적이고 동일한 분포를 이룸(independent and identically distributed)의 영어표현 頭序이다.

생물학적 동등성시험의 통계적 처리

1980년대 초까지 2x2 교차실험계획의 표준적인 통계적 절차는 Grizzle⁵⁾과 Hills와 Armitage³⁾가 제안한 분산분석법이었다. B_N과 B_S를 각각 새로운 제제와 표준적인 제제의 생체이용율을 나타낸다고 할때, 분산분석법의 가설은 다음 식(2)와 같이 구성할 수 있다.

$$H_0 : B_N = B_S \tag{2}$$

$$H_a : B_N \neq B_S$$

식(1)에서 두개의 처리효과를 τ_1, τ_2 로, 두개의 이월효과를 λ_1, λ_2 로 표기하면 식(2)의 가설을 검정하는 전통적인 F검정의 분산분석표는 다음 Table 2와 같다. Table 2의 피험자간 변동중 이월효과(Carry-over effect)에 의한 변동은 두 군간의 차이에 의한 변동, 혹은 처리와 시기의 교호작용에 의한 변동으로도 표기되며 1절에서도 언급하였듯이 이 세가지 요인에 의한 변동은 식별되지 않는다.

Grizzle⁵⁾이 분산분석법으로 제시한 이른바 이단계 검정법은 우선 $\alpha=0.1$ 수준에서 이월효과(혹은 군 효과)가 같은지를 검정한다. 이는 피험자간 이월효과와 평균제곱의 기대값(EMS)과 피

험자간 잔차의 EMS를 살펴볼때 $H_0^1: \lambda_1 = \lambda_2$ 하에서는 두개 EMS의 비가 1이 되고 $H_a^1: \lambda_1 \neq \lambda_2$ 하에서는 그 비가 1 보다 커지므로, 이월효과의 평균제곱을 분자로 잔차의 평균제곱을 분모로 한 F통계량을 검정통계량으로 사용할 수 있다. 이제 이월 효과가 같다고 판단할 때에 다음 단계로 $\alpha = 0.05$ 수준에서 처리효과를 검정할 수 있다. 즉 $H_0^1: \lambda_1 = \lambda_2$ 하에서는 피험자내 처리효과의 EMS가 다음 식(3)과 같이 된다.

$$EMSTR = \frac{2n_1 + n_2}{(n_1 + n_2)} \{(\tau_1 - \tau_2)^2 + \sigma^2\} \quad (3)$$

단 EMSTR은 처리효과의 평균제곱의 기대값을 의미한다.

피험자내 잔차의 EMS와 식(3)의 비를 구성하면 그 비가 $H_0^2: \tau_1 = \tau_2$ 하에서는 1이 되고 $H_a: \tau_1 \neq \tau_2$ 하에서는 1보다 커지는 것을 알 수 있다. 따라서 $H_0^2: \tau_1 = \tau_2$ 를 검정하는 통계량으로서 피험자내 처리의 평균제곱을 분자로 잔차의 평균제곱을 분모로 하는 F통계량을 구성할 수 있다. 이 처리된 처리효과를 검정하기 이전에 사전검정(pretest)으로서 이월효과의 차이(혹은 군 효과의 차이)를 검정하므로 이러한 절차를 Grizzle의 이단계 검정법이라고 부른다.

전통적 가설설정과 검정절차의 문제점

앞에서도 언급하였듯이 생물학적 동등성시험에서 식(2)의 가설을 설정한 후, 식(1)의 2×2 교차실험계획으로 모형을 설정하고, Grizzle이 제시한 이단계 검정법을 사용하는 것이 1980년대 초까지 표준적인 절차로 알려졌다. 그러나 생물학적 동등성시험에서 식(2)의 가설설정과 Grizzle의 이단계검정법은 모두 심각한 문제점을 안고 있다.⁶⁻¹⁰⁾ 본 절에서는 식(2)의 가설설정과 이단계검정법 각각의 문제점을 살펴보고, 이러한 문제점의 임시방편적(ad-hoc) 대안으로 제시된 "80/20"규정을 살펴보고자 한다.

통계학에서 사전검정절차의 타당성에 대하여 많은 논란이 있다. 예를 들어 정규분포 가정을 전제로 하는 이 표본 t 검정을 할 때, 주어진 자료가 정규분포를 과연 이루는지를 검정하게 되는데 이를 사전검정이라고 한다. Grizzle의 이단계검정법도 대부분의 사전 검정절차가 같은 문

제점을 안고 있다.¹¹⁾ Freeman⁸⁾은 Grizzle의 이단계검정법의 여러가지 통계적 속성을 광범위하게 조사하여 실제로 적용하기에는 허점의 가능성이 너무 많은 절차라고 결론 짓고 있다. 또한 Senn¹²⁾은 이단계검정법은 이제 과학적 흥미의 대상도 아니고 분석의 도구로도 사용되지 않는, 단지 역사적 흥미거리일 뿐이라고 지적하고 있기도 하다. 원래 통계적 가설검정의 논리는 제시하고자 하는 가설을 대립가설로 설정하도록 되어 있다. 그러나, 식(2)의 가설설정에서는 비동등성(inequivalence)이 바로 실험자가 제시(혹은 주장)하고자 하는 가설로 설정되어 있는 것이다. 그러므로 식(2)의 가설 설정하에서는 동등성이 강력한 증거를 근거로 하여 제시되는 가설이라기 보다는 증거를 제시 못할 경우에 디폴트(default)로서 얻어지는 결론이 된다.¹³⁾ 이와 관련하여 유의성검정의 개념을 정립한 Fisher¹⁴⁾도 "유의성 검정은 (귀무)가설을 채택하는 기준을 제시하지는 않는다."라는 점을 분명히 하였다.

식(2)의 가설설정의 다른 문제점으로서 다음의 두가지를 추가할 수 있다. 첫째, 표본크기가 크기만 하면 임상적으로 중요치 않은 차이가 검색되어 동등성가설이 기각된다. 둘째로 표본크기를 의도적으로 작게하거나 비효율적 실험설계로 실험오차를 크게하면 보수적인 검정이 되어서 H_0 을 기각하기가 어렵고, 따라서 생물학적 동등성을 입증하기 쉬어진다.

일반적으로 가설검정에, 세부적으로는 생물학적 동등성시험에 다음의 두가지 위험(risk)이 있다. 생물학적으로 동등하지 않는데 실험결과 생물학적으로 동등하다는 결론이 나는 소비자위험과 생물학적으로 동등한데 실험결과 생물학적으로 동등하지 않다는 결론이 나는 생산자 위험이 그것이다. 소비자 위험은 규제당국이 통제하여야 할 위험이기 때문에 흔히 규제당국의 위험이라고도 불리워지며 사회전반에 미치는 효과로 보아 생산자 위험보다는 중요하므로 제 1종오류(혹은 α 오류)가 된다. 생산자 위험은 생산자가 스스로 감당할 수 있는 위험이 된다. 즉 생산자는 실험설계에서 표본크기를 크게 함으로써 생산자위험을 줄일 수 있어야 한다. 그러나, 식(2)의 가설인 $H_0: B_N = B_S$ 대 $H_a: B_N \neq B_S$ 에서는 이러한 두가지

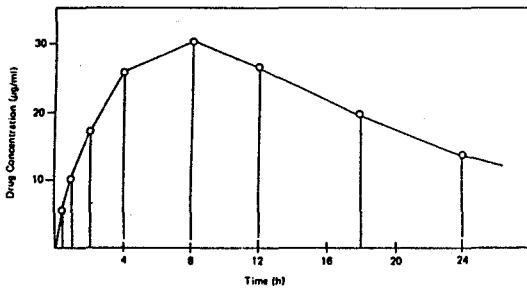


Figure 1 - 전형적인 시간 - 농도 곡선.

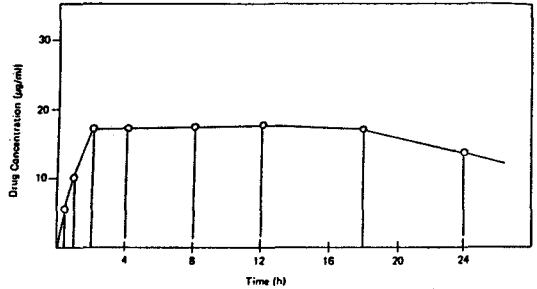


Figure 2 - T_{max} 의 범위가 큰 시간-농도 곡선.

위험이 제대로 식별이 되지않고, 소비자위험을 통제할 수도 없게 된다.

앞에서 언급하였듯이 식(2)의 가설설정에서는 표본크기를 작게하면 생물학적 동등성을 입증하기 쉬워진다. 이러한 문제점의 임시방편적(ad-hoc) 대안으로 미국의 FDA는 이미 실시된 실험으로부터 역으로 계산하여 20%의 차이를 검색해 내는 검정력이 최소한 80%가 되도록 하는 이른바 "80/20"규정을 적용하였다.¹⁰⁾ 따라서 이미 종료된 실험으로부터 20%의 차이를 검색해 내는 검정력을 역으로 계산하고 이 검정력이 80% 미만이면 그 실험결과로부터 생물학적 동등성 결론을 유도하는 것은 의문시되었다. 검정력을 사후에 역으로 계산하는 절차는 최소한의 표본크기를 유지시켜 준다는 점에서 의의를 찾을 수는 있겠으나, 검정력이 확율인 만큼 일단 시행이 끝난 실험에서 확율을 계산한다는 것은 확율의 기본 개념에 어긋나는 일이다.

"80/20"규정은 F검정을 이용한 분산분석법의 임시방편적 대안으로 제시된 내용이고, 이 두가지 절차를 종합한 방법을 흔히 검정력 접근법(power approach)이라고 부른다.

검정력 접근법의 대안 : 두개의 단측검정절차와 신뢰구간절차

Anderson과 Hauck⁶⁾은 생물학적 동등성시험의 귀무가설과 대립가설을 다음식 (4)와 같이 설정하고 있다.

$$\begin{aligned}
 H_0 : B_N - B_S \leq \theta_1 \quad \text{혹은} \quad B_N - B_S \leq \theta_2 \\
 H_a : \theta_1 < B_N - B_S < \theta_2
 \end{aligned}
 \tag{4}$$

식(4)의 가설설정은 식(2)의 가설설정이 가지고

있는 문제점을 전혀 갖고 있지 않다. 식(4)의 가설검정 절차로서 Anderson과 Hauck⁶⁾는 비심 t 분포에 기초한 비고전적 형태의 검정통계량을 제시하고 있고, Schuirmann⁷⁾은 두개의 단측 검정법을 제시하고 있다. Schuirmann⁷⁾의 두개의 단측검정절차는 $(1-2\alpha)100\%$ (혹은 90%) 신뢰구간에 기초한 절차와 일치한다.

식(2)의 가설검정이 안고 있던 문제점을 일찌기 파악한 Metzler¹⁾, Westlake¹⁶⁾와 그외에 Mandallaz와 Mau¹⁷⁾는 B_N/B_S 의 $(1-\alpha)100\%$ 신뢰구간을 구하여 그 신뢰구간이 (θ_1, θ_2) 에 포함되면 생물학적으로 동등한 것으로 간주하는 방법을 제안하기도 하였다. (θ_1, θ_2) 는 미국의 경우는 (0.8, 1.2), 캐나다나 유럽연합은 (0.8, 1.25)의 값을 취하고 있으며, B_N/B_S 의 신뢰구간 추정법으로는 Fieller의 정리¹⁸⁻¹⁹⁾를 이용하고 있다.

Senn²⁰⁾이 지적하고 있듯이 2×2 교차실험계획의 분석 방법으로서 어느 방법이 옳은지에 대하여는 놀랄 만큼의 불일치가 있다. 그리고 Schuirmann의 두개의 단측검정법과 Anderson과 Hauck의 절차를 종합하는 절차에 대하여는 최근까지도 활발한 논의가 이루어지고 있다.²¹⁻²¹⁾ 그러는 가운데 Schuirmann⁷⁾은 자신이 제시한 두개의 단측검정절차가 대부분의 경우에 대하여 검정력 접근법보다 우수한 절차임을 보였다. 미국의 FDA는 1980년대초까지 검정력 접근법과 75%의 피험자에 대하여 B_N/B_S 의 비가 (0.75, 1.25)안에 있어야 하는 이른바 75/75규정을 종종 병행하여 사용하였다. 그러나 두 방법이 모두 다 열등한 절차임을 밝혀지자, 이 방법들의 사용을 중지하였다. 그리고 1986년 부터는 두개의 단측검정절차를 의사

결정방법으로 사용하고 있다.²³⁾

AUC, C_{max}, T_{max}의 통계적 특징

생체이용율의 측도로서 흔히 AUC, C_{max}와 T_{max}를 사용하고 있다. AUC(Area Under the Curve)는 흔히 사다리꼴 공식을 사용하여 계산한다. 다음 Fig. 1은 전형적인 시간 - 농도 곡선의 형태를 나타내주고 있다.

Fig. 1과 유사하게 가령 시점 t₀, t₁, ..., t_n에서 측정된 n+1개의 관찰치가 Y₀, Y₁, ..., Y_n 이라고 할 때 AUC는 다음 식(5)와 같이 얻어진다.

$$AUC = (Y_0 + Y_1)(t_1 - t_0)/2 + (Y_1 + Y_2)(t_2 - t_1)/2 + \dots + (Y_{n-1} + Y_n)(t_n - t_{n-1})/2 = \sum_{i=0}^n W_i Y_i \dots \quad (5)$$

단 W_i는 Y_i에 대한 가중치.

식(5)에서 보듯이 AUC는 n+1개의 관찰치의 가중평균이므로 상당히 안정적이고, 통계적으로는 중심극한의 정리를 적용하여 정규분포로써 AUC의 분포를 근사할 수 있다.

C_{max}는 관찰치 Y₀, ..., Y_n의 최대값을 의미한다. 그리고 C_{max}의 정의상 n개의 관찰치에 기초한 C_{max}, n은 n+1개 관찰치에 기초한 C_{max}, n+1보다 크지 않게 된다. T_{max}는 최대농도에 도달하는 시점을 나타낸다. 그러나, 다음 Fig. 2의 경우에서 보듯이 T_{max}는 민감도에 비하여 변동이 너무 큰 단점을 가지고 있다.

생체이용율의 측도로서 C_{max}와 T_{max}, 특히 T_{max}의 사용을 주저하는 경우도 있으나, 이는 아직까지 표준적인 관습은 아니다. 그리고 아직도 당분간은 C_{max}와 T_{max}는 생물학적 동등성연구에서 사용되리라 믿는다. 단지 여러규제당국에서 C_{max}와 T_{max}에 관한 한 현재까지의 표준적 관습에 다소 유통성을 보이는 경향이 있다.²⁴⁾

문 헌

- 1) C.M. Metzler, Bioavailability : a problem in equivalence, *Biometrics*, **30**, 309-317 (1974).
- 2) W.G. Cochran and G.M. Cox, *Experimental Designs*, Wiley, New York, p. 129 (1957).
- 3) M. Hills and P. Armitage, The two - period cross - over clinical trial, *Br. J. Clin. Phar-*

- macol.*, **8**, 7-20 (1979).
- 4) B. Jones and M.G. Kenward, *Design and Analysis of Cross - over Trials*, Chapman and Hall, London, p. 40 (1989).
- 5) J.E. Grizzle, The two-period changeover design and its use in clinical trials, *Biometrics*, **21**, 467-480 (1965).
- 6) S. Anderson and W.W. Hauck, A new procedure for testing equivalence in comparative bioavailability and other clinical trials, *Communications in Statistics*, **A12**, 2663-2692 (1983).
- 7) D.J. Schuurmann, A comparison of the two one - sided tests procedure and the power approach for assessing the equivalence of average bioavailability, *J. Pharmacokin. Biopharm.*, **15**, 657-680 (1987).
- 8) P.R. Freeman, The performance of the two - stage analysis of two - treatment, two - period cross - over trials, *Statistics in Medicine*, **8**, 1421-1432 (1989).
- 9) B.E. Rodda, Bioavailability : Designs and Analysis, in *Statistical Methodology in the Pharmaceutical Sciences*, D.A. Berry (Ed), Marcel Dekker, Basel (1990).
- 10) W.W. Hauck and S. Anderson, Types of bioequivalence and related statistical considerations, *Int. J. Clin. Pharmacol., Ther. Toxicol.*, **30**, 181-187 (1992).
- 11) S. Senn, *Cross - over Trials in Clinical Research*, Wiley, N.Y., pp. 12-13 (1993).
- 12) S. Senn, *ibid* pp. 52-54 (1993).
- 13) B.E. Rodda, Bioavailability : Designs and Analysis, in *Statistical Methodology in the Pharmaceutical Sciences*, D.A. Berry, Marcel Dekker, Basel, p.73 (1990).
- 14) R.A. Fisher, *Statistical Methods and Scientific Inference*, Hafner Press, N.Y., p. 45 (1973).
- 15) B.E. Rodda, Bioavailability : Designs and Analysis, in *Statistical Methodology in the Pharmaceutical Sciences*, ed. D.A. Berry, Marcel Dekker, Barseel, p. 71 (1990).
- 16) W.J. Westlake, Symmetrical confidence intervals for bioequivalence trials, *Biometrics*, **32**, 741-744 (1976).
- 17) D. Mandallaz and J. Mau, Comparison of

- differnt methods for decision - making in bioequivalence assessment, *ibid*, **37**, 213-222 (1981).
- 18) G. Casella and R.L. Berger, *Statistical Inference*, Wadsworth & Brooks/Cole, Pacific Grove, pp. 459-460 (1990).
 - 19) J. Vourinen and J. Touminen, Fieller's confidence intervals for the ratio of two means in the assessment of average bioequivalence from crossover data, *Statistics in Medicine*, **13**, 2531-2545 (1994).
 - 20) S. Senn, *Cross - over Trials in Clinical Research*, Wiley, N.Y., p. 180 (1993).
 - 21) A. Munk, An improvement on commonly used tests in bioequivalence assessment, *Biometrics*, **49**, 1225-1230 (1993).
 - 22) Y.W. Nam, Correspondence (on the method of combining double t - test and Anderson - Hauck test), *ibid*, **50**, 884-886 (1993).
 - 23) S.V. Dighe and W.P. Adams, Bioequivalence : A United States Regulatory Perspective, in *Pharmaceutical Bioequivalence*, P.G. Welling, F. L.S. Tse and S.V. Dighe(Ed.), Marcel Dekker, Basel, pp. 367-372 (1991).
 - 24) S. Senn, *Cross - over Trials in Clinical Research*, Wiley, N.Y., p. 178 (1993).