

Assessing Bioequivalence of Variabilities in 2x2 Crossover Design¹⁾

Sang-Gue Park²⁾ · Junghoon Jang³⁾

Abstract

Several statistical procedures for assessment of bioequivalence of variabilities between two drug formulations in bioequivalence trials are reviewed and modified methods for assessing total variability are suggested. The problem of the current US FDA aggregate criterion for population bioequivalence and the necessity of disaggregate criterion are discussed with an illustrated example.

Keywords : Average Bioequivalence, Crossover Design, Inter-Variability, Intra-Variability, Population Bioequivalence

1. 서론

생물학적 동등성(bioequivalence)란 제제(drugs)간에 또는 동일 제제의 서로 다른 형식(drug formulation)간의 치료효과가 동일한 것을 의미하고, 생물학적 동등성 시험은 제제 간에 또는 동일 제제의 서로 다른 형식 간에 생물학적 동등성을 입증하는 시험을 의미한다. 즉, 두 제제 중 신약으로 허가되어 안전성과 유효성이 확인된 제제를 대조약(reference drug; *R*) 그리고 생물학적 동등성 시험의 대상이 되는 제제를 시험약(test drug; *T*)이라 할 때, 생물학적 동등성(이하 생동성) 시험이란 약제학적으로 동등한 두 제제의 치료학적 동등성을 확인하기 위해 두 제제 간의 약물동태학적 성질의 동등성을 통계적으로 확인하는 시험이라 할 수 있다.

두 제제간의 생동성을 비교 평가하는데 의약품의 약물동태학적 성질을 사용한다. 제제의 약물동태학적 성질의 중요지표로는 흡수율과 흡수량 외에 배설반감기, 배설율 및 대사율이 있으나 실제 생체이용률(bioavailability; BA)의 중요지표로 혈중농도-시

1) 본 연구는 2007년도 식품의약품안전청 자체연구사업의 연구비지원(07161유해물657)에 의해 수행되었음.

2) 서울시 동작구 흑석동 221 중앙대학교 수학과통계학부 교수
E-mail : spark@cau.ac.kr

3) 서울시 은평구 통일로 194 식품의약품안전청 국립독성연구원 응용통계팀 보건연구원
E-mail : statjjh@kfda.go.kr

간곡선하 면적(AUC) 및 최고혈중농도(C_{max}) 그리고 최고혈중농도 도달시간(t_{max})가 많이 활용된다. 2001년에 고시된 생동성 시험 기준에 의하면, 2x2 교차 설계를 원칙으로 하고, 시험을 통해 채혈된 혈액으로부터 계산된 두 제제의 피험자들의 AUC 및 C_{max} 에 대한 평균값을 비교하여 생동성을 평가하고 있으며, t_{max} 는 참고로만 활용하고 있다.

생동성 시험 기준은 1988년 10월 28일 국립보건안전연구원 고시 제 1988-1호로 처음 제정되어, 1990년 1차 개정, 1994년 2차 개정을 거쳐 최근 1998년 8월 26일에 공포된 제 7차 개정고시인 식품의약품안전청고시 제 1998-86호에 따라 생동성 시험을 실시해 왔고, 제 8차 개정안이 2001년 2월 22일에 입안 예고되어 각계의 의견 수렴을 거친 후, 2001년 8월 식품의약품안전청 고시 제 2001-21호에 의해 개정 고시되었다.

최근 미국을 위시한 일부 선진국에서 제제간의 생동성을 입증하는데 시험에서 얻어진 피험자들의 생체이용률의 평균을 비교하는데 그치지 않고, 피험자간 혹은 피험자내 변동의 개념을 도입하여 제제간의 생동성을 총체적으로 평가하는 새로운 기준을 발표하였다(US FDA, 2001). 새롭게 도입된 집단 생동성(population bioequivalence)이란 개념은 오리지널 제제와 제네릭 제제가 있을 때 두 제제의 생체이용률의 분포가 동일하기 때문에 어떤 제제라도 자유롭게 처방할 수 있다는 처방성(prescribability)을 의미하고, 개인 생동성(individual bioequivalence)이란 개념은 오리지널 제제 뿐만 아니라 다른 제네릭 제제까지도 대체할 수 있다는 전환성(switchability)을 의미한다. 이러한 개념은 피험자 변동이 동일하다는 가정 하에서 평균만 비교하는 현재 기준보다 제네릭 제제의 안전성에 크게 기여할 수 있다는 면에서 선진국을 중심으로 도입되어 활발하게 연구되고 있다. 아직 우리나라의 경우도 이러한 집단 생동성과 개인 생동성의 개념을 도입하고 있지는 못하지만 서서히 그 논의가 시작되고 있다.

우리나라의 경우 현재 기준인 평균 생동성을 평가하는 모형에서도 쉽게 확장할 수 있는 집단 생동성에 관심을 보이고 있다. 최근 우리나라에서도 Park et al.(2005)에 의해서 2x2 교차설계법을 활용하였을 때 제제간의 집단 생동성을 평가하는 통계적 방법을 US FDA 방법과 비교하여 논하였다.

집단 생동성을 평가하는 통계적 방법은 크게 평균과 분산을 의미를 가진 하나의 합수 형태로 표현하여 통계적 추론을 하는 통합적인 방법(aggregate approach)과 평균과 분산을 각각 따로 통계적으로 추론하여 제제간의 생동성 여부를 평가하는 분리적인 방법(disaggregate approach)이 사용될 수 있지만, 미국의 경우 개념의 단순성을 강조하여 통합적인 방법을 현재 활용하고 있다. 하지만 이 통합적인 방법이 오히려 평균 생동성보다 더 쉬운 기준일 수 있다는 단점이 지적되면서(Chow and Liu, 2000) 최근 들어 분리적인 방법에 대한 관심이 더 높아지고 있다.

본 연구논문은 집단 생동성이 제제간의 생동성을 평가하는 표준으로 활용될 경우, 분리적인 방법이 사용되었을 때 피험자내 변동(intra-subject variability), 피험자간 변동(inter-subject variability)을 포함한 총분산에 관한 생동성을 평가하는 통계적 추론을 논하고자 한다.

2장에서는 2x2 교차설계법에서 두 집단간 피험자간 변동을 일정하다고 놓고 피험자내 변동성의 동등성을 평가하는 통계적 방법과 피험자간 변동을 포함한 총변동성을 평가하는 통계적 방법을 논의하고, 집단 생동성에서 제안되었던 기존의 통계적 방법을 활용하여 분리적인 방법에서 활용될 수 있도록 총변동성을 평가하는 통계적 방법을 제안한다. 3장에서는 현재 US FDA에서 활용되고 있는 집단 생동성의 문제를 논

하고 분리적인 기준의 필요성과 중요성을 실제 사례를 들어 설명하고자 한다.

2. 변동의 생동성을 평가하는 통계적 방법

2.1 통계적 모형

생동성 시험에서 가장 일반적인 교차설계는 m 번 반복된 $2 \times 2m$ 교차설계법이다. $m=1$ 인 경우 현재 우리나라에서 활용하고 있는 반복이 없는 표준 2×2 교차설계법이 고, $m=2$ 인 경우 미국 등 선진국에서 많이 활용하고 있는 2×4 교차설계법이 된다. 2×4 교차설계법은 실험 수행은 조금 더 어려워지지만 제제의 평균 효과를 비롯하여 피험자간 분산, 피험자내 분산, 제제와 피험자간의 교호작용 등 제제의 효과를 다양한 관점에서 평가할 수 있다는 장점이 있다.

시험에 참여한 $(n_1 + n_2)$ 피험자 중 n_1 명을 첫 번째 순서집단에, n_2 피험자를 두 번째 순서집단에 무작위로 배정한다고 하자. 이때 y_{ijkl} 은 i -번째 순서집단의 j -번째 피험자가 k -제제($k = T, R$)를 l -번($l = 1, \dots, m$) 반복했을 때 측정된 생체이용률 값을 의미한다. y_{ijkl} 은 우리나라와 같이 로그 변환된 값을 기본으로 사용하는 경우가 많다. 이 $2 \times 2m$ 교차설계법을 통계적으로 다음과 같은 모형으로 표현할 수 있다.

$$y_{ijkl} = \mu_k + \gamma_{ikl} + S_{ijk} + \epsilon_{ijkl} \tag{1}$$

여기서 μ_k 는 k -번째 제제효과를, γ_{ikl} 은 i -번째 순서집단에서 k -번째 제제의 l -번 반복되었을 때 가지는 고정효과(fixed effect)를 의미하고, $\sum_i \sum_l \gamma_{ikl} = 0$ 과 같은 제약을 가진다. 피험자 변동 (S_{ijT}, S_{ijR})은 i -번째 순서집단에서 k -번째 제제를 복용한 j -번째 피험자에서 발생하는 피험자간 변동으로 평균이 $(0, 0)'$ 이고, 분산 공분산 행렬이 $\Sigma_S = \begin{pmatrix} \sigma_{BT}^2 & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{BR}^2 \end{pmatrix}$ 인 이변량 정규분포를 따른다. 분산 공분산 행렬에서 σ_{BT}^2 와 σ_{BR}^2 는 시험약과 대조약을 각각 복용하였을 때 피험자간 분산을 의미한다. ϵ_{ijkl} 은 i -번째 순서집단의 j -번째 피험자가 복용한 k -번째 제제에서 발생하는 제제와 관련된 피험자내 변동으로 평균은 0이고, 분산은 σ_{WT}^2 과 σ_{WR}^2 인 독립적인 정규분포를 따른다. 두 개의 변동 (S_{ijT}, S_{ijR})과 ϵ_{ijkl} 은 서로 독립으로 가정한다.

반복이 없는 표준 2×2 교차설계법에서 모형 (1)은 첨자 l 이 생략하고 변동을 하나로 묶어 다음과 같이 표현할 수 있다.

$$y_{ijk} = \mu_k + \gamma_{ik} + \epsilon_{ijk}^* \tag{2}$$

여기서 총변동 ($\epsilon_{ijT}^*, \epsilon_{ijR}^*$)은 제제변동과 피험자변동을 합한 개념으로 평균이 $(0, 0)'$ 이고 분산공분산 행렬 $\Sigma_T = \begin{pmatrix} \sigma_{WT}^2 + \sigma_{BT}^2 & \rho\sigma_{BT}\sigma_{BR} \\ \rho\sigma_{BT}\sigma_{BR} & \sigma_{WR}^2 + \sigma_{BR}^2 \end{pmatrix}$ 을 가진다.

2.2 피험자내 변동 비교

실제 제제간의 효과를 비교하는 생동성 시험을 수행할 때 시험 설계자가 제제간의 효과의 비교만큼 많은 신경을 쓰는 부분 중의 하나가 두 집단으로 나누어진 피험자내 변동의 동질성 여부이다. 이는 두 집단의 평균 비교와 관련된 통계적 추론 절차가 기본적으로 비교하려는 두 집단 간의 등분산성을 기본 가정으로 하고 있기 때문이다. 이 가정이 만족되지 않으면 평균에 관한 통계적 추론의 타당성이 의심이 되기 때문에 FDA에서도 규정을 통해 시험 설계자가 피험자를 두 집단으로 나누는 확률화 과정을 통해 두 집단 간의 등분산성을 만족하도록 권하고 있다.

일반적으로 생동성 시험은 승인된 시험계획서에 의해서 비교적 까다롭게 통제하면서 실시하기 때문에 많은 경우 시험에 참가한 피험자간 변동이 잘 통제되고 있는 경우가 많다. 만약 피험자간 분산 σ_{BT}^2 과 σ_{BR}^2 이 동일하다고 가정한다면 한 피험자의 생체이용률값 벡터 (y_{iT}, y_{iR}) 는 좀더 단순한 분산 공분산 행렬 $\Sigma_1 = \begin{pmatrix} \sigma_{WT}^2 + \sigma_B^2 & \sigma_B^2 \\ \sigma_B^2 & \sigma_{WR}^2 + \sigma_B^2 \end{pmatrix}$ 을 가지게 된다. 이럴 경우 대조약과 시험약의 평균 뿐만 아니라 피험자내 변동, 즉 제제변동까지 동일하다는 평가를 내릴 수 있다면 두 제제간의 동등성은 인정할 수 있을 것이다.

대조약과 시험약의 피험자내 변동의 생동성은 다음과 같은 가설로 평가할 수 있다.

$$H_0: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} \notin (\delta_1, \delta_2) \text{ vs. } H_1: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} \in (\delta_1, \delta_2), \quad (3)$$

여기서 $(\delta_1 < 1, \delta_2 > 1)$ 은 피험자내 변동의 동등성 기준값으로 아직까지는 규정에 의해서 구체적으로 결정되지 않고 있지만, 문헌을 통해서(Chow & Liu, 2000) 일반적으로 상위값 δ_2 의 경우 (1.25, 1.5)에서 사용되고 있다. 하위값 δ_1 은 상위값과 1을 기준으로 대칭값으로 주로 결정된다.

Liu(1991), Liu & Chow(1992)는 가설 (3)을 검정하는데 Schuirmann(1987)이 평균 생동성을 평가하기 위해 가설을 두 개의 단측검정으로 나누어 적용했듯이 다음과 같은 두 개의 가설로 나누어 각각 검정하는 방법을 제안하였다.

$$H_{01}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} \leq \delta_1 \text{ vs. } H_{11}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} > \delta_1, \quad (4)$$

$$H_{02}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} \geq \delta_2 \text{ vs. } H_{12}: \frac{\sigma_{WT}^2}{\sigma_{WR}^2} < \delta_2, \quad (5)$$

가설 (4)와 (5)는 모형 (2)를 기준으로 기간효과(period effect) 존재 여부에 따라 두 가지 방향에서 접근할 수 있다. 기간효과가 존재하지 않는 경우 다음과 같은 변환된

확률 변수를 사용하여 검정통계량을 구축할 수 있다.

$$v_{ij} = y_{ijT} - y_{ijR}, \quad u_{Lij} = y_{ijT} + \delta_1 y_{ijR}, \quad u_{Uij} = y_{ijT} + \delta_2 y_{ijR}.$$

확률변수 v_{ij}, u_{Lij}, u_{Uij} 에 대해 다음과 같은 통계적 성질은 쉽게 확인할 수 있다.

$$\begin{aligned} \text{Var}(v_{ij}) &= \sigma_{WT}^2 + \sigma_{WR}^2, \\ \text{Var}(u_{Lij}) &= \sigma_{WT}^2 + \delta_1^2 \sigma_{WR}^2 + (1 + \delta_1)^2 \sigma_B^2 \\ \text{Var}(u_{Uij}) &= \sigma_{WT}^2 + \delta_2^2 \sigma_{WR}^2 + (1 + \delta_2)^2 \sigma_B^2, \end{aligned}$$

$$\text{Cov}(v_{ij}, u_{Lij}) = \sigma_{WT}^2 - \delta_1 \sigma_{WR}^2 \quad \text{and} \quad \text{Cov}(v_{ij}, u_{Uij}) = \sigma_{WT}^2 - \delta_2 \sigma_{WR}^2.$$

이러한 통계적 성질로부터 $\text{Cov}(v_{ij}, u_{Lij})$ 와 $\text{Cov}(v_{ij}, u_{Uij})$ 는 가설 (4)와 (5)를 검정할 수 있는 검정통계량으로 활용될 수 있음을 알 수 있다. 즉, 공분산과 상관계수의 관련성을 이용해서 가설 (4)와 (5)는 다음과 같은 가설로 다시 표현할 수 있다.

$$H_{01} : \rho_L \leq 0 \quad \text{vs} \quad H_{11} : \rho_L > 0, \tag{6}$$

$$H_{02} : \rho_U \geq 0 \quad \text{vs} \quad H_{12} : \rho_U < 0, \tag{7}$$

여기서 ρ_L 와 ρ_U 는 각각 v_{ij} 와 u_{Lij} , v_{ij} 와 u_{Uij} 의 상관계수이다. 가설 (6)과 (7)은 두 집단간의 상관계수를 검정하는 잘 알려진 t -검정으로 검정될 수 있다. 가설 (6)과 (7)은 유의수준 α 에서 다음과 같이 각각 검정될 수 있다.

$$t_L = \frac{\hat{\rho}_L}{\sqrt{\frac{1 - \hat{\rho}_L^2}{n_1 + n_2 - 2}}} > t(\alpha, n_1 + n_2 - 2), \tag{8}$$

$$t_U = \frac{\hat{\rho}_U}{\sqrt{\frac{1 - \hat{\rho}_U^2}{n_1 + n_2 - 2}}} < t(1 - \alpha, n_1 + n_2 - 2). \tag{9}$$

여기서 $t(\alpha, n_1 + n_2 - 2)$ 는 자유도가 $n_1 + n_2 - 2$ 인 t -분포에서 상위 α -백분위수이다. 만약 이 두 검정이 동시에 기가될 때 우리는 유의수준 α 에서 두 제제간의 피험자 내 변동의 생동성을 인정할 수 있다.

기간효과가 존재하는 경우는 모형에서 기간효과를 제거하는 형식으로 다음과 같이 변환해 사용한다.

$$\begin{aligned} v_{ij}^* &= (y_{ijT} - \overline{y_{i.T}}) - (y_{ijR} - \overline{y_{i.R}}), \\ u_{Lij}^* &= (y_{ijT} - \overline{y_{i.T}}) + \delta_1 (y_{ijR} - \overline{y_{i.R}}), \quad u_{Uij}^* = (y_{ijT} - \overline{y_{i.T}}) + \delta_2 (y_{ijR} - \overline{y_{i.R}}). \end{aligned}$$

여기서 $\overline{y_{i.T}} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ijT}$, $\overline{y_{i.R}} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ijR}$ 이다.

변환된 확률변수들의 통계적 성질들은 다음과 같다.

$$\text{Var}(v_{ij}^*) = \frac{n_i - 1}{n_i} (\sigma_{WT}^2 + \sigma_{WR}^2),$$

$$\text{Var}(u_{Lij}^*) = \frac{n_i - 1}{n_i} (\sigma_{WT}^2 + \delta_1^2 \sigma_{WR}^2 + (1 + \delta_1)^2 \sigma_B^2).$$

$$\text{Var}(u_{Uij}^*) = \frac{n_i - 1}{n_i} (\sigma_{WT}^2 + \delta_2^2 \sigma_{WR}^2 + (1 + \delta_2)^2 \sigma_B^2).$$

기간효과가 없는 경우와 마찬가지로 가설 (6)과 (7)은 v_{ij}^* , u_{Lij}^* 와 u_{Uij}^* 의 상관계수로 검정될 수 있다. 다음과 같은 제제와 관련된 전체변동과 t -분포와 F -분포의 관계식을 이용하면 복잡한 변수변환을 피해 원변수를 그대로 이용할 수 있다.

$$\widehat{\sigma}_{TT}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ijT} - \overline{y_{i.T}})^2,$$

$$\widehat{\sigma}_{TR}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ijR} - \overline{y_{i.R}})^2,$$

$$\widehat{\sigma}_{TTR}^2 = \frac{1}{n_1 + n_2 - 2} \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ijT} - \overline{y_{i.T}})(y_{ijR} - \overline{y_{i.R}}).$$

이로부터 두 통계량

$$F_L = \frac{(n_1 + n_2 - 3)(\widehat{\sigma}_{TT}^2 - \delta_1 \widehat{\sigma}_{TR}^2 + (\delta_1 - 1)\widehat{\sigma}_{TTR}^2)^2}{(1 + \delta_1)^2 (\widehat{\sigma}_{TT}^2 \widehat{\sigma}_{TR}^2 - \widehat{\sigma}_{TTR}^2)},$$

$$F_U = \frac{(n_1 + n_2 - 3)(\widehat{\sigma}_{TT}^2 - \delta_2 \widehat{\sigma}_{TR}^2 + (\delta_2 - 1)\widehat{\sigma}_{TTR}^2)^2}{(1 + \delta_2)^2 (\widehat{\sigma}_{TT}^2 \widehat{\sigma}_{TR}^2 - \widehat{\sigma}_{TTR}^2)},$$

를 정의하고, 검정통계량 t_L 과 t_U 을 다음과 같이 정의한다.

$$t_L = (F_L)^{1/2} \quad \text{and} \quad t_U = (F_U)^{1/2}.$$

만약 $\widehat{\sigma}_{TT}^2 - \delta_1 \widehat{\sigma}_{TR}^2 + (\delta_1 - 1)\widehat{\sigma}_{TTR}^2 > 0$ 이면 유의수준 α 에서 가설 (4)는 다음과 같이 검정할 수 있다.

$$|t_L| > t(\alpha, n_1 + n_2 - 3).$$

비슷하게 만약 $\widehat{\sigma}_{TT}^2 - \delta_2 \widehat{\sigma}_{TR}^2 + (\delta_2 - 1)\widehat{\sigma}_{TTR}^2 < 0$ 이면 유의수준 α 에서 가설 (5)는 다음과 같이 검정할 수 있다.

$$|t_U| > t(\alpha, n_1 + n_2 - 3).$$

이 두 가설이 모두 기각되면 유의수준 α 에서 두 제제간의 피험자내 변동의 생동성을 인정할 수 있다.

Guilbaud(1993)는 Liu & Chow(1992)와는 달리 피험자내 변동의 생동성 검정에 불

편성을 가진 정확한 통계적 추론을 제안하였다. Guilbaud는 피험자내 변동 비 $\delta = \frac{\sigma_{WT}^2}{\sigma_{WR}^2}$ 를 $\psi = \frac{\sigma_{WT}^2 - \sigma_{WR}^2}{\sigma_{WT}^2 + \sigma_{WR}^2}$ 에 적용시켜 통계적 추론을 전개하였다. 여기서 ψ 는 $\psi = \frac{\delta - 1}{\delta + 1}$ 로 표현될 수 있고 $\delta > 0$ 의 증가함수이며, $-1 < \psi < 1$ 의 값을 가진다. ψ 의 관점에서 가설 (3)은 다음과 같이 표현할 수 있다.

$$H_0: \psi \leq \psi_U \text{ or } \psi \geq \psi_L \text{ v.s. } H_1: \psi_L < \psi < \psi_U, \tag{10}$$

여기서 ψ_L 와 ψ_U 은 ψ 에 δ_L 과 δ_U 를 적용시킨 값이다. 확률변수 v_{ij} 와 $u_{ij} = y_{ijT} + y_{ijR}$ 를 이용하면 ψ 의 자연스런 추정량은 다음과 같다.

$$\hat{\psi} = \frac{S_{vu}^2}{S_{vv}^2}.$$

여기서 $S_{vv}^2 = \frac{1}{n_1 + n_2 - 1} \sum_{i=1}^2 \sum_{j=1}^{n_i} (v_{ij} - \bar{v})^2$, $S_{vu}^2 = \frac{1}{n_1 + n_2 - 1} \sum_{i=1}^2 \sum_{j=1}^{n_i} (v_{ij} - \bar{v})(u_{ij} - \bar{u})$ 이다. 단, $\bar{v} = \frac{1}{n_1 + n_2} \sum_{i=1}^2 \sum_{j=1}^{n_i} v_{ij}$ 과 $\bar{u} = \frac{1}{n_1 + n_2} \sum_{i=1}^2 \sum_{j=1}^{n_i} u_{ij}$ 이다. ψ 의 $(1-\alpha) \times 100\%$ 신뢰구간 (L_ψ, U_ψ) 은 다음과 같이 추정된다.

$$L_\psi(U_\psi) = \hat{\psi} \pm t(\alpha/2, n_1 + n_2 - 3)S_\psi, \tag{11}$$

여기서 $S_\psi^2 = \frac{(S_{uu}^2/S_{vv}^2 - \hat{\psi}^2)}{n_1 + n_2 - 3}$ 이고 $S_{uu}^2 = \frac{1}{n_1 + n_2 - 1} \sum_{i=1}^2 \sum_{j=1}^{n_i} (u_{ij} - \bar{u})^2$ 이다.

만약 $(1-\alpha) \times 100\%$ 신뢰구간 (L_ψ, U_ψ) 이 대립가설 (10)에 속하면 피험자간 변동의 생동성은 인정된다. Guilbaud의 통계적 추론 절차는 Wang(1997)에 의해서 불편성을 가진 통계적 절차로 확장되었으나, 간혹 추정량 $\hat{\psi}$ 이 이론적인 범위 $(-1, 1)$ 을 벗어나는 경우가 많아 논리적 타당성이 떨어지는 문제점을 가지고 있다.

2.3 총변동 비교

지금까지 두 순서집단군에 피험자간 변동이 일정하다는 가정 하에서 피험자내 변동의 생동성을 통계적으로 평가하는 방법을 고찰하였다. 하지만 비교적 소수의 피험자로 이루어지는 생동성 시험에서 집단 간 피험자간 변동이 일정하다는 가정을 유지하기 힘든 경우가 많다. 이럴 경우 두 제제간의 변동의 생동성을 평가하는데 피험자간 변동까지 함께 고려해서 평가 하는 것이 타당하다.

2x2 교차설계법으로 이루어진 생동성 시험에서 i -번째 순서집단에서 k -번째 제제를 복용한 j -번째 피험자 생체이용률 벡터 (y_{ijT}, y_{ijR}) 는 모형 (2)를 따르며, 분산 공분산 행렬 Σ_T 를 가진다.

피험자간 변동과 피험자내 변동을 합한 총변동을 $\sigma_{TT}^2 = \sigma_{WT}^2 + \sigma_{BT}^2$ 와 $\sigma_{TR}^2 = \sigma_{WR}^2 + \sigma_{BR}^2$ 으로 정의하면, 대조약과 시험약의 변동의 생동성을 평가하는 가설은 다음과 같이 설정할 수 있다.

$$H_0: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \notin (\delta_1, \delta_2) \text{ vs. } H_1: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \in (\delta_1, \delta_2). \quad (12)$$

가설 (12)는 2.2장에서처럼 다음과 같은 두 단측검정 가설로 표현할 수 있다.

$$H_{01}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \leq \delta_1 \text{ vs. } H_{11}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} > \delta_1, \quad (13)$$

$$H_{11}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} \geq \delta_2 \text{ vs. } H_{11}: \frac{\sigma_{TT}^2}{\sigma_{TR}^2} < \delta_2. \quad (14)$$

가설 (13)과 (14)는 $\eta_1 = \sigma_{TT}^2 - \delta_1 \sigma_{TR}^2$ 와 $\eta_2 = \sigma_{TT}^2 - \delta_2 \sigma_{TR}^2$ 의 $(1-\alpha) \times 100\%$ 신뢰구간을 각각 추정해서 검정할 수 있다. 최근 Lee et al.(2003)는 제제간의 집단 생동성을 평가하는 문제에서 $(\mu_T - \mu_R)^2 + \eta_1$ 과 $(\mu_T - \mu_R)^2 + \eta_2$ 의 근사적인 신뢰구간을 구축하는 수정된 대표본 방법(modified large sample method)을 제안하였다. 본 연구논문에서는 Lee et al.가 제안한 방법에서 $(\mu_T - \mu_R)^2$ 부분을 제거하여 가설 (13)과 (14)를 검정할 수 있는 $(1-\alpha) \times 100\%$ 신뢰구간을 구축하는 방법을 제안한다.

$\eta_1 = \sigma_{TT}^2 - \delta_1 \sigma_{TR}^2$ 의 $(1-\alpha) \times 100\%$ 신뢰하한은 Lee et al.이 제안한 방법을 수정하여 다음과 같이 추정될 수 있다.

$$\hat{\eta}_L = \hat{\eta}_1 + \sqrt{\Delta_L}.$$

여기서 추정량 $\hat{\eta}_1 = \hat{\sigma}_{TT}^2 - \delta_1 \hat{\sigma}_{TR}^2$ 이고, 두 행렬 $\Theta_1 = \begin{pmatrix} 1 & 0 \\ 0 & -\delta_1 \end{pmatrix}$ 과

$\hat{\Sigma}_T = \begin{pmatrix} \hat{\sigma}_{TT}^2 & \hat{\sigma}_{TTR}^2 \\ \hat{\sigma}_{TTR}^2 & \hat{\sigma}_{TR}^2 \end{pmatrix}$ 의 곱행렬 $\Theta_1 \hat{\Sigma}_T$ 의 두 고유근 $\hat{\lambda}_{11} < 0 < \hat{\lambda}_{12}$ 에 대해

$$\Delta_L = \hat{\lambda}_{11}^2 \left(1 - \frac{n_1 + n_2 - 2}{\chi_{(1-\alpha, n_1 + n_2 - 2)}^2} \right)^2 + \hat{\lambda}_{12}^2 \left(1 - \frac{n_1 + n_2 - 2}{\chi_{(\alpha, n_1 + n_2 - 2)}^2} \right)^2 \text{이다. 단, } \chi^2(\alpha, n_1 + n_2 - 2)$$

는 자유도가 $n_1 + n_2 - 2$ 인 χ^2 -분포에서 상위 α -백분위수이다.

비슷하게 $\eta_2 = \sigma_{TT}^2 - \delta_2 \sigma_{TR}^2$ 의 $(1-\alpha) \times 100\%$ 신뢰상한을 추정할 수 있다.

$$\hat{\eta}_U = \hat{\eta}_2 + \sqrt{\Delta_U}.$$

여기서 $\hat{\eta}_2 = \hat{\sigma}_{TT}^2 - \delta_2 \hat{\sigma}_{TR}^2$ 이고, 두 행렬 $\Theta_2 = \begin{pmatrix} 1 & 0 \\ 0 & -\delta_2 \end{pmatrix}$ 과 $\hat{\Sigma}_T$ 의 곱행렬 $\Theta_2 \hat{\Sigma}_T$ 의 두

고유근 $\hat{\lambda}_{21} < 0 < \hat{\lambda}_{22}$ 에 대해 $\Delta_U = \hat{\lambda}_{21}^2 \left(\frac{n_1 + n_2 - 2}{\chi_{(\alpha, n_1 + n_2 - 2)}^2} - 1 \right)^2 + \hat{\lambda}_{22}^2 \left(\frac{n_1 + n_2 - 2}{\chi_{(1-\alpha, n_1 + n_2 - 2)}^2} - 1 \right)^2$

이다. 이렇게 계산된 두 신뢰구간이 $\hat{\eta}_L > 0$ 와 $\hat{\eta}_U < 0$ 을 만족할 때 유의수준 α 에서 가설 (12)의 대립가설을 채택하며 제제간의 총변동의 생동성을 인정한다.

최근 Chow et al.(2003)은 역시 집단 생동성을 평가하는 근사적인 통계적 방법을 제안하였다. Lee et al.(2003)처럼 Chow et al.의 방법을 변경하여 가설 (13)과 (14)를 검정하는 통계적 방법을 제안할 수 있다.

η_1 의 $(1-\alpha) \times 100\%$ 신뢰하한과 η_2 의 $(1-\alpha) \times 100\%$ 신뢰상한은 다음과 같이 추정된다.

$$\begin{aligned} \hat{\eta}_L^* &= \hat{\eta}_1 + t_{\alpha, n_1+n_2-2} \sqrt{V_1^*}, \\ \hat{\eta}_U^* &= \hat{\eta}_2 + t_{1-\alpha, n_1+n_2-2} \sqrt{V_2^*}. \end{aligned}$$

여기서 V_1^* 는 $\hat{\eta}_1$ 의 근사적인 분산으로 행렬 C_1 과 행렬 C_2 가 각각 $((Y_{1jT} - \bar{Y}_{1.T})^2, (Y_{1jR} - \bar{Y}_{1.R})^2), ((Y_{2jT} - \bar{Y}_{2.T})^2, (Y_{2jR} - \bar{Y}_{2.R})^2)$ 의 표본분산행렬이고, $C = \frac{(n_1-1)C_1}{(n_1+n_2-2)^2} + \frac{(n_2-1)C_2}{(n_1+n_2-2)^2}$ 일 때 $V_1^* = (1, -\delta_1) C (1, -\delta_1)'$ 로 정의된다. $\hat{\eta}_2$ 의 근사적인 분산 V_2^* 역시 비슷하게 정의된다. 위에서와 마찬가지로 추정된 신뢰구간이 $\hat{\eta}_L^* > 0$ 와 $\hat{\eta}_U^* < 0$ 을 만족할 때 유의수준 α 에서 가설 (12)의 대립가설을 채택하여 두 제제간의 총변동의 생동성을 인정한다.

2001년에 발표된 US FDA 생동성 시험 안내서에 보면 집단 생동성 평가 절차가 설명되어 있다. 현재 미국 기준은 집단 생동성을 평가하는 방법 중 전체적인 방법을 사용하고 있기 때문에 총분산을 따로 추론하지는 않지만 전체적인 방법을 추론하는 과정에서 로그 변환 기준으로 대조약과 시험약의 전체 변동의 차이가 0.02를 넘지 않아야 한다는 기준을 가지고 있다. 따라서 전체 변동성을 검정하는 기준으로 사용할 수 있을 것이다. 이를 가설로 표현해 보면 다음과 같다.

$$H_0 : \sigma_{TT}^2 - \sigma_{TR}^2 \geq 0.02 \text{ vs. } H_1 : \sigma_{TT}^2 - \sigma_{TR}^2 < 0.02. \tag{15}$$

가설 (15)는 $\delta_2 = 1$ 을 가진 가설 (14)로 표현할 수 있으며 위에서 설명한 Lee et al.(2003)의 변경된 방법이나 Chow et al.(2003)의 변경된 방법으로 신뢰구간을 구축한 후 0.02와 비교하여 가설을 검정할 수 있다.

3. 적용 사례와 논의

Liu & Chow(1992)에서 제안되었던 사례를 들어 지금까지 논의되었던 제제간의 변동성의 생동성을 평가하는 통계적 방법들을 논의하고자 한다. 24명의 건강한 피험자를 각 순서집단에 12명씩 랜덤하게 할당한 후 표준 2x2 교차실험을 수행하였다. 대조약은 5 ml (50mg/ml) 경구용 현탁액(oral suspension)이고, 시험약은 5개의 50 ml 용량의 정제(tablet)이다. [표 1]은 제제 투여시간 0-32시간에서 채취된 혈액 표본에서 측정된 AUC 값이다.

[표 1] 대조약과 시험약의 AUC(0-32)

순서집단	피험자 번호	기간	
		1	2
1 (TR)	2	74.825	37.350
	3	86.875	51.925
	7	81.675	72.175
	8	92.700	77.500
	9	50.450	71.875
	10	66.125	94.025
	13	122.450	124.975
	14	99.075	86.225
	17	86.350	95.925
	18	49.925	67.100
	21	42.700	59.425
	22	91.725	114.050
2 (RT)	1	74.675	73.675
	4	96.400	93.250
	5	101.950	102.125
	6	79.050	69.450
	11	79.050	69.025
	12	85.950	68.700
	15	69.725	59.425
	16	86.275	76.125
	19	112.675	114.875
	20	99.525	116.250
	23	89.425	64.175
	24	54.175	74.575

현재 KFDA 기준에서는 원자료의 로그변환 후에 제제간의 평균 생체이용률 차이 $\mu_T - \mu_R$ 의 90% 신뢰구간이 ($\log 0.8, \log 1.25$)안에 포함되면 두 제제간의 생동성을 인정하고 있다. [표 1]에서 계산된 $\mu_T - \mu_R$ 의 90% 신뢰구간은 $(-0.1284, 0.0716)$ 로 생동성 기준치 $(-0.2231, 0.2231)$ 안에 포함되므로 평균의 비교만으로 판단하는 두 제제간의 생동성은 인정된다고 할 수 있다.

두 제제간의 집단 생동성을 평가하는 목적으로 변동의 생동성을 평가한다고 하자. 생동성 시험이 잘 통제되어 피험자간 변동을 일정하다고 가정할 수 있다면 피험자내 변동의 동등성만 검정하면 된다. 생동성 기준 (δ_1, δ_2) 의 선택에는 여러 의견이 있지만 평균 생동성과 같이 $(0.8, 1.2)$ 를 기준으로 하는 것은 분산의 속성에 비추어 보아 너무 경직된 값이라는 것이 일반적인 논의이고, 문헌(Chow & Liu, 2000)을 보면 $(0.5, 1.5)$

의 값을 많이 사용하고 있어 본 사례 역시 그 값으로 분석하기로 한다.

[표 1]에서 주어진 자료로부터 분산분석표를 만들어 기간효과를 F-검정해보면 p-값이 0.6419로 통계적으로 기간효과가 존재하지 않음을 알 수 있다. 이로부터 피험자내 변동성의 생동성을 검정 (8)과 (9)를 이용해서 평가할 수 있다.

가설 (4)는 검정통계량 $t_L = 0.7382$ 으로 5% 유의수준 기각역 1.7171보다 작아 대립가설 H_{11} 을 기각할 수 없고, 가설 (5)는 검정통계량 $t_U = -0.5070$ 으로 역시 5% 유의수준 기각역 -1.7171 보다 커서 대립가설 H_{21} 을 기각할 수 없다. 이로부터 유의수준 5%에서 피험자간의 변동의 생동성은 인정될 수 없다.

Guilbaud(1993)의 방법으로 피험자내 변동의 생동성을 평가할 수 있다. 주어진 δ_1 과 δ_2 를 이용하면 생동성 기준 ($\psi_L = -0.3333$, $\psi_U = 0.2$)이 얻어진다. 식 (11)을 적용하여 ψ 의 90% 신뢰구간을 계산하면 $(-0.5480, 0.4767)$ 인데 이 구간은 생동성 기준을 만족하지 못하기 때문에 역시 피험자내 생동성을 인정할 수 없다.

만약 생동성 시험에서 피험자간 변동의 동등성에 대해 확신이 없다면 피험자내 변동에 피험자간 변동까지 포함해서 총변동의 생동성을 고려해 제제의 생동성을 판단해야 한다. 제제간의 총변동의 생동성을 평가하는 가설 (13)과 (14)를 검정하기 위해 Lee et al.(2003)이 제안한 집단 생동성 평가 방법을 수정한 방법을 사용할 수 있다. [표 1]의 자료로부터 계산된 $\hat{\lambda}_{11} = 0.06959$ 과 $\hat{\lambda}_{12} = -0.0324$ 을 이용하면 95% 신뢰하한 $\hat{\eta}_L = 0.0929$ 을 추정할 수 있고 이 값은 0보다 크기 때문에 생동성 가설 (13)의 대립가설을 채택할 수 있다. 가설 (14)를 검정하기 위해 계산된 $\hat{\lambda}_{21} = 0.0636$ 과 $\hat{\lambda}_{22} = -0.1062$ 를 이용해서 추정된 95% 신뢰상한 $\hat{\eta}_U = 0.0431$ 인데 이 값은 0보다 작지 않기 때문에 생동성 가설 (14)의 대립가설을 채택할 수 없다. 따라서 5% 유의수준에서 두 제제간 총변동의 생동성을 인정할 수 없다.

Chow et al.(2003)이 제안한 집단 생동성 평가 방법을 수정한 방법으로 역시 가설 (13)과 (14)를 검정할 수 있다. [표 1]로부터 계산된 행렬 $C = \begin{pmatrix} 0.00031, & -0.00004 \\ -0.00004 & 0.00056 \end{pmatrix}$ 와 $V_1^* = 0.00049$, $V_2^* = 0.00169$ 로부터 각각 95% 신뢰하한 $\hat{\eta}_L^* = -0.0047$ 과 95% 신뢰상한 $\hat{\eta}_U^* = -0.1132$ 이 추정되었다. $\hat{\eta}_L^*$ 값이 0보다 크지 않아 가설 (13)을 채택할 수 없어 $\hat{\eta}_U^*$ 이 가설 (14)의 대립가설을 채택한다 하더라도 두 제제간의 총변동성의 생동성을 인정할 수 없다. 이 두 가지 방법 중 어떤 것을 사용하는 것이 좋은 지에 대한 논의는 집단 생동성 평가의 통계적 기법을 비교한 Park et al.(2005)의 결과에서도 언급했듯이 어느 한 방법도 한 방법에 절대적이지 않아 당분간 두 가지 방법 모두 활용하는 것이 좋을 듯싶다.

현 US FDA에서 활용하고 있는 집단 생동성 기준을 활용한 방법으로 가설 (15)를 통해 제제간 총변동의 생동성을 평가해볼 수 있다. Lee et al.(2003)의 방법으로 가설 (15)를 평가해보면 95% 신뢰상한 $\hat{\eta}_U = 0.0551$ 이 추정되어 기준치 0.02보다 작지 않아 총변동의 생동성을 인정할 수 없다.

이상의 논의에서 보듯이 두 제제간의 평균 생동성을 평가하는 것으로 제제간의 생동성을 평가했을 때 인정될 수 있는 제제가 피험자간 혹은 피험자내 변동성을 함께 고려할 때 생동성을 인정할 수 없는 경우가 많이 발생한다. 두 제제간의 평균 생체이

용률만 비교해서 제네릭 의약품을 승인해주는 것보다는 변동의 생동성까지 포함해서 평가하는 것은 제네릭 의약품 품질 보증에 크게 기여할 것은 자명한 일이다. 제네릭 의약품의 품질을 보다 확실하게 보증하기 일은 크게는 국가 경제 및 복지에까지 크게 영향을 미치는 일이기 때문에 현재 많은 국가들이 집단 생동성 및 개인 생동성을 새로운 생동성의 기준으로 설정하고 있고, 우리나라도 조만간 집단 생동성을 제제간의 생동성 기준으로 추가할 것으로 예상되고 있다.

앞에서 논의했듯이 제제간의 집단 생동성은 두 가지 방향에서 평가될 수 있다. 첫 번째는 현재 US FDA에서 사용 중인 통합적인 방법으로 그 개념은 제제간의 평균과 총변동을 하나의 식 $(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)$ 으로 놓고 통계적으로 추론하는 것이다. 이러한 기준은 통계적 추론에 편리성을 제공해줄 수 있지만 총변동의 차감한 값 $(\sigma_{TT}^2 - \sigma_{TR}^2)$ 이 음수가 나올 경우 평균의 큰 차이에서도 집단 생동성을 인정할 수 있는 문제점이 있을 수 있다(Park et al., 2005). 이러한 단점을 극복하는 방법으로 평균의 차이 $\mu_T - \mu_R$ 과 총변동의 차이 $(\sigma_{TT}^2 - \sigma_{TR}^2)$ 를 분리해서 통계적으로 추론해서 판단하는 방법이 있다. 이 방법은 통합적인 방법처럼 그 의미가 모호해지는 경우는 없고, 통계적 평가 방법 역시 덜 복잡할 수 있다. 이러한 이유로 많은 국가에서 집단 생동성을 평가하는 분리적인 통계적 방법에 보다 관심이 많아지고 있다.

제제간의 집단 생동성을 평가하는 방법으로 분리적인 방법이 활용되었을 경우 제제간의 평균과 변동성을 동시에 검정하는 방법과 제제간의 평균을 먼저 평가하고 그 다음에 변동성 순차적으로 검정하는 방법이 있을 수 있다. 주어진 유의수준을 정확하게 유지하려면 제제간의 평균과 변동성을 동시에 검정하는 방법이 통계적으로 선호될 수 있으나 현실적인 여건을 감안한 순차적인 방법이 더 선호될 수 있다(Wellek, 2000). 이런 관점에서 본 연구 논문은 분리적인 방법으로 집단 생동성을 평가하는 분야에 변동의 생동성을 평가하는 통계적 이론 분야에 기여할 수 있을 것으로 생각된다. 앞으로 순차적인 방법으로 제제간의 평균과 변동성을 평가할 때 제1종 오류를 배분하는 문제, 제제간의 변동의 생동성 기준(δ_1, δ_2)의 적절한 선택 문제, 순차적인 방법으로 제제간의 집단 생동성을 평가할 때 적절한 피험자 수를 결정하는 문제 등은 분리적인 방법으로 집단 생동성을 평가하는데 추가로 연구되어야 할 분야이다.

참고문헌

1. Chow, S.C. and Liu, J.P. (2000). Design and Analysis of Bioavailability and Bioequivalence Studies. 2nd edition, Marcel Dekker, New York.
2. Chow, S.C., Shao, J., Wang, H. (2003). Statistical tests for population bioequivalence, *Statistica Sinica*, 13, 539-554.
3. FDA (2001). Guidance for industry on statistical approaches to establishing bioequivalence. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
4. Guildbaud, O. (1993). Exact inference about the within-subject variability in 2x2 crossover trials. *Journal of the American Statistical Association*, 88, 937-946.

5. Lee, Y., Wang, H., Chow, S.C. (2003). Comparing Variabilities in Clinical Research", *Encyclopedia of Biopharmaceutical Statistics*, 214-230, 2003
6. Liu, J.P. and Chow, S.C. (1992). On assessment of bioequivalence in variability of bioavailability studies. *Communications in Statistics, A-Theory and Methods*, 21, 2591-2607.
7. Liu, J.P. (1991). Bioequivalence and intrasubject variability. *Journal of Biopharmaceutical Statistics*, 1, 205-219.
8. Park, S., Lim, N., Lee, J., Kim, B. (2005). A comparative study of statistical methods for population bioequivalence in 2x2 crossover design. *Journal of Applied Statistics*, 18, 1, 159-171.
9. Schuirmann, D.J. (1987). A comparison of two one-sided tests procedures and the power approach for assessing the bioequivalence trials. *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657-680.
10. Wang, W. (1997). Optimal unbiased tests for equivalence in intrasubject variability. *Journal of the American Statistical Association*, 92, 1163-1170.
11. Wellek, S. (2000). On a reasonable disaggregate criterion of population bioequivalence admitting of resampling-free testing procedures. *Statistics in Medicine*, 19, 2755-2767.

[2007년 6월 접수, 2007년 7월 채택]