

생물학적 동등성 시험에서 추가시험의 일관성 검정에 대한 소고

하명호^a, 박상규^{1,b}

^a중앙대학교 대학원 통계학과, ^b중앙대학교 응용통계학과

요약

식품의약품안전청에서는 제제간의 생물학적 동등성의 입증이 실패했을 경우 추가시험을 허용하는 시험 기준을 2008년 7월부터 시행하고 있다. 생물학적 동등성 시험의 추가시험규정은 원시험에서의 동등성 평가가 불충분한 시험 예수 때문에 실패했을 때 별도의 재시험 시험계획서를 제출하지 않고 피험자수를 증가시켜 다시 시험하여 제제간의 동등성을 평가할 수 있는 근거를 마련한 점에 긍정적이다. 하지만 최근 추가시험의 규정에 따라 군 당 12명 이상의 피험자를 사용했을 경우 오히려 추가시험의 일관성 검정 규정을 만족하지 못하여 추가시험이 실패되는 경우가 자주 발생하고 있어 추가시험규정에 대한 불만이 고조되고 있다. 본 연구에서는 이러한 상황이 발생하는 통계적 원인을 살펴보고 현 기준을 일본 규정처럼 조금 더 일반화하게 되면 추가시험을 더 폭넓게 사용할 수 있는 것을 제안하고자 한다.

주요용어: 2×2 교차설계법, 생물학적 동등성, 유의성 검정, 추가시험의 일관성 검정.

1. 서론

생물학적 동등성 시험(이하 생동성)이란 제제학적으로 동등한 두 제제의 생물학적 동등성 입증을 위한 생체 내 시험으로, 두 제제의 생체이용률(bioavailability)을 통계학적으로 비교하는 시험이다. 즉, 두 제제를 각각 복용하고 얻은 생체이용률 지표 중 AUC(Area Under Concentration, 혈중농도-시간곡선 하면적)와 C_{max} (최고 혈중 농도)의 동등성을 통계적으로 평가하는 시험이다.

생동성 시험 기준은 1988년 10월 국립보건안전연구원 고시 제 1988-1호로 처음 제정된 후 국제적인 시험 기준의 변화에 따라 꾸준히 개정 고시되어 왔으며, 2007-65호에서는 Canada Department of Health (1992)에서 사용되고 있는 추가시험(add-on test) 규정을 추가하였다. 원시험에서 불충분한 피험자수 때문에 생동성 평가가 실패한 경우 별도의 시험계획서를 식품의약품안전청(이하 식약청)에 제출함 없이 피험자를 추가한 후 다시 시험하여 제제간의 생동성을 평가할 수 있기 때문에 생동성 시험기관에서는 추가시험 사용을 적극적으로 고려하고 있다.

하지만 최근 추가시험을 사용한 제제간의 생동성 평가가 의도된 대로 원활하게 이루어지지 않고 있다. 가장 큰 이유가 원시험과 추가시험을 합한 결과로 보면 제제간의 생동성을 충분히 입증하는데 반해 두 시험결과를 통합할 수 있는 전제 조건인 원시험과 추가시험간의 일관성 검정에서 일관성을 입증하지 못하는 경우가 자주 발생된다는 것이다. 현재 추가시험 규정에 의하면 추가시험에서 군당 12명 이상의 피험자를 대상으로 실험을 수행하는데 때로는 이 피험자 규모가 일관성 검정에 영향을 미치는 것으로 알려져 있다.

본 연구는 현재 식약청에서 규정하고 있는 생동성 추가시험의 일관성 검정의 통계적 특성과 그 의미를 살펴보고, 추가시험을 보다 폭넓게 활용하기 위해 추가시험 기준이 어떻게 완화되는 것이 바람직한지 살펴보고자 한다. 2장에서는 추가시험규정과 그 의미, 3장에서는 규정의 통계적 모형과 그

이 논문은 2009년 중앙대학교 연구장학기금 지원에 의한 것임.

¹ 교신저자: (156-756) 서울특별시 동작구 흑석동 221, 중앙대학교 응용통계학과, 교수. E-mail: spark@cau.ac.kr

미를, 4장에서는 모의실험을 통해 현 기준을 어떻게 보완하는 것이 추가시험을 더 폭넓게 사용하는 것 인지 논의하고자 한다.

2. 추가시험규정과 의미

2.1. 추가시험규정

식약청은 2008년 7월부터 개정·시행된 생동성 시험기준에서 추가시험에 관한 규정을 다음과 같이 발표하였다 (식품의약품안전청, 2007).

제 18조 2의(추가시험) 제 13조에 의거하여 수행한 시험예수로 생동성 입증이 안되는 경우에는 1회에 한하여 추가시험을 실시할 수 있으며, 그 결과는 선행된 결과와 총괄해서 최종 분석되어야 한다. 다만, 제 19조 규정에 의한 평가에 이용되기 위해서는 다음 각 호의 1을 모두 만족하여야 한다.

1. 본 시험과 동일한 시험계획서에 따라 시험.
2. 군당 최소 12명 이상.
3. 본시험과의 일관성(Consistency)을 통계적으로 입증(유의수준 0.05이하)하기 위하여 다음 각항의 1을 만족하여야 한다. 단, 통계적으로 타당성 있는 다른 방법으로 일관성을 검정할 수도 있다.
 - 가. 본 시험과 추가시험의 잔차평균제곱의 비(두 개 중 작은 값을 분모로 취함)의 값이 대응되는 자유도를 갖는 F -분포의 상위 5%값보다 작아야 한다.
 - 나. 본 시험과 추가시험 간의 제제효과의 교호작용(interaction)이 유의수준 0.05에서 존재하지 않아야 한다.
4. 추가시험 실시 여부를 시험계획서에 명기.

2.2. 일관성 검정의 의미

생동성 시험기준 제 18조 2의 3항에 제시되어 있는 일관성 검정은 추가시험을 통한 생동성 평가에서 두 시험 결과를 통합할 수 있는 조건이기 때문에 추가시험에서는 반드시 통과되어야 하는 전제로서의 의미가 있다. 하지만 일관성 검정을 자세하게 살펴보면 유의성 검정을 통해 유의하지 않으면 두 시험의 일관성이 확보된다는 것으로 규정하고 있음에 주목할 필요가 있다. 즉, 일관성 검정의 대립가설은 두 시험 간에 차이가 있다는 것인데 이 경우는 피험자수가 많으면 많을수록(통계적인 표현으로 검정력이 높은 경우) 대립가설이 지지될 확률이 높아진다. 따라서 추가시험의 피험자수의 규모가 클 경우 일관성 검정은 지지되기 어려운 속성이 있다는 것이다. 이러한 부분 때문에 현재 규정인 제 18조 2의 2항의 군당 12명 이상으로 수행하게 된 추가시험을 보다 폭넓게 사용하기 위해서 보다 완화된 규정이 필요하다고 생각된다.

3. 통계적 모형

3.1. 생물학적 동등성의 통계 모형

생동성시험 기준에 따르면 반감기가 긴 의약품은 제외하고는 표 1과 같은 표준적인 2×2 교차 설계를 원칙으로 하고 있으며, 표 1에 해당하는 통계적 모형은 (3.1)과 같다.

$$y_{ijk} = \mu + g_j + S_{l(j)} + p_k + \pi_{jk} + \epsilon_{jkl}, \quad (3.1)$$

$$S_{l(j)} \sim iid N(0, \sigma_s^2), \quad \epsilon_{jkl} \sim iid N(0, \sigma_{\epsilon_l}^2), \quad l = 1, 2, \dots, n_{1j}, \quad j, k = 1, 2,$$

표 1: 2×2 교차 설계

		기간	
		1	2
순서	1	R	T
	2	T	R

여기서 y_{jkl} 는 j -번째 순서, k -번째 기간, jk -번째 제제(11, 22는 R, 12, 21은 T)를 투여한 l -번째 피험자에서 얻은 AUC 혹은 C_{max} 등을 로그 변환한 값이며, μ 는 전체평균, g_j 는 j -번째 순서효과, p_k 은 k -번째 기간효과, π_{jk} 은 jk -번째 제제 효과이며, 식 (3.2)의 조건을 만족한다.

$$\sum_{j=1}^2 g_j = 0, \quad \sum_{k=1}^2 p_k = 0, \quad \sum_{j=1}^2 \sum_{k=1}^2 \pi_{jk} = 0, \quad \pi_{jk} = \begin{cases} \pi_R, & (j, k) = (1, 1), (2, 2), \\ \pi_T, & (j, k) = (1, 2), (2, 1). \end{cases} \quad (3.2)$$

또한 n_{1j} 는 j -번째 순서집단에 배정된 피험자 수이고, $S_{l(j)}$ 는 j -번째 순서집단에 투입된 l -번째 피험자에게 관측되는 피험자의 변동성을 나타내는 확률변수이고, ϵ_{jkl} 는 j -번째 순서에 k -번째 기간에 l -번째 피험자에게 관측되는 제제의 변동성을 나타내는 확률변수이다. $S_{l(j)}$ 와 ϵ_{jkl} 는 통계적으로 서로 독립을 가정하며, $S_{l(j)}$ 의 분산 σ_s^2 는 피험자에 의한 변동을, ϵ_{jkl} 의 분산 $\sigma_{e_l}^2$ 는 제제에 의한 변동을 나타낸다.

생동성시험 기준에서는 두 제제의 제제효과 차이 $\pi_T - \pi_R$ 의 90% 신뢰구간이 ($\log 0.8, \log 1.25$)에 포함되면 두 제제 간에 생동성을 인정한다. 제제효과 차이 $\pi = \pi_T - \pi_R$ 의 추정치와 π 의 90% 신뢰구간은 각각 다음과 같다.

$$\hat{\pi} = \hat{\pi}_T - \hat{\pi}_R = \frac{1}{2} (\bar{y}_{12} + \bar{y}_{21} - \bar{y}_{11} - \bar{y}_{22}), \quad (3.3)$$

$$\left(\hat{\pi} - t \left(\frac{\alpha}{2}, n_{11} + n_{12} - 2 \right) \sqrt{\frac{MS_{Intra}}{2} \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} \right)}, \hat{\pi} + t \left(\frac{\alpha}{2}, n_{11} + n_{12} - 2 \right) \sqrt{\frac{MS_{Intra}}{2} \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} \right)} \right), \quad (3.4)$$

여기서 $t(\alpha, n_{11} + n_{12} - 2)$ 는 자유도가 $n_{11} + n_{12} - 2$ 인 t -분포표에서 상위 α 에 해당하는 백분위수, MS_{Intra} 는 SS_{Intra} 를 그 것의 자유도 $n_{11} + n_{12} - 2$ 로 나눈 값으로 분산분석표에서 피험자 내 잔차의 평균 제곱합을 의미한다. 생동성 모형과 제제효과 추론에 관련된 보다 자세한 내용은 Chow와 Liu (2008)을 참고 바란다.

3.2. 추가시험에서의 통계 모형

식약청에서는 2×2 교차설계법을 통한 시험으로 두 제제간의 생동성을 입증하지 못했을 경우 한번의 추가시험을 수행할 수 있도록 하고 있다. 원시험과 추가시험의 모형은 (3.5)와 같다.

$$y_{ijkl} = \mu + \rho_i + g_{j(i)} + S_{l(ij)} + p_{k(i)} + \pi_{jk} + (\rho\pi)_{ijk} + \epsilon_{ijkl}, \quad (i, j, k = 1, 2, l = 1, 2, \dots, n_{ij}), \quad (3.5)$$

여기서 y_{ijkl} 은 i -번째 시험, j -번째 순서, k -번째 기간, jk -번째 제제(11, 22는 R, 12, 21은 T)를 투여한 l -번째 피험자에서 얻은 로그를 취한 생체 이용률 값이고, μ 는 전체 평균, p_i 는 i -번째 시험 효과, $g_{j(i)}$ 는 i -번째 연구에서 j -번째 순서 효과, $S_{l(ij)}$ 는 i -번째 시험에서 j -번째 순서의 l -번째 피험자 변동, $p_{k(i)}$ 은 i -번째 연구에서 k -번째 기간효과, π_{jk} 은 jk -번째 제제효과, $(\rho\pi)_{ijk}$ 는 i -번째 시험과 jk -번째 제제의 교호작용, ϵ_{ijkl} 은 제제변동을 의미한다. n_i 는 i -번째 시험에서의 2개의 군에 배정된 피험자 합계 ($n_{i1} + n_{i2}$)를 의미하며, 피험자변동 $S_{l(ij)}$ 는 $S_{l(ij)} \sim N(0, \sigma_s^2)$ 이며, 제제변동 ϵ_{ijkl} 는 $\epsilon_{ijkl} \sim N(0, \sigma_{e_l}^2)$ 이다. 피험자 변동과 제제변동은 역시 상호 독립으로 가정한다. 또한 시험효과(ρ), 순서효과(g), 기간효과(p), 제제효과

표 2: 모형 (3.5)의 분산분석표

요인	자유도	제곱합	$E(\text{평균제곱합})$
시험자 간 변동			
시험효과	1	SS_{Study}	$\frac{2n_1n_2}{n_1+n_2}(\rho_2 - \rho_1)^2 + 2\sigma_s^2 + \sigma_e^2$
순서효과	2	$SS_{Seq(Study)}$	$\sum_{i=1}^2 \frac{n_{i1}n_{i2}}{n_i} (g_{1(i)} - g_{2(i)})^2 + 2\sigma_s^2 + \sigma_e^2$
시험자간 변동	$n_1 + n_2 - 4$	SS_{Intra}	$2\sigma_s^2 + \sigma_e^2$
시험자 내 변동			
제제 효과	1	SS_{Drug}	$\frac{(\pi_R - \pi_T)^2}{1/8(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}) + \sigma_e^2}$
기간 효과	2	$SS_{Period(Study)}$	$\sum_{i=1}^2 \frac{n_{i1}n_{i2}}{n_i} (p_{2(i)} - p_{1(i)}^2) + \sigma_e^2$
시험*제제 효과	1	$SS_{Study*Drug}$	$\frac{(\rho\pi_1 - \rho\pi_2)^2}{1/8(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}) + \sigma_e^2}$
시험자 내 잔차	$n_1 + n_2 - 4$	SS_{Intra}	σ_e^2
총변동	$2(n_1 + n_2) - 1$		

(π), 시험*제제효과($\rho\pi$)는 식 (3.6)의 조건을 만족한다.

$$\sum_{j=1}^2 g_{j(i)} = 0, \quad \sum_{k=1}^2 p_{k(i)} = 0, \quad \sum_{j=1}^2 \sum_{k=1}^2 \pi_{jk} = 0, \quad \pi_{jk} = \begin{cases} \pi_R, & (j, k) = (1, 1), (2, 2), \\ \pi_T, & (j, k) = (1, 2), (2, 1), \end{cases}$$

$$\sum_{i=1}^2 \rho_i = 0, \quad \sum_{i=1}^2 \rho\pi_i = 0, \quad \rho\pi_{ijk} = \begin{cases} \rho\pi_1, & (i, j, k) = (1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1), \\ \rho\pi_2, & (i, j, k) = (1, 1, 2), (1, 2, 1), (2, 1, 1), (2, 2, 2). \end{cases}$$

생동성 추가 시험 모형 (3.5)에 대한 보다 자세한 통계적 추론에 대해서는 박상규 등 (2010)을 참고하기 바란다.

3.3. 일관성 검정

두 번 수행된 2×2 교차설계 시험을 통합하여 두 제제간의 생동성을 평가하기 이전에 각각 수행된 시험이 같은 조건에서 수행되었는가를 판단하기 위하여 일관성 검정을 수행해야하며, 일관성 검정은 다음 두 가지이다.

3.3.1. 실험 오차의 비교

일관성 검정에서 수행되어야 할 하나의 검정은 각각 수행된 시험의 실험오차를 유의수준 5%에서 비교하는 것이다. 만약 두 실험오차가 유의한 차이가 존재한다면 생동성 시험자료는 통합될 수 없으며 결국 시험은 실패한 것으로 한다.

원시험의 분산분석표를 통해 얻은 잔차의 평균 제곱합을 MSE_0 이라하고, 추가시험에서 얻은 잔차의 평균 제곱합을 MSE_a 라 하자. $MSE_1 \leq MSE_2$ 을 두 잔차의 평균 제곱합을 크기 순서대로 나열한 값이라하고, df_1, df_2 를 이에 해당되는 자유도라고 했을 때, 두 시험의 실험의 제제변동이 동등한지를 가설 $H_{10} : \sigma_{e1}^2 = \sigma_{e2}^2, H_{11} : \sigma_{e1}^2 \neq \sigma_{e2}^2$ 로 설정한 후, F -검정을 통해 검정할 수 있다.

$$\frac{MSE_2}{MSE_1} \sim F(df_2, df_1).$$

3.3.2. 연구와 제제효과의 교호작용 존재여부 검정

일관성 검증에서 수행되어야 할 두 번째 검증은 표 2에서 계산된 시험과 제제효과의 교호작용의 존재 여부를 가설 $H_{20} : \rho\pi_1 = \rho\pi_2, H_{21} : \rho\pi_1 \neq \rho\pi_2$ 로 설정한 후 F -검정하는 것이다. 만약 유의수준 5%에서 F -검정한 결과가 유의하다면 시험에 따라 제제효과가 다르다는 의미이고 따라서 두 시험자료는 통합될 수 없게 되어 생동성의 입증에 실패하게 된다.

일관성 검정의 두 단계 중 가항은 원시험과 추가시험의 오차변동의 유사성을 통해 두 시험이 유사한 시험이라는 점을 입증하라는 의미이고, 나항은 두 시험에서 제제효과가 서로 다르게 나타나지 않음을 통해 서로 다른 시점에서 수행된 두 시험이 통합될 수 있음을 입증하라는 의미로 볼 수 있다. 서로 다른 두 시험을 사후적으로 결합하여 한 시험처럼 해석할 수 있도록 이러한 전제를 검정하는 것은 바람직하다고 여겨지지만 문제는 유의성 검정을 통해 동일성을 검정하는 문제이다. 즉, 유의성 검정은 피험자 수가 많을수록 유의하게 나타날 수밖에 없기 때문에 추가시험의 피험자수를 계산할 때 일관성 검정을 통과하기 위해서는 유의성이 나타나지 않을 정도의 크지 않은 표본 크기가 필요하다.

생동성 시험을 승인받을 때 식약청에서는 적절한 피험자수를 통계적으로 결정하기를 요구한다. 추가시험의 경우도 적절한 검정력을 가진 피험자수를 계산하는 것이 바람직한데 문제는 군당 12명 이상이라는 규정이 때로는 지나치게 높은 규정일 수 있어 군당 원시험의 절반 이상의 규모로 수행한다는 보다 완화된 탄력적인 규정이 추가시험을 실시하는 목적에 더 적절할 수 있다. 다음 장에서는 현재 우리 기준과 우리보다 탄력적인 기준을 가지고 있는 Japan National Institute of Health (2006)을 고려한 기준을 모의실험을 통해 비교해보고자 한다.

A sufficient number of subjects for assessing bioequivalence should be included. If bioequivalence cannot be demonstrated because of an insufficient number, an add-on subject study can be performed using not less than half the number of subjects in the initial study.

- Japan National Institute of Health 생동성 규정 중

4. 모의실험

원 시험의 생동성 입증 비율과 원 시험에서 생동성 입증에 실패한 경우 적절한 추가시험 피험자수에 대해 알아보기 위해 모의실험을 실시하였다. SAS IML을 이용하여 식 (4.1)과 같이 이변량 정규 분포를 갖는 난수를 생성하였다.

$$Y_1 = \sigma_e U + \pi_R, \quad Y_2 = \sigma_e \rho U + \sigma_e \sqrt{1 - \rho^2} V + \pi_T, \quad (4.1)$$

여기서 $U, V \sim N(0, 1)$ 이며 $\pi_T = \pi_R + \delta$ 이다. 여기서 δ 는 두 제제간의 로그 값을 취한 평균의 차이에 해당한다. 우선 원시험에 대해 생동성 시험에서 가장 보편적으로 많이 사용되는 피험자 크기인 군당 12명을 기준으로 하였다. 만약 원시험에 대해 생동성을 입증하지 못한 경우 추가시험에서는 원시험의 절반수준인 6명부터 3명씩 증가시켜 18명까지 난수를 발생시키고, 일관성 검정과 추가시험에 대한 생동성 검정을 수행하였다. 이러한 과정을 10,000회 반복 실시하였다. 그 결과는 표 3, 표 4와 같다. 표 3, 표 4에서 원시험 합격수는 10,000회 반복 중에서 원 시험의 신뢰구간이 $(-0.2231, 0.2231)$ 에 들어온 수이며, 추가시험 수는 원시험에서 생동성을 입증하지 못하여 추가시험을 실시한 수이다. 실험오차 각각 비율은 추가시험에서 잔차평균제곱의 비의 값이 F -분포의 상위 5%값보다 큰 것의 비율이며, 교호작용 존재 비율은 추가시험에서 원 시험과 추가시험 간의 제제효과의 교호작용(interaction)이 유의수준 0.05에서 존재하는 비율이다. 추가시험 합격 비율은 실시된 추가시험에서 일관성이 입증되고, 원 시험

표 3: 모의 실험 결과($\delta = 0.00, 0.05$)

δ	σ_e	추가시험 피험자 수	원시험 합격수	추가 시험 수	실험오차 기각 비율(%)	교호작용 존재 비율(%)	추가시험 합격 비율(%)
0	0.2	6	10000	0	-	-	-
0	0.2	9	10000	0	-	-	-
0	0.2	12	10000	0	-	-	-
0	0.2	15	10000	0	-	-	-
0	0.2	18	10000	0	-	-	-
0	0.3	6	9999	1	0.0	0.0	100.0
0	0.3	9	10000	0	-	-	-
0	0.3	12	9997	3	0.0	100.0	0.0
0	0.3	15	9998	2	50.0	100.0	0.0
0	0.3	18	10000	0	-	-	-
0	0.4	6	9885	115	9.6	24.3	66.1
0	0.4	9	9884	116	6.0	31.9	64.7
0	0.4	12	9906	94	9.6	35.1	58.5
0	0.4	15	9892	108	21.3	49.1	40.7
0	0.4	18	9895	105	14.3	63.8	28.6
0	0.5	6	9120	880	9.4	15.0	66.7
0	0.5	9	9100	900	10.2	19.0	69.0
0	0.5	12	9141	859	10.0	22.9	67.4
0	0.5	15	9083	917	11.2	24.4	66.6
0	0.5	18	9091	909	11.2	26.5	65.0
0.05	0.2	6	10000	0	-	-	-
0.05	0.2	9	10000	0	-	-	-
0.05	0.2	12	10000	0	-	-	-
0.05	0.2	15	10000	0	-	-	-
0.05	0.2	18	10000	0	-	-	-
0.05	0.3	6	9972	28	7.1	32.1	67.9
0.05	0.3	9	9965	35	11.4	34.3	57.1
0.05	0.3	12	9960	40	12.5	35.0	57.5
0.05	0.3	15	9963	37	5.4	45.9	51.4
0.05	0.3	18	9956	44	11.4	68.2	29.5
0.05	0.4	6	9471	529	10.8	12.3	68.1
0.05	0.4	9	9477	523	9.9	17.8	68.8
0.05	0.4	12	9429	571	11.0	17.0	72.5
0.05	0.4	15	9461	539	11.1	26.0	64.9
0.05	0.4	18	9452	548	11.7	24.8	65.7
0.05	0.5	6	8212	1788	10.0	8.3	58.3
0.05	0.5	9	8239	1761	10.7	10.7	66.7
0.05	0.5	12	8266	1734	9.3	12.9	71.7
0.05	0.5	15	8229	1771	11.6	12.8	72.6
0.05	0.5	18	8230	1770	11.2	14.7	72.5

의 자료와 추가시험의 자료를 통합하여 제제간의 생동성을 평가하였을 때 생동성이 입증된 비율이다.

표 3, 표 4의 모의실험 결과는 실제 생동성 시험을 설계할 때 피험자 크기를 실험오차에 연관하여 계산하고 있기 때문에 해석상 주의를 요구한다. 즉, 임의로 군당 12명을 설정하여 모의실험을 했기 때문에 변동성이 큰 경우는 실제 생동성 원시험에서 보다 큰 피험자 크기로 실행했을 가능성이 높아 변동성이 크지 않은 경우만 해석하는 것이 의미가 있을 수 있다.

표 3, 표 4의 결과를 보면 제제간의 효과 차이(δ)가 0에 가깝고 제제변동을 의미하는 피험자내 변동(σ_e^2)이 작은 경우에는 원시험에서 생동성이 입증되는 비율이 높았다. 그러나 생동성이 입증되지 않

표 4: 모의 실험 결과($\delta = 0.10$)

δ	σ_e	추가시험 피험자 수	원시험 합격수	추가 시험 수	실험오차 기각 비율(%)	교호작용 존재 비율(%)	추가시험 합격 비율(%)
0.10	0.2	6	9982	18	5.6	22.2	77.8
0.10	0.2	9	9983	17	35.3	47.1	41.2
0.10	0.2	12	9981	19	26.3	47.4	47.4
0.10	0.2	15	9991	9	0.0	55.6	44.4
0.10	0.2	18	9975	25	12.0	84.0	16.0
0.10	0.3	6	9235	765	9.9	8.2	66.4
0.10	0.3	9	9284	716	11.3	12.7	71.6
0.10	0.3	12	9286	714	10.9	15.7	72.0
0.10	0.3	15	9246	754	12.2	18.8	69.4
0.10	0.3	18	9230	770	10.4	20.1	70.3
0.10	0.4	6	7389	2611	10.3	5.2	51.9
0.10	0.4	9	7462	2538	9.6	7.1	63.5
0.10	0.4	12	7484	2516	10.4	8.3	68.9
0.10	0.4	15	7413	2587	10.7	9.4	71.5
0.10	0.4	18	7494	2506	11.1	10.6	73.7
0.10	0.5	6	5839	4161	9.0	3.9	39.1
0.10	0.5	9	5715	4285	9.3	5.3	47.9
0.10	0.5	12	5832	4168	8.9	6.0	57.5
0.10	0.5	15	5780	4220	9.6	6.5	63.9
0.10	0.5	18	5819	4181	9.6	6.4	67.7

아 추가시험을 실시한 경우 추가시험의 피험자수를 크게 하면 추가시험의 합격 비율이 낮아지는 것으로 나타났으며, 이러한 현상이 발생한 원인으로는 일관성 검정의 검정력이 증가하여 일관성을 만족하지 않아 발생한 것으로 판단된다. 즉, 두 제제간의 평균의 차(δ)가 0일 때, 제제변동을 의미하는 표준편차(σ_e)가 0.4일 경우, δ 가 0.05일 때 σ_e 가 0.3일 경우, δ 가 0.10일 때 σ_e 가 0.2일 경우, 추가시험의 피험자수가 원시험의 절반일 경우가 가장 높은 일관성 검정 채택 결과를 보여준다.

최근 생동성 시험을 설계할 때 예비 실험 결과 혹은 유사실험에서 δ 와 σ_e 를 활용하여 피험자 크기를 통계적으로 결정하기 때문에 (박상규, 2007) 추가시험이 필요한 경우라 하더라도 생동성 기준에 크게 벗어나는 경우는 매우 드물다. 따라서 현 생동성 시험 규정에서처럼 추가시험을 무조건적으로 군당 최소 12명 기준을 유지한다면 시험기관에서 규정을 지켜 오히려 생동성을 확보 못하는 경우가 발생할 수 있다. 따라서 생동성 시험의 추가시험 규정은 Japan National Institute of Health (2006)처럼 원시험의 절반 이상을 피험자로 사용할 수 있도록 우리 기준을 완화한다면 추가시험을 보다 폭넓게 사용할 수 있을 것이다.

5. 결론

2008년 7월부터 식품의약품안전청은 생동성시험 기준에서 추가시험을 허용한 이래로 점차 생동성 시험기관에서 추가시험 사용 빈도가 높아지고 있다. 하지만 생동성 시험기관에서 추가시험 규정에 따라 시험을 수행했을 때 일관성 검정 때문에 제제간의 생동성 승인을 못하는 경우가 발생하고 있어 이 시점에서 일관성 검정에 관련한 규정을 통계적으로 조명해 볼 필요가 있다고 생각한다.

추가시험은 원시험에서 피험자수가 조금 모자라 제제간의 생동성을 입증 못했을 경우 추가시험을 통해 추가적인 피험자수를 확보해서 제제간의 생동성을 입증하는 방법이다. 통계적으로 부족한 검정력을 추가시험을 통해 확보하는 것이지만 문제는 추가시험이 원시험과 통합되는 전제조건인 일관성 검정은 유의성 검정이라 필요이상의 많은 피험자수는 오히려 추가시험을 실패로 유인한다. 즉, 추가시

험의 피험자수가 증가되면 일관성 검정의 검정력도 증가하기 때문에 일관성 검정이 실패할 수 있다는 것이다.

일반적으로 추가시험을 설계할 때 원시험 자료를 통해 제제간의 평균의 차와 제제 변동은 알 수 있기 때문에 적절한 추가시험의 피험자수를 추정해볼 수 있다. 하지만 현 추가시험 규정처럼 군당 12명 이상이라는 규정은 모의실험에서처럼 일관성 검정에서 문제를 야기할 수 있어 Japan National Institute of Health (2006)처럼 원시험의 절반 이상의 피험자라든지 보다 완화된 기준을 두고 탄력적으로 운영하는 것이 추가시험을 보다 폭넓게 활용하는 방안으로 생각된다.

참고 문헌

- 박상규 (2007). On sample size determination of bioequivalence trials, <한국데이터정보과학회지>, **18**, 365-373.
- 박상규, 남봉헌, 정윤노, 이재영, 정규진 (2010). 생물학적 동등성 시험의 추가시험에 대한 통계적 고찰, <한국통계학회논문집>, **17**, 107-115.
- 식품의약품안전청 (2007). 생물학적 동등성 시험기준(제 2007-65호).
- Canada Department of Health (1992). Guidance for Industry, Conduct and Analysis of Bioavailability and Bioequivalence Studies-Part A: Oral Dosage Formulations Used for Systemic Effects, Canada.
- Chow, S. C. and Liu, J. P. (2008). *Design and Analysis of Bioavailability and Bioequivalence Studies*, 3rd edition, Chapman and Hall CRC.
- Japan National Institute of Health, Division of Drugs (2006). Guideline for Bioequivalence Studies of Generic Products, Japan.

2011년 9월 접수; 2011년 9월 수정; 2011년 11월 채택

Some Remarks on Consistency Test of Add-on Test in Bioequivalence Trials

Myoung-Ho Ha^a, Sang-Gue Park^{1,b}

^aDepartment of Statistics, Graduate School of Chung-Ang University

^bDepartment of Applied Statistics, Chung-Ang University

Abstract

The newly revised bioequivalence guideline of Korea allows the add-on test since July 1, 2008 when the initial bioequivalence trial fails to show the equivalence of two drugs. The statistical model of the add-on test and its two stage testing procedures are discussed. Some statistical points of consistency test in the add-on test are considered and the issue on the sample size of add-on test is discussed. Some reasonable alternative like Japan's guideline for bioequivalence studies is recommended to secure the proper use of an add-on study through some simulation studies.

Keywords: 2×2 crossover design, add-on test of bioequivalence, consistency test, significance test.

This research was supported by the Chung-Ang University Research Scholarship Grants.

¹ Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, Seoul 156-756, Korea.
E-mail: spark@cau.ac.kr