

## Bioequivalence for $2 \times 2$ Crossover Experiment with Missing Observations<sup>1)</sup>

Sang-Gue Park<sup>2)</sup>, Young Wook Choi<sup>3)</sup>, Nam-Kyoo Lim<sup>4)</sup>, and Hyun-Chul Shin<sup>5)</sup>

### Abstract

Statistical models and their inferences in a bioequivalence trial are considered and studied when the missing observations occurred in  $2 \times 2$  crossover experiment. Patel(1985) suggested the approximate test procedures for carryover effect and drug effect in  $2 \times 2$  crossover design. Bootstrap methods are proposed to implement Patel's test to the bioequivalence trial. Patel's work and its bootstrap methods are discussed with the current KFDA rule through the simulation studies and the recommendation is made.

*Keywords* : bioequivalence, incomplete  $2 \times 2$  crossover design, maximum likelihood estimators, bootstrap confidence intervals

### 1. 서론

생물학적 동등성 시험은 화학적으로 동등한(chemically equivalent) 두 제제의 동일 유효성분을 비교할 때 제제간의 생체이용율(bioavailability)의 차이가 없고 치료 유효성분이 동일한 양과 속도로 전신순환 내에 도달하는지를 확인하는 시험이다.

생물학적 동등성 시험에서 일반적으로 채택하고 있는 교차설계법(crossover design)은 평행설계법(parallel design)에 비해서 통계적 추론이 비교적 복잡한 연구 방법이다. 특히  $2 \times 2$  교차설계법은 모형의 모수들이 서로 교락되어 있기 때문에 순서효과나 잔존효과 등의 일부 모수가 존재하지 않는다는 가정 하에서 나머지 모수들의 통계적 추론을 하게 된다(Chow and Liu, 2000). 따라서 이들 모수들이 존재하지 않는다고 가정할 때 이것이 타당성을 갖기 위해서는 실험이 계획되고 수행될 때 통계적 가정에 만족될 수 있는 실험이 되어야 한다. 예를 들어 프로토콜 상에 정의되어 있는 피험자 선택 기준에 의해서 피험자를 모집한 후에 확률화(randomization)를 반드시 거쳐야 한다는 것과 적당한 휴약기간(washout period)을 주어 잔존효과(carryover effect)가 존재하지 않도록 해야 한다는 것이 대표적인 것들이다.

- 
- 1) 이 연구는 2002년 한국과학재단 특정기초연구연구비(R01-2002-000-00076-0)에 의해서 수행되었음.
  - 2) Professor, Department of Statistics, Chung-Ang University, Seoul, Korea. E-mail: spark@cau.ac.kr
  - 3) Professor, College of Pharmacy, Chung-Ang University, Seoul, Korea.
  - 4) Full-time lecturer, Department of Information and Statistics, Daejeon University, Daejeon, Korea.  
Head of research center, R & D center, Pharmacodesign Co, Ltd., Daejeon, Korea.
  - 5) Post-Doc, School of Pharmacy, Rutgers University, Piscataway, New Jersey, USA.

생물학적 동등성 시험은 프로토콜에 따라서 수행되어야 한다. 하지만 피험자가 프로토콜을 위반하는 경우라든지 피험자가 시험을 포기하는 경우 등이 생길 수 있다. 예를 들면 프로토콜에서 금기시 하고 있는 사항을 위반했다든지 혹은 일부 피험자가 2번째 기간에 실험에 참여하지 못하게 되어 결측이 발생하는 경우가 바로 그것이다. 현재 식품 의약품 안전청에서 이런 경우에 추천하고 있는 방법은 일부 결측치가 존재하는 피험자의 자료를 모두 삭제하고 생물학적 동등성 시험을 평가하도록 하고 있다.  $2 \times 2$  교차설계에 의해 수행되는 생물학적 동등성 시험은 보통 20명에서 30명 사이의 피험자로 수행되는데 이런 결측치가 존재할 경우에 표본의 수가 작아 그 결과에 크게 영향을 미칠 가능성이 높다.

본 연구논문은  $2 \times 2$  교차설계법에서 생물학적 동등성 시험을 수행할 때 두 번째 기간에 피험자의 일부가 사정상 참여하지 못했을 때 생물학적 동등성의 통계적 분석에 관한 통계적 검정 절차에 관한 것이다. 통계적으로 결측치가 존재하는 문제는 매우 중요한 연구 주제로 연구되어 왔지만  $2 \times 2$  교차설계법에서는 결측치에 관한 문제가 그리 많이 다루어지지 않았다. Patel(1985)은  $2 \times 2$  교차설계법을 이용한 임상시험 결과를 분석하는데 결측치가 존재할 때 분석하는 방법을 처음으로 제시하였다. Patel이 제시한 결과는 생물학적 동등성 시험의 결과 분석에도 적용할 수 있지만 작은 피험자 수로 진행되는 생물학적 동등성 시험의 경우 대표본 근사식에 의존한 Patel의 결과를 직접 적용시키는 것에는 의문이 제기되어 왔다. 본 논문은 Patel의 검정을 확장한 결과가 그리 크지 않은 수의 표본에서 실시되는 생물학적 동등성 시험에서도 유효한 결과를 가지는지와 Patel의 통계량을 붓스트랩화(bootstrapping)할 때 유효성이 있는지를 고찰하고자 한다.

## 2. 통계적 방법론

생물학적 동등성 시험은 일반적으로  $2 \times 2$  교차설계법에 의해서 이루어진다. 실험에 참여하는  $n$ 명의 피험자를 확률화를 통해 대조약  $R$ 을 먼저 먹고 나중에 시험약  $T$ 를 먹는 첫 번째 순서에  $n_1$ 명을 할당하고, 시험약  $T$ 를 먼저 먹고 대조약  $R$ 을 나중에 먹는 두 번째 순서에  $n_2$ 를 할당한다. 실험을 수행하는 과정에서 첫 번째 기간에 실험을 성공적으로 끝마쳤던 피험자들 중 일부가 2번째 기간에 여러 가지 이유로 해서 실험에 참여하지 못했다 가정해보자. 즉, 첫 번째 기간에 참여했던 피험자의 수는  $n_1 + n_2 (= n)$ 이고 이 가운데 2번째 기간에 참여한 피험자의 수는  $m_1 + m_2 (= m)$ 이다.

행렬  $Y$ 를  $m \times 2$  차원의 결측없이 관찰된 관찰치 행렬이고, 벡터  $y_1^*$ 은 기간 1에서만 관찰된 관찰치 벡터이다. 이때 결측치를 포함하는 자료들은 완전하게 관찰된 자료들과 독립임을 가정한다.

$$Y = \begin{pmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \end{pmatrix}, \quad y_1^* = \begin{pmatrix} y_{11}^* \\ y_{12}^* \end{pmatrix}.$$

확률변수  $(y_{1k}, y_{2k})$ ,  $(i=1, \dots, m_k; k=1, 2)$ 는 서로 독립인 이변량 정규분포를 따른다고 하자. 즉,

$$(y_{1k}, y_{2k}) \sim IN((\mu_{1k}, \mu_{2k}), \Gamma).$$

<표 1> 결측치가 존재할 때 생물학적 동등성 시험 자료구조

		기간 (period)	
		1	2
순서 (sequence)	1	$y_{11} = \begin{pmatrix} y_{111} \\ y_{211} \\ \vdots \\ y_{m_1 11} \end{pmatrix}$	$y_{21} = \begin{pmatrix} y_{121} \\ y_{211} \\ \vdots \\ y_{m_1 21} \end{pmatrix}$
		wash-out	
		$y_{11}^* = \begin{pmatrix} y_{m_1+1, 11}^* \\ \vdots \\ y_{n_1 11}^* \end{pmatrix}$	missing
	2	$y_{12} = \begin{pmatrix} y_{112} \\ y_{212} \\ \vdots \\ y_{m_2 12} \end{pmatrix}$	$y_{22} = \begin{pmatrix} y_{122} \\ y_{222} \\ \vdots \\ y_{m_2 22} \end{pmatrix}$
	wash-out		
		$y_{12}^* = \begin{pmatrix} y_{m_2+1, 12}^* \\ \vdots \\ y_{n_2 12}^* \end{pmatrix}$	missing

생물학적 동등성 시험에서  $\mu_{jk}$ 의 통계적 모형과 분산 공분산 행렬  $\Gamma$ 는 일반적으로 다음과 같이 정의된다(Chow and Liu, 2000).

$$\mu_{jk} = \mu + P_j + F_{(j,k)} + C_{(j-1,k)}, \quad (j=1, 2; k=1, 2), \tag{1}$$

$$\Gamma = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{11} \end{pmatrix}. \tag{2}$$

모형 (1)에서  $P_j$ 는 기간효과를,  $F_{(j,k)}$ 는 제제효과를,  $C_{(j-1,k)}$ 는 잔존효과를 의미하고, 제제 효과와 잔존효과는 다음과 같이 다시 표현될 수 있다.

$$F_{(1,1)} = F_{(2,2)} = F_R; \quad F_{(2,1)} = F_{(1,2)} = F_T,$$

$$C_{(0,1)} = C_{(0,2)} = 0, \quad C_{(1,1)} = C_R, \quad C_{(1,2)} = C_T.$$

[표 1]과 같은 생물학적 동등성 시험에서 제제효과와 잔존효과 등 관심있는 모수들의 통계적 추론을 위해 확률벡터  $u$ 와  $v$ 를 다음과 같이 정의한다.

$$u' = (y_{11}', y_{12}', y_{11}^{*'}), \quad v' = (y_{21}', y_{22}')$$

식 (2)를 Patel(1985)에서 제시되었던 우도함수에 적용하여 다시 계산하면 다음과 같은 모수들의 최우추정량들을 얻을 수 있다.

$$\hat{\mu}_{1k} = \bar{y}_{1k}^{(n_k)}, \quad \hat{\mu}_{2k} = \bar{y}_{2k}^{(m_k)} + \hat{\rho} \left( \hat{\mu}_{1k} - \bar{y}_{1k}^{(m_k)} \right),$$

$$\begin{aligned} \bar{y}_{jk}^{(c)} &= \sum_{i=1}^c y_{ijk}/c, \quad s_{uv} = \sum_{k=1}^2 \sum_{i=1}^{m_k} (y_{iuk} - \bar{y}_{uk}^{(m_k)})(y_{ivk} - \bar{y}_{vk}^{(m_k)})/m, \\ &\quad (c = n_k \text{ or } m_k; u, v = 1, 2), \\ \hat{\lambda}_{11} &= \frac{n\hat{\lambda}_a + m\hat{\lambda}_b/(1-\hat{\rho}^2)}{n+m}, \quad \hat{\rho} = 2s_{12}/(s_{11} + s_{22}), \\ \hat{\lambda}_a &= \sum_{k=1}^2 \sum_{i=1}^{n_k} (y_{i1k} - \hat{\mu}_{1k})^2/n, \quad \hat{\lambda}_b^2 = \sum_{k=1}^2 \sum_{i=1}^{m_k} \{(y_{i2k} - \hat{\mu}_{2k}) - \hat{\rho}(y_{i1k} - \hat{\mu}_{1k})\}^2/m, \\ \hat{\lambda}_{12} &= \frac{\hat{\rho}}{1-\hat{\rho}^2} \hat{\lambda}_b. \end{aligned}$$

여기서  $\hat{\rho}$ 은 Morrison(1973)에 의해서 제시된 추정량이고,  $(\hat{\mu}_{1k}, \hat{\mu}_{2k})$ 의 근사 분산 공분산 행렬  $\Gamma_k$ 는 다음과 같이 얻어진다.

$$\Gamma_k = \begin{pmatrix} \lambda_{11}/n_k & \lambda_{12}/n_k \\ \lambda_{12}/n_k & \lambda_{11}[(1-\rho^2)/m_k + \rho^2/n_k] \end{pmatrix}. \tag{3}$$

### 3. 통계적 가설 검정과 신뢰구간

생물학적 동등성 시험 결과를 분석할 때 가장 관심의 대상이 되는 것은 제제간의 생물학적 동등성을 입증하는 것이지만 2x2 교차설계법에서는 제제효과의 통계적 추론을 다루기 전에 제제효과와 교락되어 있는 잔존효과에 대한 통계적 추론이 선행되어야 한다. 즉, 2x2 교차설계법을 이용해서 두 제제간에 생물학적 동등성을 판단하고자 할 때 그 판단이 유효성을 가지려면 먼저 실험에서 잔존효과가 존재하지 않음을 입증해야 한다. 잔존효과에 대한 가설은 다음과 같다.

$$H_{01}: C_R - C_T = 0. \tag{4}$$

그리고 위 가설 (4)를 다시 표현하면 다음과 같다.

$$H_{01}: \theta = (\mu_{11} + \mu_{21}) - (\mu_{12} + \mu_{22}) = 0. \tag{5}$$

Patel(1985)은 가설 (5)에 대한  $\theta$ 의 최대우도 추정량에 기초한 검정통계량을 제안하였다.

$$Z_1 = \frac{\hat{\mu}_{11} + \hat{\mu}_{21} - \hat{\mu}_{12} - \hat{\mu}_{22}}{\sqrt{\sum_{k=1}^2 (\hat{\Gamma}_{11k} + 2\hat{\Gamma}_{12k} + \hat{\Gamma}_{22k})}}.$$

여기서  $\hat{\Gamma}_{11k}, \hat{\Gamma}_{12k}$ 와  $\hat{\Gamma}_{22k}$ 는 식 (3)에 주어진  $(\hat{\mu}_{1k}, \hat{\mu}_{2k})$ 의 근사 분산 공분산 행렬  $\Gamma_k$ 의 원소들의 일치추정치이다. 통계량  $Z_1$ 은 귀무가설 하에서  $m_1$ 과  $m_2$ 가 충분히 클 때, 근사적인 정규분포를 따르게 된다. Patel은 소표본일 경우  $(n+m-5)/2$ 인 자유도를 이용해서  $t$ -분포를 사용하기를 제안하였다. 이 방법은 상당히 임의적이긴 하지만 만약  $n = m$ 이면 통계량  $Z_1$ 은 자유도  $n-2$ 인 정확한  $t$ -분포를 따르기도 하고, Patel이 제시한 모의실험 결과를 보면  $n$ 과  $m$ 이 그리 차이가 나지 않는다면 비교적 정확한 것으로 보인다.

교차설계법에서 두 제제간의 잔존효과가 없다는 것이 판단되면 그 다음으로 중요한 것이 제제효과를 판단하는 것이다. 제제효과에 대한 가설은 다음과 같다.

$$H_{02}: F_R - F_T = 0 \tag{6}$$

위 가설 (6)을 다시 정리하면

$$H_{02}: \Delta = (\mu_{11} + \mu_{22}) - (\mu_{12} + \mu_{21}) = 0 \tag{7}$$

와 같이 표현할 수 있고 Patel(1985)은 가설 (7)에 대한 검정통계량을 다음과 같이 제안하였다.

$$Z_2 = \frac{\hat{\mu}_{11} + \hat{\mu}_{22} - \hat{\mu}_{12} - \hat{\mu}_{21}}{\sqrt{\sum_{k=1}^2 (\hat{\Gamma}_{11k} - 2\hat{\Gamma}_{12k} + \hat{\Gamma}_{22k})}}$$

통계량  $Z_2$  역시 귀무가설 하에서 표본의 크기가 클 때 근사적으로 정규분포를 따르게 되지만 표본의 크기가 그리 크지 않을 경우 Patel은 Morrison(1973)의 결과를 이용하여 자유도가  $m - 2$  인  $t$ -분포를 사용하기를 제안하였다.

Patel의 모의실험 결과를 보면 잔존효과에 대한 검정이나 제제효과에 대한 검정에 Patel의 제안을 사용하여  $t$ -분포를 이용하면 주어진 유의수준을 비교적 잘 유지하고 있음을 볼 수 있지만 기간 1과 기간 2의 상관계수  $\rho$ 가 증가하면 추정된 유의수준이 감소하는 경향이 나타나고 있다는 점은 유의해야 할 일이다. 또한 이것은 보다 다양한 조건 하에서 Patel의 제안이 유효한지를 이론적 혹은 모의실험으로 확인해야할 필요성이 제기된다 하겠다.

지금까지 언급한 Patel의 검정은 일반적인 임상시험에서 처리효과의 유효성을 검정하기 위해서  $2 \times 2$  교차설계법을 활용할 때 적용할 수 있는 방법이다. 제제간의 생물학적 동등성을 판단할 때는 Patel의 검정을 그대로 사용하지 않는다. Westlake(1976)과 Metzler(1974)는 제제간의 생물학적 동등성을 평가하는데 신뢰구간을 이용한 방법이 적당하다고 밝혔고, 현재 식품의약품안전청에서도 신뢰구간 방법에 의한 생물학적 동등성 여부를 판단할 것을 규정하고 있다. 따라서 결측자료가 존재할 때 Patel의 방법을 현재 식품의약품안전청에서 규정하는 방법처럼 사용하려면 다음과 같이 수정해서 사용할 수 있을 것이다. 실험에서 얻은 원자료에 로그를 취한 후

(i) Patel의 통계량  $Z_1$ 을 계산해서 Patel의 검정통계량을 이용해서 유의수준 5%에서 잔존효과 존재 여부를 검정한다.

(ii) 만약 잔존효과가 존재하지 않는다면 Patel의 통계량  $Z_2$ 를 이용해서 제제효과에 대한 90% 신뢰구간을 계산할 수 있다. 이렇게 구성된 제제효과의 90% 신뢰구간을  $\log(0.8)$  과  $\log(1.25)$ 와 비교하여 제제간의 생물학적 동등성을 판단하게 된다.

생물학적 동등성 시험에서 잔존효과가 존재하는 경우는 그리 많지 않다. 이는 대부분의 경우 제제들의 반감기(half life)를 계산해서 충분히 휴약기간을 두기 때문이다. 만약 잔존효과가 존재한다면  $2 \times 2$  교차설계법에서는 제제효과와 잔존효과를 분리할 수 없기 때문에 두 번째 기간에 얻어진 자료를 삭제하고 분석을 해야한다. 따라서 첫 번째 기간에서 얻어진 자료만을 이용해서 제제간의 생물학적 동등성을 확인할 수 있지만 이 경우는 일반적으로 제제간의 생물학적 동등성을 입증하는데 실패한 것으로 보는 경우가 많다. 왜냐하면 첫 번째 기간만으로 생물학적 동등성을 입증한다고 해도 표본의 수가 감소하여 추론의 검정력이 매우 떨어지게 되기 때문에 그 결과를 채택하기가 쉽지 않다.

제제간의 생물학적 동등성을 판단하는 방법으로는 Schuirmann(1987)의 TOST(two one-sided test)를 사용할 수 있지만 이 방법은 위에서 언급한 신뢰구간 방법과 동일한 방법임이 알려져 있

다(Berger and Hsu, 1996). 제제간의 생물학적 동등성을 판단하는데 사용되는 로그 변환은 통계적인 고려에 의한 것으로 보는 시각이 많다. 즉, 대조약과 시험약간의 동등성을 입증하는데 사용되는 검정통계량이 두 제제간의 효과의 비율로 구성되어 있고 이 비율이 0.8과 1.25 사이에 존재하는지 여부를 따지게 되는데 비율 통계량의 분포가 점근적인 분포로 존재하므로 특히 소표본에 의존하는 생물학적 동등성 시험에서 통계량의 확률계산이 부정확할 수 있다. 하지만 자료에 로그를 취한 후 정규성을 가정하게 되면 비율 통계량을 차이 통계량으로 변환시킬 수 있고 차이 통계량에 대한 확률분포는  $t$ -분포를 따르기 때문에 정확한 확률 계산을 할 수 있다는 장점이다.

미국이나 한국의 경우 제제간의 생물학적 동등성 판단 기준으로 제제간의 차이가 0.8과 1.25로 사용하지만 유럽공동체의 경우는 0.7과 1.43를 사용해서 비교적 제제간의 생물학적 동등성 판단을 다른 지역보다 용이하게 허가해주고 있다.

#### 4. 붓스트랩 신뢰구간 방법

Patel(1985)의 검정에 기초한 두 제제간의 생물학적 동등성 판단 방법이 근사적인 방법이고 소표본일 경우에 이론적으로 정확한 방법이 아니기 때문에 보다 이론적인 접근 방법이 모색되어야 한다. 하지만 소표본의 경우에 통계량의 정확한 분포를 찾는 것이 한계가 있기 때문에 그에 대한 대안으로 붓스트랩(Bootstrap) 방법을 활용할 수 있다.

붓스트랩 방법은 Efron(1979)이 그 이론적 타당성을 제안한 이래로 소표본에서 통계량의 분포를 근사시킬 때 많은 분야에서 활용되고 있다. 특히 결측치를 포함하고 있는 자료의 분석의 경우 이론적 접근의 복잡성 때문에 붓스트랩 방법이 많이 활용되고 있고 그 타당성 역시 인정되고 있다. 제제간의 생물학적 동등성 시험에서도 Chow(1990) 이래로 붓스트랩 방법이 이용되고 있다.

결측치가 존재하는 경우 분석에서 붓스트랩 방법을 사용하는 경우는 크게 두 가지가 있다. 첫 번째 방법은 결측치를 붓스트랩 방법으로 대체(imputation)한 후에 추정량을 계산하는 방법이고, 두 번째 방법은 결측치가 발생한 자료와 그렇지 않은 자료를 둘로 구분해서 각각 붓스트랩 방법을 적용한 후 추정량을 계산하는 방법이다(Bello 1994; Efron 1994; Shao and Sitter 1996). 건강한 피험자를 대상으로 하는 생물학적 동등성 시험에서는 자료의 수도 적고 피험자간의 차이도 어느 정도 존재하며 기간 효과도 어느 정도 존재할 가능성이 있어 결측치를 대체하는 첫 번째 방법보다는 두 번째 방법이 논란의 여지를 방지할 것으로 보인다.

생물학적 동등성 시험에서 결측치가 있을 때 붓스트랩 방법을 적용하는 것은 Bello(1994)가 제안한 비례 추출 붓스트랩 방법(proportional bootstrap method)을 사용하는 것이 타당해 보인다. 즉, 먼저 주어진 표본에서 각 순서에서 결측치가 존재하지 않는 부분과 존재하는 부분으로 나눈 후 각 부분으로부터 반복을 허용하고 재추출(resampling)을 한다. 첫 번째 기간에서의 데이터가 선택되면 두 번째 기간의 데이터는 자동적으로 선택되는 것으로 한다. 따라서 붓스트랩 표본의 경우에도 표본에서 결측치의 비율은 원래 데이터에서의 결측치 비율과 항상 동일하게 된다. 이렇게 얻어진 표본을 대상으로 통계량  $\hat{\Delta}$ 를 계산하고 이를  $\hat{\Delta}^1$ 으로 놓는다. 이런 과정을  $B$ 번 반복해서  $B$ 개의 통계량  $\hat{\Delta}^1, \hat{\Delta}^2, \dots, \hat{\Delta}^B$ 을 얻는다. 만약  $\Delta$ 의  $100 \times (1 - 2\alpha)\%$  신뢰구간을 구한다면  $B$ 개의  $\hat{\Delta}$ 값들을 순서대로 놓고  $\hat{\Delta}^{(B \times \alpha)}, \hat{\Delta}^{(B \times (1 - \alpha))}$ 을 계산하여 제제효과의 신뢰구간을 계산할 수 있다. 이러한 방법을 백분위수(percentile) 방법이라 한다. 이 방법은 단순하지만 가장

많이 활용되는 방법이다.  $B$ 번을 몇 번으로 해야 하는가는 문제에 따라 조금씩 다르긴 하지만 최소 1000번 이상해야 하는 것으로 알려져 있다.

붓스트랩 신뢰구간을 계산하는 방법으로 백분위수 방법의 편의를 수정하고 가속화한(biased corrected and accelerated :  $BC_a$ ) 붓스트랩 방법을 사용하기도 하는데 이를 사용하면 붓스트랩 표본이 한쪽으로 치우쳐진 경우에 포함확률을 잘 유지하는 것으로 알려져 있다(Efron and Tibshirani, 1993).  $BC_a$  방법은 백분위수 방법이 신뢰구간을 계산할 때 양쪽에 같은 확률을 부여하는 것을 수정하여 분포의 치우침에 따라 조정할 수 있도록 해 보다 신뢰수준에 적합한 신뢰구간을 찾을 수 있도록 하는 방법이다. 만약  $\Delta$ 의  $100 \times (1 - 2\alpha)\%$  신뢰구간을 계산한다면  $B$ 개의  $\hat{\Delta}$  값들을 순서대로 놓고  $\hat{\Delta}^{(B\alpha_1)}$ ,  $\hat{\Delta}^{(B\alpha_2)}$  을 계산하여 제제효과와 근사적인 신뢰구간을 계산할 수 있다. 여기서  $\alpha_1$ 과  $\alpha_2$ 는 다음과 같이 계산될 수 있다.

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z_{(\alpha)})}\right), \quad \alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z_{(1-\alpha)})}\right),$$

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\text{No. of } \{\hat{\Delta}^i < \hat{\Delta}\}}{B}\right), \quad \hat{a} = \frac{\sum_{i=1}^n (\hat{\Delta}_{(\cdot)} - \hat{\Delta}_{(i)})^3}{6\left\{\sum_{i=1}^n (\hat{\Delta}_{(\cdot)} - \hat{\Delta}_{(i)})^2\right\}^{3/2}},$$

$z_{(\alpha)}$ 는 표준정규분포의  $100 \times \alpha$  백분위수를 의미하고,  $\hat{\Delta}_{(i)}$ 는 원 표본에서  $i$ 번째 관측치를 제거하고 계산한 jackknife 추정치이고  $\hat{\Delta}_{(\cdot)} = \sum_{i=1}^n \hat{\Delta}_{(i)} / n$ 이다. 만약  $\hat{a}$ 와  $\hat{z}_0$ 가 0이라면  $BC_a$  방법과 백분위수 방법은 동일하게 된다.

$BC_a$  방법은 주어진 붓스트랩 표본이 치우쳐진 분포를 가질 때 주어진 신뢰수준을 잘 유지할 수 있는 방법으로 특히 표본의 수가 그리 크지 않게 실행되는 생물학적 동등성 시험에서 제제간의 생물학적 동등성을 판단하는데 유용하게 사용될 수 있을 것으로 판단된다. Czado and Munk(2001)은 제제간의 모집단 생물학적 동등성을 비모수적인 방법으로 평가하는데 붓스트랩 방법을 이용하였고 매우 다양한 상황에서 백분위수 방법과  $BC_a$  방법을 비교하고 두 방법론의 유효성을 논의하였다.

### 5. 모의 실험

지금까지 논의된 방법론과 본 논문에서 제시된 붓스트랩 접근 방법의 유효성을 비교하기 위해 다음과 같은 환경에서 모의실험을 실시한다. Patel(1985)이 제안한 모형을 생물학적 동등성 모형으로 변환시키면 다음과 같은 이변량 정규분포의 평균과 분산-공분산 행렬을 가정할 수 있다.

$$(y_{1k}, y_{2k}) \sim N\left((\mu_{1k}, \mu_{2k}), \begin{pmatrix} \lambda_{11} & \rho\lambda_{11} \\ \rho\lambda_{11} & \lambda_{11} \end{pmatrix}\right),$$

여기서  $\rho$  는 두 제제간의 상관계수를 의미한다.

모형 (1)에서 기간효과  $P_j$ 나 잔존효과  $C_{(j-1, k)}$ 는 존재하지 않는 것으로 가정하였고 비교를

위하여 제제 효과  $\mu + F_R$  값은 100으로 고정하였다. 두 제제 평균의 비율  $(\mu + F_T)/(\mu + F_R)$ 을 0.7부터 1.3까지 변화를 주었고, 상관계수는  $\rho = 0.2, 0.5, 0.8$ 로 그리고 변동계수(CV)는 10%와 20%로 설정하였다. 또한 각 순서의 표본의 수는  $n_1 = n_2 = 9, 12, 15$ 로 하였고 결측치를 가지는 데이터의 수는 현실을 고려해서 1, 2로 하였다.

이변량 정규분포로부터 생성된 난수에 대해서 논의된 네 가지 방법을 이용해 제제간의 90% 신뢰구간을 계산하였다. 즉, 결측치를 버리고 계산된 신뢰구간(GM), Patel의 통계량으로부터 계산된 신뢰구간(PM), 백분위수 붓스트랩 신뢰구간(PB), 그리고 편의를 수정하고 가속화한 붓스트랩 신뢰구간(BC)을 각각 구한 후에 그 신뢰구간이 생물학적 동등성 기준 구간인 (0.8, 1.25) 안에 포함되는지를 확인하였다. 모의실험 반복은 1000번을 기준으로 하였으며 붓스트랩 표본을 얻을 때 반복은 1000번으로 설정하였다.

## 6. 비교 분석

Grizzle(1965)의 방법은 결측치를 포함하고 있는 피험자의 자료를 제거하고 분석하는 현재 식품의약품 안전청에서 채택하고 있는 규정이다. 모의실험 결과는 표본의 크기에 따라 전체 표본의 크기가 각 순서에 9명씩 18명인 경우를 [표 2]에, 각 순서에 12명씩 24명인 경우가 [표 3]에, 각 순서에 15명씩 30명인 경우를 [표 4]에 정리하였다. 결측치의 수는 현실적인 해석을 부여하기 위해 각 순서에 1명 혹은 2명인 경우를 가정하였다.

방법론 비교의 판단기준은 제제와 시험 제제의 비율이 0.8과 1.25사이에서는 실제 제제들이 생물학적으로 동등하다고 해석되는 경우이므로 이때는 동등하다고 판단하는 횟수가 더 큰 것이, 그렇지 않은 범위에서는 생물학적으로 동등하다고 판단하는 횟수가 작은 것을 좋은 방법으로 간주한다.

제시된 세 개의 표를 이용하여 Patel의 방법과 Grizzle의 방법을 비교해 보면 큰 차이는 나지 않고 있지만 대체적으로 Patel의 방법이 Grizzle 방법보다는 실제 동등성인 경우에 더 많은 동등성 판정 비율을 보여주고 있다.

한편 제제 간 비율이 0.7과 1.3인 경우에는 두 제제가 생물학적으로 동등하지 않다고 판단하는 경우를 의미한다. 표를 통해서 판단해보면 작은 차이지만 Patel의 방법이 Grizzle의 방법보다 평균적으로 적게 생물학적 동등성을 판단하고 있음을 볼 수 있다. 이런 경우는 Patel의 방법이 Grizzle의 방법보다 정확한 의사결정을 잘 내린다고 판단할 수 있다. 이 결과는 식품의약품안전청에서 제안하고 있는 Grizzle방법을 Patel의 방법으로 대체해야 함을 의미한다.

붓스트랩 방법은 백분위수 방법과 편의를 수정하고 가속화한 방법을 수행했다. 이 두 가지 방법을 Patel의 방법과 비교해보면 실제 동등성인 경우에 Patel의 방법보다 동등하다고 판단하는 횟수가 더 많음을 볼 수 있다. 그러나 생물학적으로 동등하지 않은 경우 (비율이 0.7과 1.3의 경우)에서는 동등하다고 판단하는 횟수가 붓스트랩 방법이 Patel의 방법보다도 그리 큰 차이는 아니지만 생물학적 동등성을 더 많이 판단하는 것으로 나타나고 있다. 특히 표본의 크기가 9개로 표본의 크기가 아주 작을 때 이런 현상이 나타나고 있음을 볼 수 있지만 현실적으로 생물학적 동등성 시험에서 사용하고 있는 12명 이상의 경우에는 이런 현상이 급속도로 감소하고 있음을 볼 수 있다. 이



렇게 볼 때 붓스트랩 방법이 다른 방법들에 비해 생물학적 동등성을 판단하는데 가장 효과적이라고 볼 수 있다. 수행된 두 가지 붓스트랩 방법의 결과를 비교해보면 두 결과가 거의 유사함을 볼 수 있다. 특히 표본의 크기가 커지면서 거의 같은 결과를 주고 있다. 따라서 분포의 치우침이 크게 나타날 것으로 예측되는 경우를 제외하고는 백분위수 붓스트랩 방법을 사용하는 것이 바람직해 보인다.

<표 2> 모의실험(n=9)에서 생물학적 동등성 판단 비율

sample size	n <sub>1</sub> =n <sub>2</sub> =9 m <sub>1</sub> =8 m <sub>2</sub> =8				n <sub>1</sub> =n <sub>2</sub> =9 m <sub>1</sub> =7 m <sub>2</sub> =7				n <sub>1</sub> =n <sub>2</sub> =9 m <sub>1</sub> =8 m <sub>2</sub> =7						
	CV	Rho	Ratio	GM	PM	BP	BC	GM	PM	BP	BC	GM	PM	BP	BC
0.2	0.7			0	0	0	0	0	0	0	0	0	0	0	0
	0.8			36	36	58	58	32	33	60	60	36	32	54	54
	0.9			870	889	936	936	813	849	910	910	846	865	912	912
	1			1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1			988	990	995	995	971	979	992	992	976	982	990	990
	1.2			379	386	466	467	363	374	463	470	378	388	470	473
1.3			4	1	3	3	7	5	14	14	6	4	8	9	
0.1	0.7			0	0	0	0	0	0	0	0	0	0	0	0
	0.8			39	37	60	60	29	33	58	58	33	30	48	48
	0.9			858	881	922	922	837	843	890	890	840	853	907	907
	1			1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1			987	986	995	995	977	980	994	994	972	975	988	989
	1.2			387	395	476	478	360	376	461	466	366	394	470	473
1.3			6	5	6	6	5	3	4	4	9	7	9	9	
0.8	0.7			0	0	0	0	0	0	0	0	0	0	0	0
	0.8			47	42	63	63	41	45	71	71	50	48	67	67
	0.9			872	875	918	918	815	835	885	885	823	832	884	884
	1			1000	1000	1000	1000	999	999	999	999	1000	1000	1000	1000
	1.1			984	982	990	991	960	961	980	980	980	981	989	990
	1.2			392	386	465	471	365	373	440	445	401	412	479	482
1.3			5	7	11	11	8	12	17	18	7	7	13	13	
0.2	0.7			0	0	1	1	2	1	2	2	2	1	2	2
	0.8			37	42	67	67	33	37	62	62	32	28	48	48
	0.9			335	350	453	453	329	354	457	457	355	369	459	459
	1			834	857	916	916	739	796	868	868	758	803	871	872
	1.1			603	630	707	708	600	632	710	712	592	627	711	715
	1.2			163	159	217	219	183	188	257	260	188	191	253	254
1.3			18	16	25	27	17	19	27	28	24	25	33	33	
0.2	0.7			0	0	0	0	1	1	2	2	0	0	1	1
	0.8			42	43	74	74	36	35	57	57	40	43	65	65
	0.9			367	378	472	472	330	345	439	439	323	334	427	427
	1			796	813	879	881	729	753	830	831	768	786	852	854
	1.1			625	636	705	709	562	597	672	680	599	617	702	707
	1.2			199	198	255	256	168	168	241	244	194	189	258	261
1.3			18	18	29	29	22	26	33	33	16	18	29	29	
0.8	0.7			3	1	6	6	1	1	3	3	1	0	3	3
	0.8			47	48	73	73	38	36	64	64	33	36	58	58
	0.9			391	383	478	478	312	327	417	418	348	354	460	460
	1			797	810	877	878	717	743	815	816	736	744	826	827
	1.1			608	616	685	690	535	562	632	636	586	594	660	663
	1.2			204	200	254	254	187	204	241	243	197	192	257	259
1.3			24	20	35	35	18	21	33	34	26	27	37	37	

### 7. 결론

제제간의 생물학적 동등성 시험에서 실험 도중에 결측치 발생하게 되면 결측치를 가지는 피험자를 제외하고 생물학적 동등성을 입증하는 것이 현재 식품 의약품 안전청의 규정이다. 하지만 적은 표본으로 실험을 수행하는 생물학적 동등성 시험의 경우 결측치를 가지는 피험자 자료의 제외

가 결론에 크게 영향을 줄 수 있다는 것은 알려져 왔다.

<표 3> 모의실험(n=12)에서 생물학적 동등성 판단 비율

sample size	n <sub>1</sub> =n <sub>2</sub> =12 m <sub>1</sub> =11 m <sub>2</sub> =11				n <sub>1</sub> =n <sub>2</sub> =12 m <sub>1</sub> =10 m <sub>2</sub> =10				n <sub>1</sub> =n <sub>2</sub> =12 m <sub>1</sub> =11 m <sub>2</sub> =10			
	GM	PM	BP	BC	GM	PM	BP	BC	GM	PM	BP	BC
CV Rho Ratio	0.7	0	0	0	0	0	0	0	0	0	0	0
	0.8	38	41	56	56	38	35	47	47	48	41	57
	0.9	948	954	968	968	935	945	966	966	947	957	969
	1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1	1000	1000	1000	1000	997	999	999	999	999	997	998
	1.2	477	484	546	548	461	474	540	540	481	502	547
	1.3	5	5	5	5	4	4	4	4	2	2	4
0.1	0.7	0	0	0	0	0	0	0	0	0	0	0
	0.8	43	43	57	57	34	31	47	47	37	34	43
	0.9	967	971	982	982	932	937	956	956	945	952	968
	1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1	1000	999	999	999	993	996	996	997	994	996	997
	1.2	476	495	543	543	465	492	543	543	473	486	530
	1.3	3	2	5	5	2	2	5	5	4	4	5
0.8	0.7	0	0	0	0	0	0	0	0	0	0	0
	0.8	35	33	51	51	47	52	65	65	42	46	57
	0.9	956	956	967	967	942	943	959	959	929	938	953
	1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1	995	996	997	997	993	995	997	997	994	994	996
	1.2	500	503	565	567	437	439	479	486	441	459	500
	1.3	4	4	6	6	5	4	5	5	2	2	6
0.2	0.7	0	0	0	0	0	0	0	0	0	0	0
	0.8	39	40	55	55	43	39	58	58	39	38	48
	0.9	500	513	591	591	448	467	552	552	429	444	532
	1	946	948	965	965	916	932	961	961	927	940	959
	1.1	747	756	809	811	696	723	774	776	738	745	797
	1.2	230	234	279	282	215	220	261	262	217	227	277
	1.3	6	9	14	15	13	15	24	24	12	11	17
0.2 0.5	0.7	0	0	0	0	0	0	0	0	1	1	1
	0.8	33	32	48	48	34	35	46	46	35	41	55
	0.9	482	493	561	561	439	466	533	533	488	485	552
	1	923	927	946	946	905	915	944	944	923	932	953
	1.1	745	748	785	791	689	710	763	767	730	737	783
	1.2	230	229	271	273	223	228	265	271	210	216	259
	1.3	23	25	30	30	13	12	16	16	21	22	31
0.8	0.7	0	0	0	0	1	1	1	1	1	0	2
	0.8	49	53	68	68	40	37	53	53	45	38	57
	0.9	486	479	555	555	455	443	521	521	476	478	555
	1	913	913	941	942	899	911	935	935	902	906	933
	1.1	722	728	765	768	717	722	773	774	730	739	779
	1.2	243	236	285	289	233	240	287	288	231	225	267
	1.3	16	12	20	20	26	24	29	29	14	17	24

본 연구에서는 생물학적 동등성 시험에서 결측치가 발생했을 때 고려할 수 있는 통계적 방법으로 Patel의 방법과 붓스트랩 방법을 고려하였다. 모의실험을 통한 비교에서 Patel의 방법은 현 기준으로 사용하고 있는 Grizzle 방법에 비해서 우월한 방법으로, 붓스트랩 방법은 Patel의 방법보다 우월한 방법으로 나타났다.

붓스트랩 방법은 제제 효과에 대한 신뢰구간을 간단한 재표본 추출방식으로 손쉽게 구할 수 있는 방법으로 생물학적 동등성 시험에서도 널리 사용되어져 왔던 방법 중의 하나이고(Shao et. al. 2000), 표본분포가 정규분포라는 가정 없이도 사용할 수 있다는 장점이 있어 현재 식품 의약품 안전청에서 사용하고 있는 기준을 대체하는 방안으로 생각된다.

한편 최근 미국을 비롯한 선진국들에서 제제간의 생물학적 동등성 기준으로 자리 잡고 있는 모집단 생물학적 동등성(population bioequivalence)이나 개인 생물학적 동등성(individual

bioequivalence) 시험에서 결측치가 생겼을 때 이를 처리하는 방법에 대한 연구도 시급히 이루어져야 한다고 생각된다.

<표 4> 모의실험(n=15)에서 생물학적 동등성 판단 비율

sample size	n <sub>1</sub> =n <sub>2</sub> =15 m <sub>1</sub> =14 m <sub>2</sub> =14				n <sub>1</sub> =n <sub>2</sub> =15 m <sub>1</sub> =13 m <sub>2</sub> =13				n <sub>1</sub> =n <sub>2</sub> =15 m <sub>1</sub> =14 m <sub>2</sub> =13				
	GM	PM	BP	BC	GM	PM	BP	BC	GM	PM	BP	BC	
CV Rho Ratio	0.7	0	0	0	0	0	0	0	0	0	0	0	
	0.8	35	29	48	48	28	28	38	38	30	34	51	51
	0.9	986	986	991	991	970	982	984	984	986	987	993	993
	0.2 1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1	999	1000	1000	1000	998	999	999	999	999	999	999	999
	1.2	544	551	587	588	515	527	567	574	573	566	621	624
	1.3	3	3	3	3	1	1	1	1	3	4	5	5
0.1 0.5	0.7	0	0	0	0	0	0	0	0	0	0	0	
	0.8	31	29	42	42	34	38	47	47	46	40	54	54
	0.9	982	984	988	988	974	981	984	984	982	982	987	987
	1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1	1000	1000	1000	1000	1000	1000	1000	1000	999	999	1000	1000
	1.2	539	547	585	585	553	572	610	612	535	541	577	579
	1.3	0	0	0	0	1	3	3	3	3	4	4	4
0.8	0.7	0	0	0	0	0	0	0	0	0	0	0	
	0.8	41	40	57	57	50	51	63	63	36	32	48	48
	0.9	984	985	989	989	971	974	977	977	983	986	991	991
	1	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
	1.1	1000	1000	1000	1000	999	1000	999	999	998	997	999	999
	1.2	585	584	628	631	563	567	613	616	521	520	565	571
	1.3	4	4	5	5	2	2	2	2	0	1	2	2
0.2	0.7	0	0	0	0	0	0	0	0	0	1	1	1
	0.8	51	44	62	62	40	31	49	49	51	47	63	63
	0.9	565	572	621	621	531	552	604	604	547	557	628	628
	1	972	976	984	984	972	985	989	989	975	979	984	984
	1.1	828	837	861	863	794	818	845	845	835	842	867	868
	1.2	255	256	291	294	252	256	297	301	241	265	307	307
	1.3	10	12	19	19	10	9	16	17	7	8	12	12
0.2 0.5	0.7	0	0	0	0	0	0	0	0	1	1	1	1
	0.8	32	28	38	38	36	34	43	43	29	25	39	39
	0.9	579	582	639	639	561	556	619	619	559	563	622	622
	1	984	983	986	986	975	979	984	984	971	977	985	985
	1.1	846	849	874	875	828	836	860	861	833	841	864	864
	1.2	247	255	287	288	244	255	286	291	258	258	295	295
	1.3	10	11	12	12	10	11	14	14	12	11	14	15
0.8	0.7	0	0	0	0	0	0	0	0	0	0	0	
	0.8	40	42	54	54	32	33	43	43	43	42	51	51
	0.9	566	575	625	625	537	539	595	595	537	531	576	576
	1	976	979	984	984	973	969	980	980	969	970	978	978
	1.1	831	836	850	852	797	798	829	830	829	830	852	854
	1.2	279	282	311	316	248	251	287	291	231	243	275	278
	1.3	22	20	25	25	19	18	20	20	12	9	13	13

참고문헌

[1] Bello, A.L. (1994). A bootstrap method for using imputation techniques for data with missing values. *Biometrical Journal*, 36, 453-464.  
 [2] Berger, L.R. and Hsu, J.C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11, 283-302.

- [3] Czado, C. and Munk, A. (2001). Bootstrap methods for the nonparametric assessment of population bioequivalence and similarity of distribution. *Journal of Statistical Computation and Simulation*, 68, 243-280.
- [4] Chow, S.C. (1990). Alternative approaches for assessing bioequivalence regarding normality assumptions. *Drug Information Journal*, 24, 753-762.
- [5] Chow, S.C. and Liu, J.P. (2000). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, New York.
- [6] Efron, B. (1979). Bootstrap methods : another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- [7] Efron, B. (1994). Missing data, imputation, and the bootstrap. *Journal of the American Statistical Association*, 89, 463-479.
- [8] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- [9] Grizzle, J.E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, 21, 467-480.
- [10] Metzler, C.M. (1974). Bioavailability : a problem in equivalence. *Biometrics*, 30, 309-317.
- [11] Morrison, D.F. (1973). A test for equality of means of correlated variates with missing data on one response. *Biometrika*, 60, 101-105.
- [12] Patel, H.I. (1985). Analysis of incomplete data in a two-period crossover design with reference to clinical trials. *Biometrika*, 72, 411-418.
- [13] Schuirman, D.J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657-680.
- [14] Shao, J., Chow, S.C. and Wang, B. (2000). The Bootstrap procedure in individual bioequivalence. *Statistics in Medicine*, 19, 2741-2754.
- [15] Shao, J. and Sitter, R.R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, 91, 1278-1288.
- [16] Westlake, W.J. (1976). Symmetrical confidence intervals for bioequivalence trials. *Biometrics*, 32, 741-744.

[ 2004년 1월 접수 , 2004년 4월 채택 ]