

동등성 시험을 신뢰구간을 사용하여 검정하는 경우 왜 신뢰도 90%인 신뢰구간을 사용하는가?

강승호¹

¹연세대학교 응용통계학과

(2008년 3월 접수, 2008년 4월 채택)

요약

임상시험의 자료에 대하여 통계적 검정을 실시하는 경우 유의수준 5%를 사용하는 것이 관례이다. 하지만 동등성 시험을 신뢰구간을 사용하여 검정하는 경우, 신뢰도 90%인 신뢰구간이 사용되고 있다. 흔히 신뢰도 $1 - \alpha$ 인 신뢰구간을 검정에 사용하는 경우, 그 검정법의 유의수준은 α 이다. 이 때문에 동등성 검정에서 신뢰도 90%인 신뢰구간을 사용하게 되면, 유의수준은 10%가 되는 것이 아닌가 하는 혼란을 일으켰다. 본 논문에서는 이와 관련된 이슈들을 관련 문헌의 검토와 시뮬레이션을 통하여 명확하게 하여, 제약회사, CRO, 대학병원 등에 종사하는 통계전문가들에게 도움을 주고자 한다.

주요용어: 생물학적 동등성 시험, INTERSECTION UNION TEST, TWO ONE-SIDED TESTS.

1. 서론

두 처리를 비교하는 임상시험에서 많은 경우 임상시험의 목적은 두 처리의 효과가 차이가 있다는 것을 밝히는 것이다. 하지만, 생물학적 동등성 시험처럼, 임상시험의 목적이 두 처리의 효과 차이가 없음을 보이고자 하는 경우도 종종 있다. μ_T 와 μ_S 를 각각 시험약과 대조약의 주평가변수의 모평균이라 하고, $\delta > 0$ 를 동등성 마진이라고 하자. 즉 μ_T 와 μ_S 의 차이가 δ 보다 작으면, 시험약은 대조약과 동등하다고 판단하게 된다. 임상시험에서는 이러한 동등성 검정을 실시하는 경우가 종종 발생하는데, 구체적으로 가설은 다음과 같이 주어진다.

$$H_0 : |\mu_T - \mu_S| \geq \delta \quad \text{vs.} \quad H_1 : |\mu_T - \mu_S| < \delta, \quad (1.1)$$

이러한 동등성 가설을 검정하는 방법으로 신뢰구간을 사용하는 방법이 종종 사용되는데, 신뢰구간이 $(-\delta, \delta)$ 안에 들어가면, H_0 를 기각하게 된다. 여기서 문제가 되는 것은, 흔히 임상시험의 자료에 대하여 통계적 검정을 실시하는 경우 유의수준 5%를 사용하는 것이 관례임에도 불구하고 (ICH, 1998), 신뢰도 95%인 신뢰구간이 아니고 신뢰도 90%인 신뢰구간이 종종 사용된다는 점이다. 예를 들어 생물학적 동등성 판정기준은 AUC와 같은 값을 로그 변환하여 평균의 차이에 대한 90% 신뢰구간을 구하고 이것이 $\log(0.8) = -0.223$ 에서 $\log(1.25) = 0.223$ 사이에 놓이면, 생물학적으로 동등하다고 판단한다 (FDA, 1992; Chow와 Liu, 2000; 권광욱 등, 2006; 식품의약품안전청, 2005). 하지만 신뢰도 $1 - \alpha$ 인 신뢰구

본 논문은 2006년도 과학기술부의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(R01-2006-000-101650-2006)

¹(120-749) 서울 서대문구 성산로 262. 연세대학교 응용통계학과, 부교수. E-mail: seungho@yonsei.ac.kr

간을 검정에 사용하는 경우, 흔히 그 검정법의 유의수준은 α 이다. 이 때문에 동등성 검정에서 신뢰도 90%인 신뢰구간을 사용하는 것은 많은 혼란을 일으켰다. 예를 들면, 신뢰도 90%인 신뢰구간을 사용하는 것은 유의수준을 10%나 허용한다는 말인가? 만일 그렇다면 왜 생물학적 동등성 시험의 경우에만 더 큰 유의수준을 허락하는 것일까? 또는 유의수준이 5%가 되기 위하여 신뢰도 90%인 신뢰구간이 아니라 신뢰도 95%인 신뢰구간이 사용되어야 하는 것이 아닐까? 저자 또한 생물학적 동등성 시험에 대한 통계 방법을 처음 공부할 때 이런 질문을 혼자 던진 적이 있었고, 그 후에도 제약회사, CRO, 대학병원에서 근무하는 분들로부터 비슷한 질문을 종종 받았다. 그래서 이 문제에 대한 해답을 명확히 하고 관련된 문헌들을 정리할 필요성을 느끼게 되어 본 논문을 쓰게 되었다.

2. TOST 방법과 신뢰구간을 사용하는 방법

설명의 편의상 2군 평행설계인 경우를 고려하겠다. μ_T 와 μ_S 를 각각 시험약과 대조약의 주평가변수의 모평균이라 하고, $\delta > 0$ 가 동등성 마진인 경우, 동등성 검정의 가설은 (1.1)처럼 주어진다. 가설 (1.1)을 검정하는 방법은 TOST(two one-sided tests)와 신뢰구간을 사용하는 방법이 널리 사용되고 있다.

TOST 방법은 (1.1)의 귀무가설이 다음의 두 개의 귀무가설의 합집합으로 표현된다는 사실을 이용한다.

$$H_{01} : \mu_T - \mu_S \geq \delta \quad \text{그리고} \quad H_{02} : \mu_T - \mu_S \leq -\delta.$$

만일 H_{01} 과 H_{02} 를 각각 유의수준 α 에서 모두 기각한다면, 우리는 (1.1)에 있는 H_0 를 유의수준 α 에서 기각하게 된다. 자료가 정규분포를 따른다고 가정하고, \bar{y}_T 와 \bar{y}_S , n_1 과 n_2 를 각각 시험약군과 대조약군의 표본평균과 표본크기라고 하는 경우

$$\frac{\bar{y}_T - \bar{y}_S + \delta}{se} > t_\alpha(n_1 + n_2 - 2) \quad \text{그리고} \quad \frac{\bar{y}_T - \bar{y}_S - \delta}{se} < -t_\alpha(n_1 + n_2 - 2) \quad (2.1)$$

를 만족하면, TOST는 유의수준 α 에서 (1.1)에 있는 H_0 를 기각하게 된다. 여기서 se 는 $\bar{y}_T - \bar{y}_S$ 의 표준 오차이며, (2.1)에 있는 두 조건이 모두 만족되어야 (1.1)에 있는 H_0 를 기각하게 됨을 명심해야 한다.

반면에 신뢰구간을 사용하는 방법은 신뢰도 $1 - 2\alpha$ 인 다음의 신뢰구간을 구하여

$$CI_{2\alpha} = (\bar{y}_T - \bar{y}_S - t_\alpha(n_1 + n_2 - 2)se, \bar{y}_T - \bar{y}_S + t_\alpha(n_1 + n_2 - 2)se)$$

이 신뢰구간이 $(\delta, -\delta)$ 안에 들어가면, (1.1)에 있는 H_0 를 기각하게 된다. 조건 (2.1)가 만족된다는 것은, $CI_{2\alpha}$ 가 $(\delta, -\delta)$ 안에 들어간다는 것과 필요충분조건임을 쉽게 보일 수 있다. 예를 들면,

$$\frac{\bar{y}_T - \bar{y}_S + \delta}{se} > t_\alpha(n_1 + n_2 - 2)$$

이 만족된다는 것은, $CI_{2\alpha}$ 의 하한이 $-\delta$ 보다 크거나 같다는 동일한 조건이 된다. 이처럼 두 방법이 동일하다는 것은 다음과 같은 혼란을 낳았다.

(Q1) TOST 방법의 유의수준은 α 인가?

(Q2) 신뢰도 $1 - 2\alpha$ 인 신뢰구간 $CI_{2\alpha}$ 을 사용하여 H_0 와 H_1 를 검정하는 경우, 이 검정방법의 유의수준은 얼마인가?

(Q3) 신뢰구간 $CI_{2\alpha}$ 의 신뢰도는 얼마인가?

(Q4) 흔히 신뢰도 $1 - \alpha$ 인 신뢰구간을 검정에 사용하는 경우, 그 검정법의 유의수준은 α 이다. 하지만 본 논문에서 고려하는 $CI_{2\alpha}$ 인 경우는 신뢰도는 $1 - 2\alpha$ 인데, $CI_{2\alpha}$ 을 사용하는 검정법의 유의수준은 α 이다. 이처럼 모순되어 보이는 현상을 어떻게 설명할 수 있는가?

(Q5) 임상시험에서 통계적 검정의 유의수준은 5%를 사용하는 것이 관례이고, 식품의약품안전청에 의하여 요구되는 바이다. 유의수준 5%인 검정을 하고 싶고, $CI_{2\alpha}$ 을 사용하는 검정법을 사용하고자 한다면, 신뢰도가 얼마인 $CI_{2\alpha}$ 를 사용해야 하는가?

본 논문에서는 위 5가지 질문에 대한 답을 제시하는 것이 목적이다. Berger와 Hus (1996)과 Chow와 Shao (2002)에는 위 5가지 질문에 대한 답이 어느 정도 제시되어 있으나, 난해한 설명으로 이해가 쉽지 않다. 본 논문에서는 이를 보다 더 쉽게 설명하여 실제 임상시험 자료 분석에 종사하는 통계진공자들이 쉽게 이해를 할 수 있도록 하고, 시뮬레이션으로 수학적으로 증명된 사실들을 다시 한 번 확인하고자 한다.

3. 동등성 검정에 사용되는 통계 방법들의 통계적 성질

본 절에서는 2절에서 제기된 문제들에 대한 답을 찾아 정리해보고자 한다.

(Q1) TOST 방법의 유의수준은 α 인가?

통계적 검정방법 T 에 대하여 T 의 size는 다음처럼 정의된다 (Lehmann, 1986).

$$\alpha_T = \sup_{H_0} P(T \text{ rejects } H_0)$$

여기서 supremum은 귀무가설에 속하는 모수공간에서 취하게 되며, 귀무가설에서 값이 지정되는 관심 모수는 값이 지정되지 않지만, 귀무가설에 값이 지정되지 않은 장애모수는 주어진 장애모수공간을 마음껏 움직이게 되므로, supremum은 사실상 장애모수공간에서 취하는 것과 같게 된다. 정의에 의하면 size란 가장 커질 수 있는 1종의 오류 확률로서, 다시 말하면 최악의 경우의 1종의 오류 확률을 나타낸다. T 의 유의수준은 $\alpha_1 \geq \alpha_T$ 를 만족하는 임의의 실수 α_1 으로 정의된다. size와 유의수준을 차이를 아는 것은 대단히 중요한데, 그 차이가 본 논문의 핵심 주제이기도 하다. 예를 들어, $\alpha_T = 5\%$ 인 경우, 5% 보다 크거나 같은 모든 실수값은 유의수준이 될 수 있으므로, 유의수준을 5%라고 말해도 되지만, 유의수준이 10%라고 말해도 무방하다.

TOST 방법은 다음의 Intersection-Union Tests(IUT)의 특별한 경우이다 (Berger와 Hsu, 1996). 귀무가설 H_0 에서 지정하는 모수 공간 Θ_0 가 k 개의 집합 Θ_i 의 합집합으로 표현된다고 가정하자.

$$H_0 : \theta \in \Theta_0 = \bigcup_{i=1}^k \Theta_i \quad \text{vs.} \quad H_1 : \theta \in \Theta_0^c = \bigcap_{i=1}^k \Theta_i^c,$$

각 i ($i = 1, 2, \dots, k$)에 대하여, R_i 는 다음 가설을 검정하는 기각역이라고 하자.

$$H_{0i} : \theta \in \Theta_i \quad \text{vs.} \quad H_{1i} : \theta \in \Theta_i^c.$$

H_0 와 H_1 에 대한 IUT의 기각역은 $\bigcap_{i=1}^k R_i$ 로 주어진다. 즉, IUT에서는 모든 H_{0i} ($i = 1, 2, \dots, k$)가 기각될 때, H_0 를 기각하게 된다.

다음의 정리는 IUT의 유의수준을 보이는데 유용한 수단이다. 이미 알려진 사실이지만, 증명과정은 간단하고, 또한 증명과정까지 이해하는 것이 IUT의 성질을 이해하는데 도움이 된다고 판단하여, 정리뿐만 아니라 증명까지 여기에서 설명하고자 한다. 아래의 두 정리가 본 논문에서 다루는 문제의 핵심을 해결하는 내용이다.

정리 3.1 만일 R_i ($i = 1, 2, \dots, k$)가 H_{0i} vs. H_{1i} 를 검정하는 유의수준 α 인 검정의 기각역이라면, 기각역 $\bigcap_{i=1}^k R_i$ 를 갖는 IUT는 H_0 vs. H_1 을 검정하는 유의수준 α 인 검정이다 (BERGER와 HSU, 1996).

증명: $\theta \in \Theta_0 = \bigcup_{i=1}^k \Theta_i$ 인 θ 는 $i = 1, 2, \dots, k$ 에 속하는 어떤 j 에 대하여 $\theta \in \Theta_j$ 를 만족한다.

$P_\theta(\bigcap_{i=1}^k R_i) \leq P_\theta(R_j) \leq \alpha$ 를 만족하므로, 양변에 supremum 을 취하면 최종 결론을 얻는다. \square

정리 3.2 $i \in \{1, 2, \dots, k\}$ 인 어떤 i 에 대하여, R_i 는 H_{0i} vs. H_{1i} 를 검정하는 size- α 인 검정의 기각역이라고 하자. 또한 $j = 1, 2, \dots, k, j \neq i$ 인 모든 j 에 대하여, R_j 는 H_{0j} vs. H_{1j} 를 검정하는 유의수준 α 인 검정의 기각역이라 하자. 만일 모수들의 어떤 수열 $\theta_l, l = 1, 2, \dots \in \Theta_i$ 에 대하여, $\lim_{l \rightarrow \infty} P_{\theta_l}(R_i) = \alpha$ 이 만족되고, 또한 $j = 1, 2, \dots, k, j \neq i$ 인 모든 j 에 대하여, $\lim_{l \rightarrow \infty} P_{\theta_l}(R_j) = 1$ 이 만족된다면, 이 경우 기각역 $\bigcap_{i=1}^k R_i$ 를 갖는 IUT는 H_0 vs. H_1 을 검정하는 size- α 인 검정이다 (BERGER와 HSU, 1996).

증명: Bonferroni inequality로부터

$$\begin{aligned} P\left(\bigcup_{i=1}^k R_i^c\right) &\leq \sum_{i=1}^k P(R_i^c) \\ 1 - P\left(\bigcap_{i=1}^k R_i\right) &\leq \sum_{i=1}^k [1 - P(R_i)] \\ P\left(\bigcap_{i=1}^k R_i\right) &\geq \sum_{i=1}^k P(R_i) - (k-1) = \alpha + (k-1) - (k-1) = \alpha. \end{aligned}$$

그러므로 동등성 검정이라 모수공간 Θ_1 과 Θ_2 을 다음처럼 가지는 IUT이다.

$$\Theta_1 = \{(\mu_T, \mu_S, \sigma^2) : \mu_T - \mu_S \geq \delta\}, \quad \Theta_2 = \{(\mu_T, \mu_S, \sigma^2) : \mu_T - \mu_S \leq -\delta\}.$$

자료가 정규분포를 따른다고 가정하면, $(\bar{y}_T - \bar{y}_S + \delta)/se > t_\alpha(n_1 + n_2 - 2)$ 을 만족되면 H_{01} 을 기각하는 검정방법은 유의수준이 α 이고, $(\bar{y}_T - \bar{y}_S - \delta)/se < -t_\alpha(n_1 + n_2 - 2)$ 을 만족되면 H_{02} 을 기각하는 검정방법 또한 유의수준이 α 이므로, TOST 방법의 유의수준은 정리 3.1에 의하여 α 가 된다. 정리 3.2를 이용하기 위하여 $\mu_T - \mu_S = \delta$ 이고 $\sigma^2 \rightarrow 0$ 인 모수들의 수열을 고려하자. 이런 모수들은 Θ_1 의 경계점에 위치하게 되며, 모든 $\sigma^2 > 0$ 에 대하여, $P(R_1) = \alpha$ 이고, $\sigma^2 \rightarrow 0$ 에 따라 $P(R_2) \rightarrow 1$ 을 만족한다. 그러므로 정리 3.2에 의하여 TOST 방법의 size는 α 이다. \square

(Q2) 신뢰도 $1 - 2\alpha$ 인 신뢰구간 $CI_{2\alpha}$ 을 사용하여 H_0 와 H_1 를 검정하는 경우, 이 검정방법의 유의수준은 얼마인가?

신뢰구간을 사용하는 방법은 신뢰도 $1 - 2\alpha$ 인 다음의 신뢰구간을 구하여

$$CI_{2\alpha} = \left(\bar{y}_T - \bar{y}_S - t_\alpha(n_1 + n_2 - 2)se, \bar{y}_T - \bar{y}_S + t_\alpha(n_1 + n_2 - 2)se\right)$$

이 신뢰구간이 $(-\delta, \delta)$ 안에 들어가면, H_0 를 기각하게 된다. 조건 (2.1)가 만족된다는 것은, $CI_{2\alpha}$ 가 $(-\delta, \delta)$ 안에 들어간다는 것과 필요충분조건임을 쉽게 보일 수 있다. 예를 들면,

$$\frac{\bar{y}_T - \bar{y}_S + \delta}{se} > t_\alpha(n_1 + n_2 - 2)$$

이 만족된다는 것은, $CI_{2\alpha}$ 의 하한이 $-\delta$ 보다 크거나 같다고 동일한 조건이 된다. 그러므로 신뢰도 $1 - 2\alpha$ 인 신뢰구간 $CI_{2\alpha}$ 을 사용하여 H_0 와 H_1 를 검정하는 경우, 이 방법은 TOST와 동일한 방법이므로 (Q1)의 결과에 의하여, 이 검정방법의 유의수준은 α 이며, size 또한 α 이다.

표 4.1. 시뮬레이션에 의한 경험적 유의수준과 신뢰도

$n = n_1 = n_2$	TOST의 유의수준	$CI_{2\alpha}$ 의 신뢰도	$CI_{2\alpha}$ 의 유의수준
50	0.049	0.904	0.049
100	0.050	0.903	0.050
200	0.047	0.907	0.047

(Q3) 신뢰구간 $CI_{2\alpha}$ 의 신뢰도는 얼마인가?

$$CI_{2\alpha} = (\bar{y}_T - \bar{y}_S - t_{\alpha}(n_1 + n_2 - 2)se, \bar{y}_T - \bar{y}_S + t_{\alpha}(n_1 + n_2 - 2)se)$$

신뢰구간 $CI_{2\alpha}$ 의 오른쪽과 왼쪽으로 각각 α 만큼의 확률이 빠지므로, 신뢰도는 $1 - 2\alpha$ 이다.

(Q4) 흔히 신뢰도 $1 - \alpha$ 인 신뢰구간을 검정에 사용하는 경우, 그 검정법의 유의수준은 α 이다. 하지만 본 논문에서 고려하는 $CI_{2\alpha}$ 인 경우는 신뢰도는 $1 - 2\alpha$ 인데, $CI_{2\alpha}$ 을 사용하는 검정법의 유의수준은 α 이다. 이처럼 모순되어 보이는 현상을 어떻게 설명할 수 있는가?

우선 $CI_{2\alpha}$ 을 사용하는 검정법의 size가 얼마인지를 알아내는 것이 혼동을 정리하는 길이다. $CI_{2\alpha}$ 을 사용하는 검정법은 TOST 방법과 일치하므로, 정리 3.2에 의하여 size는 α 이다. 유의수준의 정의에 의하여 size α 보다 큰 값은 모두 유의수준이라고 불러도 된다. 즉 $CI_{2\alpha}$ 을 사용하는 검정법의 유의수준은 α 라고 해도 되고, 2α 라고 해도 된다. 그러므로 신뢰도 $1 - 2\alpha$ 인 $CI_{2\alpha}$ 을 사용하는 검정법의 유의수준은 2α 가 된다고 말할 수 있으므로 모순될 것이 없다. 이러한 사실은 이미 정리 3.1과 3.2에서 수리적으로 증명이 되어 있으므로, 본 논문에서는 시뮬레이션을 통하여 이 사실을 다시 한 번 확인해보고자 한다. 이론적 배경이 강하지 않은 실무적인 일을 하는 통계전문가들에게는 어려운 이론보다는 쉬운 시뮬레이션을 통하여 이 사실을 확인하는 것이 더 쉬운 방법이 될 수도 있을 거라고 생각하기 때문이다. 4절에서는 시뮬레이션을 통하여 $CI_{2\alpha}$ 을 사용하는 검정법의 size가 α 임을 확인하였다.

(Q5) 임상시험에서 통계적 검정의 유의수준은 5%를 사용하는 것이 관례이고, 식품의약품안전청에 의하여 요구되는 바이다. 유의수준 5%인 검정을 하고 싶고, $CI_{2\alpha}$ 을 사용하는 검정법을 사용하고자 한다면, 신뢰도가 얼마인 $CI_{2\alpha}$ 를 사용해야 하는가?

답은 신뢰도 90%인 $CI_{2\alpha}$ 이다. 즉 $\alpha = 5\%$ 이다. 이 신뢰구간의 신뢰도는 90%이지만, 그 신뢰구간을 사용하는 검정방법의 size는 5%가 되어 유의수준은 5%가 된다. 이것이 바로 가이드라인에 나와 있는 내용이다 (FDA, 1992; 식품의약품안전청, 2005).

4. 시뮬레이션 결과

설명의 편이상 2군 평행설계인 경우를 고려하였다. $\alpha = 0.05$ 인 경우, 동등성 마진을 $\delta = 10$ 으로 잡고, $X_1 \sim N(110, 5^2)$ 이고 $X_2 \sim N(100, 5^2)$ 인, 서로 독립인 확률변수 X_1 과 X_2 를 따르는 크기가 각각 $n (= n_1 = n_2)$ 인 난수들을 발생시켜, TOST 방법에 의하여 귀무가설이 기각되는지, 신뢰구간 $CI_{2\alpha}$ 가 두 모평균의 차이인 $110 - 100 = 10$ 을 포함하는지, $CI_{2\alpha}$ 을 사용하는 검정방법에 의하여 귀무가설이 기각되는지, 총 10,000을 반복하여 조사하여 다음의 결과를 얻었다. 이 시뮬레이션 결과는 3절의 내용이 옳다는 것을 다시 한 번 확인해준다. 즉, $\alpha = 0.05$ 인 경우, TOST 방법의 유의수준은 5%이고, 신뢰구간 $CI_{2\alpha}$ 의 신뢰도는 90%, $CI_{2\alpha}$ 를 사용하는 검정방법의 유의수준은 5%임을 보여준다. 물론 $CI_{2\alpha}$ 를 사용하는 검정방법의 유의수준이 10%라고 말해도 된다.

5. 결론 및 고찰

본 논문에서는 IUT를 소개하고, TOST 방법을 IUT로 설명한 후, TOST의 유의수준과 size는 α 이고, 신뢰도 $1 - 2\alpha$ 인 신뢰구간 $CI_{2\alpha}$ 을 사용하여 동등성을 검정하는 경우, 이 검정방법의 유의수준과 size는 α 이며, 신뢰구간 $CI_{2\alpha}$ 의 신뢰도는 $1 - 2\alpha$ 임을 명확히 하고, 시뮬레이션을 통하여 이런 사실을 재확인 하였다. 이처럼 동등성 검정에서 발생하는 혼동을 정리하여, 실무에 종사하는 통계전문가들이 쉽게 이 문제에 대한 의문점을 해소할 수 있도록 하였고, 시뮬레이션 프로그램을 제공하여 직접 여러 모수 값에서 다양한 시뮬레이션을 시도해 볼 수 있도록 하였다. 시뮬레이션 프로그램은 SAS로 작성되었으며, 저자에게 이메일로 요청하면 얻을 수 있다.

동등성 검정을 하는데 사용되는 신뢰구간은 $CI_{2\alpha}$ 말고도 다음의 신뢰구간들이 제안되어 있다. Westlake (1976)가 제안한 신뢰구간은 $(\bar{y}_T - \bar{y}_S - t_{\alpha_1}se, \bar{y}_T - \bar{y}_S + t_{\alpha_2}se)$, $\alpha_1 + \alpha_2 = \alpha$ 이고, Liu (1990)는 신뢰구간 $(-|\bar{y}_T - \bar{y}_S| - t_{\alpha}se, |\bar{y}_T - \bar{y}_S| + t_{\alpha}se)$ 을 제안했으며, Hsu (1984)는 $(\min(0, \bar{y}_T - \bar{y}_S - t_{\alpha}se), \max(0, \bar{y}_T - \bar{y}_S + t_{\alpha}se))$ 를 제안하였다. 이 세 신뢰구간은 모두 신뢰도가 $1 - \alpha$ 이며, 신뢰구간이 $(-\delta, \delta)$ 사이에 들어가는 경우 H_0 를 기각하면 유의수준이 α 인 검정방법이 된다. 즉 이 세 신뢰구간은 $CI_{2\alpha}$ 가 갖는 혼란스러운 점을 가지고 있지 않다.

참고문헌

- 권광욱, 김옥희, 김종국, 김주일, 배균섭, 사흥기, 이용복, 이화정, 최선옥 (2006). <생물학적 동등성의 이해>, 신일상사.
- 식품의약품안전청 (2005). 고시 제 2005-31호 생물학적 동등성시험기준.
- Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection union-tests and equivalence confidence sets (with discussion), *Statistical Science*, **11**, 283-319.
- Chow, S. C. and Liu, J. P. (2000). *Design and Analysis of Bioavailability and Bioequivalence Studies*, 2nd edition, Marcel Dekker, New York.
- Chow, S. C. and Shao, J. (2002). *Statistics in Drug Research: Methodologies and Recent Developments*, Marcel Dekker, New York.
- FDA (1992). Guidance on statistical procedures for bioequivalence studies using a standard two-treatment crossover design, *Office of Generic Drugs*, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
- Hsu, J. C. (1984). Constrained simultaneous confidence intervals for multiple comparisons with the best, *The Annals of Statistics*, **12**, 1136-1144.
- ICH (1998). Statistical principles for clinical trials, *Tripartite International Conference on Harmonization Guideline*, E9.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd edition, John Wiley & Sons, New York.
- Liu, H. K. (1990). Confidence intervals in bioequivalence assessment, In *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 51-54.
- Westlake, W. J. (1976). Symmetric confidence intervals for bioequivalence trials, *Biometrics*, **32**, 741-744.

Why is 90% Confidence Interval Used When Confidence Interval Approach is Used for Testing Equivalence?

Seung-Ho Kang¹

¹Dept. of Applied Statistics, Yonsei University

(Received March 2008; accepted April 2008)

Abstract

It is a convention to use 5% significance level when a statistical test is employed for clinical data. But when a confidence interval is used for testing equivalence, 90% confidence interval has often been used. When $1 - \alpha$ confidence interval is used for hypothesis testing, its significance level is often α . So it makes a confusion that the significance level is 10% if 90% confidence interval is employed for testing equivalence. In this paper I will clarify this issue by reviewing relevant papers and conducting simulation studies. I hope that it will be beneficial to statisticians in pharmaceutical companies, CROs, university hospitals.

Keywords: Bioequivalence, intersection union test, two one-sided tests.

¹Associate Professor, Dept. of Applied Statistics, Yonsei University, Seoul 120-749, Korea.
E-mail: seungho@yonsei.ac.kr