임상적 동등성 평가를 위한 임상시험에서 동등성범위의 설정과 가설검정

서울대학교 의과대학 예방의학교실¹⁾, 서울대학교병원 임상시험센터²⁾

박병주^{1,2)}, 정귀옥²⁾

Abstract

Equivalence Margin and Hypothesis Testing in the Equivalence Trial

Byung-Joo Park^{1,2)}, Kui-Oak Jung²⁾

Department of Preventive Medicine, Seoul National University College of Medicine¹⁾
Clinical Trial Center, Seoul National University Hospital²⁾

Interest in equivalence trials has been increasing for many years because ethical requirements preclude the use of a placebo. Equivalence trials aim to show that two treatments have equivalent therapeutic effects. The design of a clinical trial to test the equivalence of two treatments differs from that of testing efficacy. The conventional null hypothesis of no difference between two treatments is not appropriate for equivalence trials. So, we must use non-zero (inequivalence) null hypothesis to test clinical equivalence between two treatments. It's very important but difficult to specify a threshold value(δ) which is clinically insignificant in advance for defining equivalence. Statistical analysis should be based on confidence intervals. When we have predefined a range of equivalence, we can check whether the confidence interval obtained on the observed difference lies entirely within the range. If it does, equivalence is demonstrated. In this paper, the several issues of equivalence clinical trials are discussed: choice of equivalence margin, testing statistical hypothesis, and estimating confidence intervals.

KEY WORDS : Clinical trials · Equivalence · Non-inferiority · Delta · Statistical hypothesis.

1. 서 론

새로운 약물의 효능과 안전성을 평가할 수 있는 가장 과학적인 방법은 임상시험을 실시하는 것이며, 최근 들어 임상시험의 중요성이 부각되면서 이에 대한 다양한 방법론적인 문제가 제기되고 있다. 이러한 임상시험 방법 중 가장 과학적이고 신뢰할 수 있는 방법은 무작위배정을 통한 위약대조 · 양측 눈가림 임상시험을 실시하는 것이지만, 해당 적응증에 따라 이미 기존에 공인된 표준치료제가 존재하는 경우에는 위약 대

조 임상시험을 실시하는 것은 비윤리적이어서¹⁻⁴⁾, 이를 해결하기 위하여 해당 적응증에 대하여 이미 시판되고 있는 표준치료약물이 존재하는 경우에는 새로운 치료제의 유효성과 안전성을 평가할 때 위약 대신 표준치료제 대조군을 사용하여야 한다고 주장하였다. 이러한 임상시험을 표준치료제 대조 임상시험(active controlled clinical trial)이라 한다.

새로운 치료제가 표준 치료제보다 더 나을 것이라고 기대하여 이를 증명하고자 하는 경우는 위약 대조 임상시험과 방법론적인 측면에서 차이가 없다. 그러나

교신저자 : 박병주

소 속 : 서울대학교 의과대학 예방의학교실

주 소 : 서울시 종로구 연건동 28 ☎110-799

전 화 : (02) 740-8325 FAX : (02) 747-4830 E-mail : bjpark@snu.ac.kr

새로운 치료제를 기존의 표준치료제와 비교할 때, 유효성 측면에서는 두 치료제가 서로 유사하지만 안전성, 복용의 편리성, 비용 그리고 삶의 질(quality of life) 등 여러 다른 측면에서는 새로운 치료제가 더 나을 것이라고 기대하는 임상시험을 수행할 수 있다. 이와 같이 새로운 치료제의 유효성이 기존 치료제와 임상적으로 유사하거나 혹은 동등하다는 것을 확인하고자 수행하는 임상시험을 ‘동등성 임상시험’이라고 한다^{3, 5, 6)}.

동등성 임상시험은 약물의 투여 후 시간에 따른 혈중농도의 변화양상을 비교하여 기존 약물과 새로운 약물의 동등성을 입증하기 위한 ‘생물학적 동등성 임상시험(bioequivalence trials)’과 임상적인 측면에서 약물의 유효성과 안전성이 동등함을 증명하기 위한 ‘임상적 동등성 임상시험(clinical equivalence trials)’ 그리고 새로운 치료제의 유효성이 기존 치료제의 유효성보다 나쁘지 않음을 보이기 위한 ‘비열등성 임상시험(non-inferiority trials)’으로 구분할 수 있다. 이들 중 ‘생물학적 동등성 임상시험’에 대한 연구의 설계, 수행 및 결과 해석에 대한 연구는 비교적 많은 경험을 가지고 있지만, ‘임상적 동등성 임상시험’과 ‘비열등성 임상시험’에 관한 연구는 동등성 인정 한계의 설정을 포함한 연구 방법론적인 문제가 아직까지 명확하게 정립되어 있지 않은 실정이다⁷⁾.

우리 나라에서도 의약분업이 2000년 7월에 실시될 예정이다. 의약분업이 실시될 경우에는 새로 개발된 신약이나 외국에서 수입해 오는 약물뿐만 아니라, 이미 시판되고 있는 약물에 대하여도 약효의 동등성에 대한 명확한 정의가 내려져야 하므로 이들에 대한 생물학적 동등성 임상시험이 필수적으로 요구되고 있으며, 또한 신약 개발이 활발히 추진되고 있는 우리나라의 상황을 감안할 때, 기존에 존재하는 약물과 유사한 성분의 신약을 개발하는 경우에 약효는 유사하지만, 비용-효과와 안전성 측면에서는 신약이 더 낫다는 것을 증명하기 위한 임상적 동등성 임상시험의 필요성은 더욱 증가될 것으로 사료된다.

그러므로 최근 많은 연구자들의 관심이 집중되고 있

으며, 연구의 설계와 수행과정에서 주의가 요구되는 동등성 임상시험의 방법론을 검토하고, 이미 출판된 외국 문헌에서 동등성 임상시험 결과를 검토하여 최근 많은 연구자들의 관심의 대상이 되고 또한 논란을 일으키고 있는 동등성 인정 한계의 결정과정을 살펴보고, 동등성을 판단하기 위한 기준 및 통계적 검정법 등을 살펴봄으로써 향후 실시될 동등성 임상시험의 준비에 도움을 주고자 한다.

2. 동등성과 차이성 평가에 대한 통계적 가설의 비교

임상시험의 통계적 가설을 설정하기 전에 연구자는 제1종 오류 및 제2종 오류의 발생 가능성을 미리 정해 두어야 한다^{3, 5)}. 제1종 오류를 범하게 될 확률은 ‘잘못해서 귀무가설을 기각할 확률’로 정의하는데, 이를 통계적 유의수준이라 하며 α 로 표시한다. 그리고 ‘실제로 차이가 있는데 잘못해서 차이가 없다’고 판정 내리게 되는 오류를 제2종 오류라 하며 β 로 표시한다. 여기서 $1-\beta$ 는 ‘실제로 치료군간에 어느 정도 이상의 차이가 존재할 때 귀무가설을 기각해서 차이가 있음을 관찰하게 될 확률’을 의미한다는 뜻에서 통계적 검정력(statistical power)이라고 한다. α 와 β 는 임상시험을 시작하기 전에 연구자가 미리 결정하는 것으로 교과서적인 기준은 없다. 하지만, 두 종류의 오류를 동시에 최소화할 수는 없는 것이기 때문에 연구의 특성에 맞게 설정하여야 하며, 일반적으로 α 를 5%, β 를 20%로 정한다⁸⁾.

통계적 가설이란 확률변수에 대한 분포의 형태나 분포와 관련된 모수 등에 관한 연구자의 주장으로 정의되는데, 임상시험에서의 모수는 두 치료군에서 관찰되는 치료효과의 차이가 된다. 차이를 보이기 위한 비교 임상시험에서의 귀무가설은 ‘두 약물의 치료효과에 차이가 없다’는 것이고, 이에 대한 대립가설은 ‘두 약물의 치료효과에 차이가 있다’는 것으로 정의한다^{3, 9, 10)}. 그러나 동등성 임상시험에서도 차이성 임상시험에서 정의된 것처럼 ‘차이가 없다’는 귀무가설을 사용한다

표 1. 차이성 임상시험, 동등성 임상시험, 그리고 비열등성 임상시험에서의 통계적 가설

	차이성 임상시험	동등성 임상시험	비열등성 임상시험
귀무가설	$H_0 : \mu_N - \mu_C = 0$	$H_0 : \mu_N - \mu_C \geq \delta$	$H_0 : \mu_N - \mu_C \leq -\delta$
대립가설	$H_1 : \mu_N - \mu_C > 0$	$H_1 : \mu_N - \mu_C < \delta$	$H_1 : \mu_N - \mu_C > -\delta$

μ_N, μ_C 은 각각 신약과 기존치료제 대조군에서의 치료율

δ : 동등성을 인정할 수 있는 한계치

표 2. 동등성 임상시험에서의 검정통계량과 가설 검정

	동등성 임상시험	비열등성 임상시험
검정 통계량	$z_1 = \frac{(\mu_N - \mu_C) - \delta}{s_d}$ $z_2 = \frac{\delta - (\mu_N - \mu_C)}{s_d}$	$z = \frac{(\mu_N - \mu_C) + \delta}{s_d}$
가설 검정	여기서 d = 관찰된 평균 차이, s_d : 관찰된 d 의 표준오차 If z_1 and $z_2 > z_{1-\alpha}$ then reject H_0 . If $z > z_{1-\alpha}$, then reject H_0	

면 통계적 해석상에 심각한 문제가 발생한다.^{3,11-13)} 즉, '차이가 없다'라는 귀무가설을 기각하지 못하면 이것을 받아들여지게 되어 두 약물의 효과에 차이가 없다, 즉 동등하다는 잘못된 결론을 내리게 되는 것이다. 그러므로 두 치료제의 치료적 동등성을 확인하기 위한 동등성 임상시험에서의 귀무가설은 '두 약물의 치료효과 차이는 동등성 인정 한계인 δ (non-zero)보다 크거나 같다'로 설정하며, 이에 대한 대립가설은 '두 약물의 치료효과 차이는 동등성 인정 한계인 δ 미만이다'로 설정한다. 그리고 두 치료제의 비열등성을 확인하기 위한 임상시험에서의 가설은 동등성 인정한계의 하한치 ($-\delta$)를 기준으로 하며, 귀무가설은 '두 약물의 치료효과 차이가 $-\delta$ 보다 작거나 같다'로, 대립가설은 '두 약물의 치료효과 차이는 $-\delta$ 보다 크다'로 설정한다(표 1).

3. 임상적 동등성을 인정할 수 있는 한계(δ)의 설정

위약 대조군을 사용하는 차이성 임상시험에서의 치료효과 차이는 임상적으로 명백하게 의미가 있는 값으로 설정하여야 한다. 즉, 차이성 임상시험에서 치료

효과 차이는 최소한의 유효성보다 더 큰 치료효과 차이를 기대하는 과학적인 근거가 되는 것이다. 그러나 동등성 임상시험에서는 새로운 치료제를 기존 치료제의 유효성과 비교하였을 때, 관찰된 차이가 임상적으로 인정할 수 없을 정도로 작다는 것을 보이는 것이 목적 이므로 비교 임상시험에서 차이를 보이기 위한 목적으로 사용하였던 δ 보다 더 작은 값을 사용하여야 한다³⁾.

이러한 δ 의 선택은 동등성 임상시험의 계획단계에서 가장 어려운 과정중의 하나로 역학자나 생물통계학자 혹은 한 사람의 임상연구자가 결정하는 것이 아니라, 충분한 전문지식을 가진 여러 임상전문가들과의 광범위한 토론을 통하여 결정하여야 한다. 임상적으로 동등하다고 인정할 수 있는 치료효과 차이는 대상 적응증에 따른 약물의 약효군에 따라 상이할 수 있다. 따라서, 이러한 범위를 확실적인 기준에 의하여 기계적으로 결정할 수는 없다. 동등성 범위를 결정하기 위하여는 전임상자료를 포함하여 신약 개발과정에서 생산된 모든 자료들과 외국에서 이미 시판되고 있는 경우에는 외국에서 수행된 시판후 연구자료들을 필요로 한다. 이러한 자료들에서 제시된 치료효과에 대한 결과들을 자료로 통계적인 방법을 이용하여 동등성 범위를 결정하게 된다. 충분한 근거자료가 없는 경우에는

해당 분야 임상전문가들의 합의에 의하여 결정할 수도 있는데, 이렇게 결정된 값은 물론 임상적으로 타당한 의미를 가지고 있어야 한다.

임상적 동등성을 인정할 수 있는 범위는 0을 중심으로 대칭인 경우와 그렇지 않은 경우로 나눌 수 있다. 먼저, 동등성 인정 범위가 0을 중심으로 대칭인 경우는 관찰된 결과가 동등성 범위의 양쪽 바깥에 있을 경우는 두 치료제의 효과에 차이가 있다, 즉 두 치료제가 동등하지 않다고 평가하고, 대칭인 범위내에 들어오는 경우에는 두 치료제의 동등성을 인정하는 것으로써, 양측 가설 검정을 사용한다. 또 다른 경우는 0을 중심으로 대칭이 아닌 동등성 한계를 선택하는 것이다. 이것은 새로운 약제가 더 좋다는 것을 입증하는 데는 관심이 없으나 새로운 약이 기존약제의 치료효과보다 나쁘지 않음을 증명하고자 하는 경우에 사용된다. 이때는 단측검정이 사용되어 관찰된 치료효과와의 차이가 미리 설정된 동등성 한계의 하한치보다 크다는 것을 입증하고자 하는 것이다. 이것은 개발된 약물이 부작용 측면에서 많은 개선을 보였을 경우나 임상시험의 목적이 새로운 약제의 유효성이 표준치료제의 유효성보다 나쁘지 않음을 증명하고자 하는 경우에 사용된다.

이러한 두 가지 방법은 동등성 임상시험의 목적에 따라 다르게 적용될 수 있다. 즉, 일반적인 동등성 임상시험에서는 관찰된 치료효과와의 차이가 동등성 인정 범위의 하한치보다 작을 경우뿐만 아니라 상한치보다 클 경우에도 효과는 좋지만 동등하지 않다고 정의하여 0을 중심으로 대칭인 동등성 인정범위를 사용한다. 비열등성 임상시험에서는 치료효과와의 차이가 동등성 인정 범위의 하한치보다 크기만 하면 기존 약제보다 나쁘지 않음을 인정하는 것으로써 동등성 인정범위의 하한치만을 설정한다.

동등성 인정범위의 결정과정을 가상의 예를 들어 살펴본다.

천식 환자를 대상으로 beta-agonist 약제의 기존 제형과 새로운 제형의 효과가 동등함을 평

가하기 위한 임상시험을 수행하려고 한다. 주 결과변수는 환자 자신이 측정한 peak expiratory flow rate(PEF)이다. 이 때 두 가지 제형의 효과가 임상적으로 동등함을 입증하기 위하여 필요한 동등성 인정한계를 결정하고자 한다.

임상시험을 계획하는 단계에서 사용 가능한 정보는 다음의 3가지이었다.

첫째, 이전에 수행된 차이성 임상시험에서 기존제형과 위약과의 효과차이는 37 l/min 이었다.

둘째, beta-agonist 약제의 단기효과에 의한 PEF 증가량의 최대치는 70 l/min이었다.

셋째, 동등성 인정 한계값은 기존 제형에 대한 위약과의 효과차이보다는 작은 값으로 결정하여야 하며, 그 차이가 임상적으로 받아들일 수 있는 값이어야 하므로 이에 대한 의견을 천식에 대한 전문지식을 가지고 있는 임상 전문가들로부터 수렴하였다.

위의 세 가지 정보를 이용하여 연구진들은 기존 제형과 위약과의 차이인 37 l/min의 대략 1/2, 평균 최대 증가량인 70 l/min의 대략 1/4에 해당하는 값을 동등성 인정범위로 사용하고자 하였으며, 이 값을 이용하여 최종적으로 전문가들이 임상적으로 인정할 수 있는 차이를 15 l/min으로 결정함으로써 동등성 인정범위를(-15, 15)로 잡아 최종 결과변수의 치료차이에 대한 신뢰구간이 설정된 범위 내에 들어오면 기존 제형과 새로운 제형의 효과가 동등하다는 결론을 내리고자 하였다.

4. 동등성의 판정을 위한 검정통계량과 가설 검정

일반적으로 임상시험에서의 가설 검정은 치료효과와의 차이를 검정하는 통계적 방법 (예, t-test, χ^2 -test 등)을 이용하여 p-값만으로 결과를 판정하였다. 그러나 이러한 방법은 두 치료군 사이에 관찰되는 차이 여부에 대한 평가는 가능하지만, 차이의 크기를 추정하기는 힘들다¹⁴⁾. 그러므로 동등성 임상시험에서의

그림 1. 동등성과 비열등성의 평가.

검정은 일반적인 가설 검정법과 신뢰구간을 이용한 검정법을 함께 실시하며(표 2), 동등성에 대한 최종 판정은 치료효과 차이에 대하여 미리 정해진 신뢰수준(90%, 95% 혹은 99%)에서 산출된 신뢰구간을 사용한다. 동등성 임상시험의 경우 관찰된 신뢰구간이 미리 정의한 동등성 인정 한계($\pm\delta$)내에 완전하게 포함될 경우에 두 치료법이 동등하다고 판정하고, 새로운 치료제가 표준치료제보다 열등하지 않음을 증명하고자 하는 비열등성 임상시험의 경우에는 계산된 신뢰구간의 하한값이 동등성 인정한계의 하한치($-\delta$)를 초과하면 새로운 치료제가 기존치료제보다 열등하지 않다고 판단한다.^{3,5-7)}(그림 1).

동등성 임상시험의 자료분석에서 어떠한 피험자를 결과분석에 포함시킬지를 결정하는 통계적 분석 원칙에는 피험자들이 실제로 그 치료를 받았는지 여부에 관계없이 무작위 배정된 모든 피험자를 대상으로 분석하는 '의도된 대로 분석(Intention-to-Treat Analysis, 이하 ITT)'법과 등록기준을 만족하고 임상시험 계획서를 일정 수준이상 준용한 피험자만을 분석에 포함시키는 '계획서 준용 분석(Per-Protocol Analysis, 이하 PP)'법이 있다. 이 중 ITT 분석법은 약물을 일반인들에게 실제로 투여하였을 때 발생할 수 있는 여러 가지 문제를 반영할 수 있기 때문에 일반적으로 차이성 임상시험 보다 보수적인 방법으로 알려져 있다. 그러나 동등성 임상시험에서의 ITT분석은 프로토콜 위반자와 탈락자를 모두 분석에 포함시킴으로써 두 치료제 사이에 존재하는 실제 차이를 희석시켜서

동등성을 입증할 수 있는 확률을 증가시킬 것이므로 더 이상 보수적인 방법이라고 할 수 없다^{4,8)}. 그리고 PP분석법은 중도 탈락자나 프로토콜 위반자를 분석에서 제외하기 때문에 두 군간에 존재하는 치료효과와의 차이를 오히려 증가시킬 것이다. 그러나 이 방법만으로 평가하게 되면 한 피험자가 치료법에 반응을 보이지 않을 경우 그 피험자를 조기에 탈락시키는 상황 등이 발생할 수 있다는 단점이 존재한다. 그러므로 최종 결론을 내리기 전에 이러한 탈락자들에 의한 비뚤림이 발생하였는지를 파악하기 위하여 PP 분석에서 제외된 피험자들의 특성을 세심하게 검토하여야 한다. 실제로 두 군에서 관찰된 부작용 발생률이나 탈락 양상이 현저하게 다르다면 이것 자체로도 두 치료제가 완전히 동등하지 않다는 결론을 내릴 수 있기 때문이다.

차이성 임상시험에서와 마찬가지로 동등성 임상시험에서도 통계적 분석과정에 두 가지 분석법을 모두 포함하는 것이 바람직하다. 두 가지 분석결과가 모두 동등성을 입증할 수 있는 일관된 결과가 나올 수 있도록 환자의 순응도를 높이고, 프로토콜 위반을 줄여야 한다. 이를 위하여 임상시험 계획단계로부터 연구를 종료하기까지의 전체 수행과정에서 연구자는 최선을 다하여야 할 것이다.

5. 외국에서 실시된 동등성 임상시험의 사례

지금까지 동등성 임상시험을 설계하는 과정에서 고

표 3. 동등성을 증명하기 위하여 설계된 외국 임상시험의 결과 요약

Topics	Design	Sample size	Delta	Outcome	Result
Comparison of the safety and efficacy of the combination and the concomitant administration of dorzolamide/timolol ¹⁵⁾	Multicenter Randomized Double-blind Parallel	Enrolled 299 Complete 290	1.5mmHg	• Mean change in IOP* from the baseline	<IOP change> • at trough 0.01 (-0.52, 0.55) • at peak 0.08 (-0.45, 0.60)
Fixed dose of perindopril+indapamide vs. fixed dose of captopril +hydrochlorothiazide and enalapril+hydrochlorothiazide in the treatment of hyper-tension ¹⁶⁾	Multinational Multicenter Randomized Double-blind Parallel	Enrolled 527 Complete 492	4mmHg	• Change DBP from baseline	• Captopril+HCTZ** ⇒ 0.3 (-1.1, 1.7) • Enalapril+HCTZ ⇒ 1.1 (-0.4, 2.6)
Cefaclor and cefuroxime axetil in the treatment of acute otitis media with effusion in children who failed amoxicillin ¹⁷⁾	Single-blind (Investigator) Randomized Parallel	Enrolled 205 Complete 205	15%	• Pneumatic otoscopy • tympanogram	<after 1-6 days> • Cefaclor : 92.3% • Cefuroxime : 89.1% • Diff=3.2(-4.8 - 11.2) <after 10-16days> • Cefaclor : 82.7% • Cefuroxime : 83.2% • Diff=0.5(-10.8-9.9)

*IOP = Intraocular pressure

**HCTZ = Hydrochlorothiazide

려되어야 하는 문제들과 최종 결과분석에 관한 문제들을 살펴보았다. 외국에서 실시된 동등성 임상시험의 사례를 대상으로 동등성 임상시험에서의 동등성 범위의 설정과 통계적 분석 결과에 대한 해석 방법을 살펴보기로 한다(표 3).

Hutzelmann 등¹⁵⁾의 논문은 open angle glaucoma 혹은 ocular hypertension 환자의 치료를 위하여 dorzolamide/timolol 혼합용액과 dorzolamide와 timolol을 각각 처방하는 방법의 유효성과 안전성을 평가하기 위한 다기관, 무작위, 양측 눈가림, 병행설계 임상시험이다. 최종 결과변수는 기초검사값으로부터 IOP (Intraocular Pressure)의 평균 변화정도이었으

며, 변화정도 차이의 신뢰구간이 $\pm 1.5\text{mmHg}$ 이내일 경우에 동등성을 인정하기로 하였다($\delta=1.5\text{mmHg}$). 논문에서 제시된 동등성 판정은 치료차이에 대한 90% 신뢰구간을 사용하여 관찰된 신뢰구간이 (-1.5mmHg, 1.5mmHg)내에 완전하게 포함될 경우에 두 치료법이 동등하다고 판정하기로 하였다. 최종 계산된 치료차이는 0.01mmHg이었으며, 이에 대한 90% 신뢰구간은 (-0.52, 0.55)로 관찰되어 미리 정의한 신뢰구간내에 완전히 포함되므로 동등성을 인정하였다.

Roger 등¹⁶⁾의 연구는 경증부터 중등증까지의 고혈압 환자를 대상으로 고정 용량 혼합약물 perindopril

+indapamide (4+1.25mg)와 두 혼합약제 captopril+hydrochlorothiazide(50+25mg)와 analapril+hydrochlorothiazide (20+12.5mg)의 동등성을 평가하기 위한 다기관, 무작위, 양측 눈가림, 병행설계 임상시험으로 피험자들은 무작위 배정된 치료제를 8주 동안 투여 받았다. 최종 결과변수는 약물 투여 후 24 ± 3 시간에 편안한 상태에서 측정된 이완기 혈압에 대한 기초검사값으로부터의 변화정도이었다. 동등성 판정 기준은 평균 치료효과에 대한 90% 신뢰구간을 사용하여 관찰된 신뢰구간이(-4mmHg, 4mmHg) 이 내에 완전하게 포함될 경우에 두 치료법이 동등하다고 판정하기로 하였으며, 최종 관찰된 치료효과 차이는 perindopril+indapamide(4+1.25mg)를 기준으로 captopril + hydrochlorothiazide(50 + 25mg) 와의 평균차이를 평가하였을 때 0.3mmHg(90% CI : -1.1, 1.7)이었으며, analapril+hydrochlorothiazide (20+12.5mg)와의 차이는 1.1 mmHg(90% CI : -0.4, 2.6)로 관찰되어 모두 미리 정의된 신뢰구간인 (-4mmHg, 4mmHg) 내에 완전하게 포함되므로 이들 세 혼합약물들의 동등성을 인정하고 있다.

Turik 와 Johns¹⁷⁾은 amoxicillin 치료에 실패한 급성 중이염을 앓고 있는 어린이 중 삼출액이 있는 환자를 대상으로 cefaclor와 cefuroxime axetil의 치료적 동등성을 검정하기 위하여 연구자가 치료내용을 모르도록 단측 눈가림, 무작위, 병행설계 임상시험을 시행하였다. 최종 결과변수는 완전치료와 개선을 포함한 중이염 치료율이었으며, 두 약물간의 치료율 차이(%)를 비교하였다. 이 때 두 약물의 임상적 치료율에 대한 차이의 95% 신뢰구간이 (-15%, 15%) 이내일 경우에 두 약물의 치료적 동등성을 인정하기로 하였다. 치료효과는 치료 종료 후 1~6일 후와 10~16일 후로 나누어 평가하였다. 그 결과 첫 1~6일 후에는 cefaclor 치료군이 92.3%, cefuroxime 치료군이 89.1%의 치료율을 보였고, 이들 차이에 대한 95% 신뢰구간은 (-4.8%, 11.2%)이었다. 그리고 치료 종료 후 10~16일 후에 관찰된 결과는 cefaclor 치료군이 82.7%, cefuroxime 치료군이 83.2%의 치료율

을 보였고, 이들 차에 대한 95% 신뢰구간은 (-10.8, 9.9%)이었다. 두 결과 모두 미리 정의된 신뢰구간 (-15%, 15%)내에 완전하게 포함되므로 두 약물의 치료적 동등성을 인정하고 있다.

6. 결 론

위약대조 임상시험의 비윤리성을 해결하고, 향후 시행될 의학분업에 대비하기 위하여, 유효성 측면에서는 두 치료제가 서로 유사하지만 안전성, 복용의 편리성, 삶의 질 및 비용 등의 측면에서는 새로운 치료제가 더 나을 것이라고 기대하는 동등성 임상시험의 필요성과 다양한 접근 방법론에 대하여 살펴보았다.

본 연구를 통하여 얻은 결론은 동등성 임상시험의 성공적인 수행을 위하여는 적절한 귀무가설의 설정과 과학적인 근거를 이용한 동등성 인정 한계 등을 임상시험 계획단계에서부터 충분히 고려하여야 한다는 것이다. 이미 언급하였듯이 동등성을 증명하기 위한 임상시험에서의 귀무가설은 전형적인 차이성 임상시험에서 사용되어온 귀무가설과는 달리 미리 설정된 0이 아닌 차이에 대한 귀무가설을 수립하는 것이 적절하다. 그러나 외국에서조차도 이러한 귀무가설에 대한 통계적 검정은 임상시험을 계획하고 분석하는 단계에서 일반적인 통계적 가설과의 구분이 사실상 간과되어져 왔던 것이 사실이며, 임상시험에 대한 역사가 그리 길지 않은 우리 나라의 실정에서는 더욱 그러하였다. 또한 동등성 임상시험을 수행하는 과정에서 자주 발생하는 중대한 오류 중의 하나는 차이성 임상시험을 수행하는 것으로 임상시험을 계획하고 자료를 분석한 다음 치료효과의 차이가 통계적인 유의성을 확보하지 못한 경우에 그 결과를 마치 치료효과가 동등한 것처럼 해석하는 것이다. 이러한 경우는 차이성에 관한 귀무가설을 기각하지 못하게 되므로 그 결과에 대한 해석을 “두 가지 약물간의 치료효과에 있어 유의한 차이를 입증하지 못하였다”로 국한하여야지 “두 가지 약물의 치료효과가 동등하다”로 확대 해석하면 안된다는 것이다. 그러므로 저자들은 이러한 오류의 발생을

최소화시키고 지금까지 간과되어져 왔던 동등성 임상 시험에 대한 여러 가지 방법론적 문제의 정리를 통하여 향후 수행될 동등성 임상시험에 일조할 수 있기를 기대한다.

참 고 문 헌

1. Rothman KJ, Michels KB. *The continuing unethical use of placebo controls.* *N Engl J Med* 1994;331:394-8
2. Taubes G. *Use of placebo controls in clinical trials disputed.* *Science* 1995;267:25-6
3. Jones B, Jarvis P, Lewis JA, Ebbutt AF. *Trials to access equivalence: the importance of rigorous methods.* *BMJ* 1996;313:36-9.
4. Ebbutt AF, Frith L. *Practical issues in equivalence trials.* *Stat Med* 1998;17:1691-701
5. Blackwelder WC. "Proving the null hypothesis" in *Clinical trials.* *Controlled Clin Trials* 1982;3:345-53.
6. Blackwelder WC. *Sample size and power for prospective analysis of relative risk.* *Stat Med* 1993;12(7):691-8
7. Garbe E, Rohmel J, Gundert-Remy U. *Clinical and statistical issues in therapeutic equivalence trials.* *Eur J Clin Pharmacol* 1993;45:1-7.
8. CPMP Working Party. *'Biostatistical methodology in clinical trials in applications for marketing Guidance III/3630/92-EN,* *Stat Med* 1995;14:1658-82
9. Rodary C, Com-Nougue C, and Tourade M. *'How to establish equivalence between treatments: A one-sided clinical trial in paediatric oncology.'* *Stat Med* 1989;8:593-8
10. Farrington CP and Manning G. *'Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk.'* *Stat Med* 1990;9:1447-54
11. Westlake WJ. *Statistical aspects of comparative bioavailability trials.* *Biometrics* 1979;35(1):273-80
12. Gart J and Nam JM. *Approximate interval estimation of the difference in binomial parameters: Correction for skewness and extension to multiple tables.* *Biometrics* 1990;46:637-43.
13. Nam JM. *Sample size determination in stratified trials to establish the equivalence of two treatments.* *Stat Med* 1995;14(18):2037-49
14. Cox DR. *Statistical significance tests.* In: *Chaput de Saintong DM, Vere DW, Eds. Current Problems in Clinical Trials.* 1984;41-7.
15. Hutzelmarm J, Owens S, Shedden A, Adamsons I, Vargas E. *Comparison of the safety and efficacy of the fixed combination of dorzolamide/timolol and the concomitant administration of dorzolamide and timolol: a clinical equivalence study.* *International Clinical Equivalence Study Group. Br J Ophthalmol* 1998;82(11):1249-53
16. Roger L, Peter SS, Tullio DP, Jose R, Isabell L, Yves B, Claude C and David G. *An equivalence study of the safety and efficacy of a fixed-dose combination of perindopril with indapamide versus fixed-dose combination of captopril with hydrochlorothiazide and enalapril with hydrochlorothiazide in the treatment of the hypertension.* *J Hypertension* 1995;13: 1847-51
17. Turik MA, Johns D Jr. *Comparison of cefaclor and cefuroxime axetil in the treatment of acute otitis media with effusion in children who failed amoxicillin therapy.* *J Chemother* 1998;10:306-12