

## 바이오시밀러 개발에 필요한 통계방법들에 대한 고찰

강승호<sup>1</sup> · 남주선<sup>2</sup>

<sup>1</sup>연세대학교 응용통계학과

<sup>2</sup>이화여자대학교 통계학과, 식품의약품안전청 바이오생약국 바이오의약품정책과

(2011년 12월 접수, 2012년 1월 수정, 2012년 1월 채택)

---

### 요약

최근 들어 바이오시밀러에 대한 국내외 관심이 매우 증가하고 있다. 바이오시밀러가 오리지널 생물의약품과 효능과 안전성이 유사함을 보이기 위해서는 최종적으로 임상시험을 수행하여야 한다. 본 논문에서는 이러한 임상시험의 수행과 통계적 분석에 필요한 여러 방법들과 외국의 사례 그리고 관련된 가이드라인들을 살펴볼 것이다.

주요어: 생물의약품, 생물학적동등성, 동등성시험.

---

### 1. 서론

의약품은 약사법상 사람이나 동물의 질병을 진단·치료·경감·처치 또는 예방할 목적으로 사용하는 물품 중 기구·기계 또는 장치가 아닌 것 또는 사람이나 동물의 구조와 기능에 약리학적(藥理學的) 영향을 목적으로 사용하는 물품 중 기구·기계 또는 장치가 아닌 것을 의미한다. 이 의약품에는 크게 합성의약품과 생물의약품으로 구분할 수 있는데, 합성의약품은 쉽게 표현하여 우리에게 친숙한 알약이나 캡슐처럼 화학적으로 합성하여 만들어진 약을 말할 수 있다. 반면 생물의약품은 사람이나 다른 생물체에서 유래된 것을 원료 또는 재료로 하여 제조한 의약품으로 보건위생상 특별한 주의가 필요한 의약품을 말하며 생물학적제제, 유전자재조합의약품, 세포배양의약품, 세포치료제, 유전자치료제 등을 포함한다.

합성의약품의 복제약을 만드는 경우에는, 복제약과 오리지널 약(brand-name drug)이 혈액 내에 흡수되는 정도와 비율이 같은지를 보이는, 생물학적 동등성 시험을 실시하여, 혈액 내에 흡수되는 정도와 비율이 같음을 증명해야 심사기관으로부터 판매승인을 받을 수 있다. 복제약은 연구개발비가 들지 않았으므로, 오리지널 약보다 가격이 저렴하며, 정부 입장에서는 의료보험 재정에 부담을 덜 수 있다는 장점이 있다.

또한, 블록버스터의 생물의약품의 특허만료기간이 다가옴에 따라, 국내외 제약업계에서는 생물의약품의 복제약을 개발하고 있으며, 이런 의약품을 바이오시밀러(Biosimilar)라 한다. 이를 국내 규정에서는 동등생물의약품이라 명명하며, 이미 제조판매·수입품목 허가를 받은 품목과 품질 및 비임상·임상적 비교동등성이 입증된 생물의약품으로 정의하고 있다 (식품의약품안전청, 2009). 이미 유럽에서는 특허만료된

---

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2010-0009224). 본 논문의 내용은 저자들의 개인적인 의견으로 우리나라 심사기관의 공식의견을 반영하는 것은 아니다.

<sup>1</sup>교신저자: (120-749) 서울시 서대문구 연세로50, 연세대학교 응용통계학과, 교수.

E-mail: [seungho@yonsei.ac.kr](mailto:seungho@yonsei.ac.kr)

표 1.1. 주요국의 바이오시밀러 허가현황

허가국	제품명	성분명	제조사	허가일
유럽	Omnitrope	소마트로핀	Sandoz	2006.04.12.
	Valtropin	소마트로핀	BioPartners	2006.04.24.
	Binocrit	에포에틴 알파	Sandoz	2007.08.28.
	Epoetin alfa Hexal	에포에틴 알파	Hexal	2007.08.28.
	Abseamed	에포에틴 알파	Medice Arzneimittel Putter	2007.08.28.
	Retacrit	에포에틴 제타	Hospira	2007.12.18.
	Silapo	에포에틴 제타	Stada	2007.12.18.
	Tevagrastim	필그라스티姆	Teva	2008.09.15.
	Ratiograstim	필그라스티姆	Ratiopharm	2008.09.15.
	Filgrastim Ratiopharm	필그라스티姆	Ratiopharm	2008.09.15.
	Biograstim	필그라스티姆	CT Arzneimittel	2008.09.15.
	Filgrastim Hexal	필그라스티姆	Hexal	2009.02.06.
	Zarzio	필그라스티姆	Sandoz	2009.02.06.
	Nivestim	필그라스티姆	Hospira	2010.06.08.
	일본	Somatropin BS	소마트로핀	Sandoz
Epoetin Alfa BS		에포에틴 카파 (에포에틴 알파 바이오시밀러-1)	JCR	2010.01.20

표 1.2. 세계 바이오시밀러 시장규모 예측 (항체의약품 제외) (단위: 억원)

국가	2009년	2011년	2013년	2015년	2017년	2019년
미국	382	821	9,672	16,342	14,632	14,486
유럽	1,338	2,362	3,644	4,161	4,364	4,746
일본	11	90	146	191	281	360
합계	1,732	3,273	13,463	20,694	19,277	19,592

생물의약품에 대하여 바이오시밀러가 개발 되어왔으며, 식품의약품안전청 보도자료에 따르면, 유럽에는 이미 14개, 일본은 2개의 바이오시밀러가 허가된 상태이다 (표 1.1 참조). 또한 세계 바이오시밀러의 시장의 성장률은 2014년 이후 급속도로 성장할 것이라 예상을 하고 있다 (식품의약품안전청, 2011).

표 1.2와 같이 항체의약품을 제외한 바이오시밀러 시장은 급속도로 성장하고 있음에도 불구하고, 국내외 많은 회사들이 바이오시밀러에 관심을 보이는 진정한 이유는 전 세계적으로 블록버스터라 불리우는 엄청난 액수가 팔리는 생물의약품의 특허가 몇 년 내에 만료되어서, 다른 회사들도 블록버스터 생물의약품의 바이오시밀러를 만들어 팔 수 있기 때문이다. 표 1.3에는 몇 년 내에 특허가 만료되는 생물의약품의 이름, 원개발사, 연매출액이 요약되어 있다 (식품의약품안전청, 2009).

바이오의약품은 암, 류마티스관절염 등 난치성, 만성질환에 좋은 치료효과를 기대할 수 있으나, 고가의 치료비용이 소요된다. 표 1.4에는 바이오의약품의 연간 치료비용이 요약되어 있다.

이에 바이오시밀러의 개발은 난치성, 만성질환 환자에 대한 치료기회를 확대시킬 수 있을 뿐만 아니라, 의료비 부담을 경감시킬 것이라 기대하고 있다. 이에 우리나라 또한 ‘신성장 스마트 프로젝트’ 중 하나로 바이오의약산업을 지정하고 삼성전자 등 총 4개의 컨소시엄에 지원하고 있다. 또한 식품의약품안전청에서는 바이오시밀러의 신속한 산업화 촉진을 체계적으로 지원하기 위하여 ‘바이오시밀러(동등생물의약품) 민·관 실무협의체’를 구성하여 운영하고 있다. 전체 10개의 실무협의체가 구성되었고, 동 협의체를 통해 바이오시밀러의 개발 초기부터 품목허가에 이르기까지 품질, 비임상, 임상, 제조 및 품질관리 등을 포괄하는 전 주기적인 허가 지원 프로그램을 운영하고 있다 (식품의약품안전청, 2011). 바이오시

표 1.3. 몇 년 내에 특허가 만료되는 생물의약품

제품명	성분명	개발사	적응증	시장규모 (억달러/년)	특허만료 (미국)
엔브렐	에타너셉트	암젠	류마티스관절염 등	50	2012
에포젠	에포에틴알파	암젠	빈혈	53	2013
레미케이드	인플릭시맵	존슨앤존슨	류마티스관절염 등	44	2013
에포넥스	인터페론베타-1a	바이오제아이텍	다발성경화증	18	2013
레비프	인터페론베타-1a	세로노	다발성경화증	16	2013
휴마로그	인슐린리스프로	릴리	당뇨병	14	2013
뉴포젠	필그라스티م	암젠	호중구감소증	12	2013
세레자임	이미글루세라제	젠자임	고셔병	11	2013
리특산	리특시맵	제넨텍	비호지킨림프종 등	45	2015
란투스	인슐린글라진	사노피아벤티스	당뇨병	27	2015
휴미라	아달리무맵	애보트	류마티스관절염 등	30	2016

표 1.4. 바이오의약품의 연간 치료비용

	제품	적응증	1인당 연간 치료비용
미국	엔브렐, 레미케이드, 휴미라	류마티스관절염	\$15,000-22,450
	아보넥스, 레비프, 베타페론	다발성경화증	\$12,700-28,400
	아바스틴, 허셉틴, 엘비투스	암	\$5,500-32,500
	페가시스, 페그인트론	C형간염	\$6,400-15,700
	포스테오	골다공증	\$6,700
	노보린, 란투스, 휴마로그	당뇨병	\$1,100-2,000
한국	엔브렐	류마티스관절염	약 1,400만원
	리특산	비호지킨림프종	약 1,700만원
	허셉틴	유방암	약 3,800만원
	레미케이드	류마티스관절염	약 1,600만원
	아바스틴	직장암	약 5,100만원(비보험)
	엘비투스	대장암	약 6,000만원(비보험)

밀러에 대한 규정은 유럽이 불과 몇 년 전에, 그리고 우리나라는 2009년 7월에 만들었고, 미국은 지금 만들고 있는 중이다. 그 만큼 바이오시밀러는 이제 막 시작이고, 앞으로 더 많은 논의가 있으리라 예상된다.

## 2. EMEA의 접근방법

바이오시밀러는 생물의약품의 복제약이므로 가장 먼저 떠오르는 생각은, 합성의약품의 복제약을 평가하는데 사용하는 생물학적동등성 시험을 바이오시밀러에도 그대로 적용이 가능 하느냐는 것이다. EMEA는 이러한 과제를 가장 먼저 해결하고자 시도한 심사기관인데, 결론은 적용이 가능하지 않다는 것이다 (Fox, 2010). 그 이유는 합성의약품은 일반 화학구조를 이용하여 합성하는 의약품인데 반해, 생물의약품은 살아있는 생물체(세포)를 이용하여 제조하여 그 제조방법 등이 환경에 민감하기 때문이다. 또한 생물의약품 복용 이후 상당한 시간이 지난 후에야 평가할 수 있는 면역원성(immunogenicity)과 같은 항목들도 있기 때문이다.

또한 EMEA는 생물의약품 복제약은 오리지널 생물의약품과 동일할 수 없으며, 단지 비슷(similar)할 수만 있다고 결론 내렸다. 그래서 EMEA는 생물의약품 복제약, 즉 바이오시밀러(biosimilar)의 승인

을 위한 새로운 제도를 마련하였다 (EMA, 2006a, 2006b). 이 제도의 핵심 내용은 바이오시밀러 제조회사는 비임상시험과 임상시험을 실시하여, 바이오시밀러와 오리지널 생물의약품의 효능과 안전성이 비슷하다는 것을 보이는 것을 요구하고 있다. 생물의약품은 각기 고유의 특성이 매우 달라서, EMA는 바이오시밀러의 품질, 비임상시험 및 임상시험에 대한 전반적인 가이드라인을 제시한 후, 각 제품별(erythropoietin, granulocyte colony-stimulating factor(G-CSF), insulin, human growth hormone, interferon alpha, low-molecular-weight heparin)로 별도의 가이드라인을 제시하였다 (EMA, 2006c, 2006d, 2006e, 2006f, 2006g, 2007, 2009). 다시 말하면, EMA가 내린 결론은 모든 바이오시밀러에 적용될 수 있는 한 가지 기준은 만들 수 없다는 것이다. 왜냐하면 각각의 바이오시밀러는 benefit/risk profile, 유효성을 평가할 대리변수의 존재여부, 부작용의 빈도, 적응증의 넓이와 속성 등에서 매우 다양하기 때문이다. 그렇기 때문에 EMA는 각각의 바이오시밀러 제품별로 가이드라인을 만든 것이다 (Fox, 2010).

EMA는 바이오시밀러와 오리지널 생물의약품의 유사성을 보이는 방법으로 비열등성 시험이 아니라, 동등성 시험을 채택하였다. 즉 바이오시밀러는 오리지널 생물의약품에 비하여 열등해도 안 되지만, 더 우월해서도 안 된다는 것이다. 이는 우리나라가 채택한 방법과 동일하다 (식품의약품안전청, 2009). 동등성 시험에 대해서는 다음 절에서 더 자세하게 다룰 것이다. 동등성 시험과 비열등성 시험의 차이에 대해서는 강승호 (2010)의 1.9절에 더 자세하게 설명되어 있다. EMA가 채택한 바이오시밀러 평가방법은 캐나다, 일본, 호주, 스위스, WHO 등에 의해서도 비슷하게 채택되었다.

### 3. 동등성시험

동등성 검정은 두 치료제(treatment) 사이에 임상적으로 의미 있는 차이가 없음을 보이는 것을 목표로 한다. 설명의 편이상 주평가변수가 연속형 변수이고, 시험약과 대조약의 모평균을 각각  $\mu_1$ 과  $\mu_2$ 로 표기하는 경우, 동등성 시험의 귀무가설과 대립가설은 다음처럼 주어진다.

$$H_0 : |\mu_1 - \mu_2| \geq \Delta \quad vs. \quad H_A : |\mu_1 - \mu_2| < \Delta, \quad (3.1)$$

여기서  $\Delta$ 는 양수로서 동등성 마진이라고 불리는데, 주평가변수에서  $\Delta$  차이는 임상적으로는 의미있는 차이가 아님을 나타낸다. 동등성 시험에서 통계 비전문가가 범하기 쉬운 실수는 가설은 다음처럼 세운 후

$$H_0 : \mu_1 = \mu_2 \quad vs. \quad H_A : \mu_1 \neq \mu_2 \quad (3.2)$$

귀무가설을 채택하게 되면 두 치료제가 동등함을 보였다고 주장하는 경우이다. 하지만 이는 통계학적으로 오류이다. 그 이유는 임상시험을 통하여 입증하고 싶은 가설을 대립가설로 놓아야 옳은 방법이기 때문이다. 식 (3.2)에 있는 가설은 우월성 임상시험에 적절한 가설형태이다. 식 (3.2)의 형태의 가설을 세운 경우 만일 귀무가설이 채택되었다면, 이는 대립가설인  $H_A : \mu_1 \neq \mu_2$ 을 지지할 증거가 부족하다는 말이지, 귀무가설인  $H_0 : \mu_1 = \mu_2$ 에 대한 증거가 존재함을 의미하지는 않는다.

이는 검정력을 이용하여 설명할 수 있다. 검정력이란 대립가설이 참일 때 대립가설을 채택하게 되는 확률로서, 다른 요소(1종 오류 확률, 시험약의 약효의 크기, 모분산 등)들이 고정되어 있는 경우, 표본크기가 증가하면 검정력도 증가하게 된다. 이는 다시 말하면 만약 표본크기가 작다면, 비록 대립가설이 참이라 하더라도 검정력이 부족하여, 대립가설을 채택하지 못하고 귀무가설을 채택하게 된다는 의미이다. 즉 만일 동등성 시험에서의 가설을 식 (3.2)처럼 세우게 되면, 표본크기만 작게 하면 검정력이 작아져서 대립가설이 참인 경우에도 귀무가설인  $H_0 : \mu_1 = \mu_2$ 을 채택하게 되는 모순이 발생하게 된다. 표본크기

가 작으면  $\mu_1$ 과  $\mu_2$ 가 다름에도 불구하고, 이를 보여줄 기회가 적어서 귀무가설을 채택하게 되는 현상이 발생하게 되는 것이다.

이러한 현상은 2종의 오류를 사용하여서도 설명이 가능하다. 검정력이란 100%에서 2종 오류 확률을 빼 값이므로, 검정력이 작다는 것은 2종 오류가 크다는 의미이다. 2종의 오류란 대립가설이 참임에도 불구하고 귀무가설을 채택하게 되는 오류이므로, 표본크기가 작으면 2종 오류가 증가하여 대립가설이 참임에도 불구하고 귀무가설인  $H_0 : \mu_1 = \mu_2$ 을 채택하게 될 가능성이 높아지게 되는 것이다.

식 (3.2)처럼 가설을 세운 후 귀무가설을 채택하게 되면 두 치료제가 동등함을 보였다고 주장하는 것이 잘못되었음을 ICH E9은 다음처럼 지적하고 있다 (ICH E9, 1998).

“Concluding equivalence or non-inferiority based on observing a non-significant test result of the null hypothesis that is no difference between the investigational product and the active comparator is inappropriate.”

#### 4. EMEA의 바이오시밀러 승인 실제 사례에 대한 분석

지금까지 바이오시밀러에 대한 승인 경험이 가장 많은 EMEA가 판매 승인한 바이오시밀러에 대해 어떠한 임상시험이 수행되고 어떠한 통계방법이 사용되었는지를 살펴보는 것은 매우 유익할 것이다. 바이오시밀러를 만드는 회사들은 오리지널 생물의약품을 개발한 회사의 제조 공정 노하우에 대한 자세한 정보를 얻을 수 없기 때문에, 바이오시밀러에 생산된 제조 공정은 오리지널 생물의약품을 생산하는데 사용된 제조 공정과는 다소 다를 수 밖에 없다. 바이오시밀러의 질적 특성은 어떤 제조 공정에서 생산되느냐에 따라 매우 달라지기 때문에, 바이오시밀러와 오리지널 생물의약품간에 어느 정도의 구조적 차이가 존재할 것이라는 것은 예상되는 일이다. 그러므로 철저한 물리화학적 생물학적 비교분석이 바이오시밀러의 유사성을 증명하는 첫 번째 단계가 될 것이다. 그 후 EMEA 가이드라인에 의하면, 비임상시험, PK/PD 시험, 효능과 안전성을 확인하는 임상시험이 순서대로 실시되어야 한다. 본 논문에서는 임상시험에서 발생하는 통계적 이슈들을 중점적으로 살펴볼 것이다.

본 논문에서 사용할 사례는 2008년 9월 EMEA로부터 바이오시밀러로 판매승인을 받은 Filgrastim Raiopham에 대한 임상시험 사례이다. Filgrastim Raiopham은 암환자들에게 세포독성치료제를 투여 후 종종 발생하는 백혈구 감소증 회복에 사용되는 성장인자(G-CSF; granulocyte colony-stimulating factor)로서, 오리지널 생물의약품은 Amgen의 Neupogen이었다. 이 바이오시밀러와 오리지널 생물의약품의 효능과 안전성을 평가하기 위하여 세 번의 임상시험이 수행되었다. 이절에서 살펴보는 내용들은 EMEA가 2008년에 작성한 보고서에 근거한 것이다 (EMEA, 2008). Filgrastim Raiopham의 판매승인을 위하여 모두 세 번의 임상시험(study XM02-02-INT, study XM02-03-INT, study XM02-04-INT)이 각각 유방암 환자, 폐암환자, non-Hodgkin's lymphoma를 대상으로 실시되었다. study XM02-03-INT와 study XM02-04-INT는 위약군 없이 시험군과 대조군만을 비교한 반면, study XM02-02-INT에는 추가로 위약군까지 포함되어 있다. study XM02-02-INT에 더 많은 통계적 이슈들이 포함되어 있으므로, 본 논문에서는 study XM02-02-INT만을 살펴볼 것이다.

임상시험 study XM02-02-INT은 유방암 환자들을 대상으로 실시되었다. 10개국 52개 센터에서 수행된 다국적 다기관 임상시험이었으며, 시험군, 대조군, 위약군 이렇게 세 그룹에 피험자들이 2:2:1의 비율로 무작위배정 되었다. 위약에 배정된 피험자는 첫 번째 사이클 후에 시험약 그룹으로 배정되었고, 총 4번의 사이클이 투여되었다. 주평가변수는 첫 번째 사이클 동안 심각한 백혈구 감소증이 지속된 기간(DSN; duration of severe neutropenia)이다. 이 임상시험에서 통계적으로 주목할 만한 사항은 다음과 같다.

표 4.1. 시험약과 위약에 대한 공분산분석

변동	df	p-value	least square means			
			시험약	대조약	점추정치	95% 신뢰구간 상한
baseline ANC	1	0.6039				
country	8	0.0145				
therapy	1	0.8642				
treatment	1	<0.0001	1.141	3.823	-2.682	-2.151

## (a) assay sensitivity

바이오시밀러와 오리지널 생물의약품이 유사함(similar)을 임상시험을 통하여 보이기 위해서는, 시험군(바이오시밀러)과 대조군(오리지널 생물의약품)만을 두고, 동등성 시험을 실시하면 될 것 같지만 항상 그렇지는 않다. 그전에 확인해야 하는 것이 바로 assay sensitivity이다. 질병 분야에 따라 대조약이 위약보다 우월함을 과거에 실시된 임상시험을 통하여 입증하였음에도 불구하고, 현재의 임상시험에서는 그 우월함이 재현되지 않는 경우가 있음이 경험적으로 알려져 있다. 만일 현재의 임상시험에서 대조약이 위약보다 우월함이 입증되지 않는다면, 설사 시험약과 대조약이 동등함을 보인다고 해도 이는 시험약이 위약과 별반 차이가 없음을 의미하게 된다. 시험약의 효능이 위약과 별 차이가 없다면 그런 시험약을 판매 승인할 수 없음은 당연한 일이다. 그러므로 시험약과 대조약의 동등함을 평가하기 이전에, 먼저 현재 임상시험에서 시험약 (또는 대조약)이 위약보다 우월함을 확인하여야 한다. 이러한 필요에 의하여 위약군이 포함되어 있는 것이다.

ICH E10에 의하면 assay sensitivity의 정의는 다음과 같다 (ICH E10, 2001).

“Assay sensitivity is a property of a clinical trial defined as the ability to distinguish an effective treatment from an ineffective treatment.”

assay sensitivity를 염려하는 이유는, 비록 시험약이 대조약과 동등함에도 불구하고, 시험약이 위약에 비하여 효능이 우월하지 않은 경우가 발생할 가능성이 있기 때문이다. 그러므로 이러한 우려를 불식시킬 수 있는 가장 확실한 방법은, 현재 임상시험에 위약군을 포함시키고, 시험약이 위약보다 효능이 우월함을 직접 보이는 것이다. 표 4.1에는 assay sensitivity를 평가하기 위하여, 즉 시험약(Filgrastim Raiopham)이 위약보다 우월함을 입증하기 위하여 full analysis set에 대해 공분산분석을 사용하여 분석한 결과가 요약되어 있다.

표 4.1을 보면 시험약과 위약의 차이에 대한 95% 신뢰구간의 상한이 -2.151로서 0보다 작으므로, 시험약의 효능은 5% 유의수준에서 위약보다 우월하다고 결론 내릴 수 있다. 다시 말하면 이 임상시험은 assay sensitivity를 갖추었다고 말할 수 있다.

윤리적인 이유 때문에 모든 임상시험에서 assay sensitivity를 확인하기 위하여 위약군을 항상 포함시킬 수 있는 것은 아니다. 위약군을 포함시키지 않은 경우에도 간접적인 방법을 통하여 assay sensitivity가 만족되는지를 확인하여야 한다. 이러한 내용들은 강승호 (2010)의 9.3절에 더 자세하게 설명되어 있다.

## (b) 공변량보정

두 모평균에 대한 차이에 대한 검정을 실시할 때, 단순히 이표본 *t*-검정을 사용하는 대신, 주평균변수에 영향을 미치리라 예상되는 공변량을 추가하여 공분산분석을 실시하면 더 유의한 결과를 얻을 가능성이 높아짐은 강승호 (2010) 6장에서 지적한 바 있다. 이 임상시험에서는 baseline ANC(absolute neutrophil count), country, therapy가 공변량으로 사용되었다. 강승호 (2010) 6.3절에서도 강조한 바와 같이 여기서 명심할 것은 공분산분석에 사용할 공변량을 사전에 임상시험계획서에 명기해야만 한다는

표 4.2. 시험약과 대조약에 대한 동등성 분석

	변동	df	p-value	신뢰구간	
				하한	상한
PP	baseline ANC	1	0.6245		
	country	9	0.0042		
	therapy	1	0.7583		
	treatment	1	0.8305	-0.262	0.325
FA	baseline ANC	1	0.4400		
	country	9	0.0034		
	therapy	1	0.4183		
	treatment	1	0.8508	-0.261	0.316

점이다. 임상시험이 완료되고 모든 분석을 다 마친 후 공변량을 추가로 공분산분석에 넣으면 분석자의 입맛에 맞는 결과가 나오도록 결과를 조작할 수 있는 가능성이 있기 때문이다.

표 4.1에서 주의 깊게 볼 또 하나의 사항은 baseline ANC와 therapy가 5% 유의수준에서 유의하지 않음에도 불구하고, 제거하지 않고 그대로 공분산분석 모형에 두고 있다는 점이다. 이는 강승호 (2010) 6.3절에서 지적한 바와 같이 사전에 임상시험계획서에 공분산분석 모형에 넣겠다고 명기한 공변량은 나중에 얻어지는 통계적 유의성에 관계없이 그대로 모형에 두어야 하기 때문이다.

(c) 동등성 분석

Per protocol set에 대한 시험약과 대조약의 약효의 차이에 대한 95% 신뢰구간을 구한 결과 (-0.262, 0.325)가 표 4.2에 제시되어 있다. 이 신뢰구간은 사전에 정한 동등성 마진 (-1.0, 1.0)안에 들어오므로, 우리는 5% 유의수준에서 시험약과 대조약이 효능 면에서 동등하다고 결론을 내리게 된다. 여기서 주목할 점은 신뢰구간을 구할 때도 여전히 사전에 정한 공변량을 보정하고 있다는 점이다. 게다가 5% 유의수준에서 유의하지 않은 공변량도 공분산분석 모형에서 제거하지 않고 그대로 두고 있다는 점이다.

(d) 다중검정

지금까지 assay sensitivity를 확인하기 위하여 대조약과 위약을 5% 유의수준으로 비교하였고, 그 후에 시험약과 대조약을 5% 유의수준에서 동등성 검정을 수행하였다. 이렇게 5% 유의수준에서 통계적 검정을 여러 번 수행하면 실제로는 유의하지 않음에도 불구하고 우연에 의해 유의한 결과를 얻게 되는 familywise type I error(모든 귀무가설이 참임에도 불구하고 우연에 의해 적어도 하나의 귀무가설을 기각하게 되는 오류)가 5%를 초과하게 됨은 잘 알려진 사실이다 (강승호, 2010, 8.1절). 하지만 임상시험 study XM02-02-INT에서는 이렇게 두 번의 검정을 실시하였지만 familywise type I error는 여전히 5% 이하로 잘 통제된다. 그 이유는 assay sensitivity를 통과한 경우에만 시험약과 대조약의 동등성 검정이 의미가 있으며, assay sensitivity를 보이지 못한 경우에는 시험약과 대조약의 동등성 검정은 의미가 없기 때문이다. 이는 첫 번째 단계(assay sensitivity)를 통과해야만 두 번째 단계(시험약과 대조약의 동등성검정)로 넘어갈 수 있는 구조로서, 이는 gatekeeping 방법의 한 예이기 때문이다. gatekeeping 방법에서는 다중성 보정 없이 각 단계에서 5%의 유의수준으로 가설검정을 하여도 familywise type I error가 5% 이하로 잘 통제됨이 알려져 있다. gatekeeping 방법에 대해서는 강승호 (2010)의 8.14절에 더 자세하게 설명되어 있다. 이 문제를 검정순서의 중요성에 의미를 두지 않고 다른 시각으로 보는 방법은, assay sensitivity를 보이는 것과 시험약과 대조약이 동등함을 보이는 것, 이 두 가지가 모두 만족되어야만 이 임상시험이 성공한 것으로 보는 것이다. 이처럼 두 가지 조건이 모두 만족되어야만 하는 경우에는 다중성 보정 없이 각 조건을 5% 유의수준으로 검정하여도 familywise type I error가 5% 이하로

잘 통제됨이 알려져 있다 (강승호, 2010, 8.8절).

동등성 검정은 표 4.2에서 보듯이 per protocol set과 full analysis set 모두에 대해 수행되었다. 하지만 동등성 검정에서는 per protocol set을 분석한 것을 주분석으로 삼겠다고 사전에 명시했으므로, full analysis set에 대한 분석은 탐색적 분석이 되므로 역시 다중검정에 따른 다중성 보정은 필요하지 않다. 이처럼 탐색적 분석에 대해 다중성 보정이 필요 없는 점에 대해서는 강승호 (2010)의 8.3절에 더 자세하게 설명되어 있다.

(e) 분석군(analysis set)

임상시험에서는 도중에 중도탈락하거나 임상시험계획서를 위반한 피험자들이 발생하기 때문에, 임상시험을 종료한 후에 어느 피험자들을 분석에 포함시켜야 하는지가 중요한 이슈이다. 임상시험 study XM02-02-INT에서 assay sensitivity를 평가하기 위하여 대조약과 위약을 비교하는 것은 우월성 임상시험으로 full analysis set을 분석대상으로 삼았는데, 이는 ICH E9에서 권고하는 바이다 (ICH E9, 1998). full analysis set이란 ITT(intention to treat) 원칙을 가능한 가깝게 지키면서 무작위배정된 피험자들 중에서 분석에서 제외되는 피험자의 수를 최소화하고 제외되는 경우에는 그 타당성을 입증한 피험자들의 집합이다. 우월성 임상시험에서 full analysis set을 주분석군으로 삼는 것이 권고되는 이유에 대해서는 강승호 (2010)의 7.7절에 더 자세하게 설명되어 있다.

반면에 강승호 (2010)의 7.7절에서 언급한 바와 같이 시험약과 대조약을 비교하는 동등성 시험에서는 full analysis set과 per protocol set중에 어느 분석군을 주분석으로 삼아야 하는지에 대해서는 일치된 의견이 존재하지 않는다. per protocol set은 full analysis set의 부분집합으로서 임상시험계획서를 보다 더 잘 지킨 피험자들의 집합이다. 임상시험 study XM02-02-INT에서는 per protocol set을 주분석으로 삼고, full analysis set에 대한 분석은 단지 탐색적 분석으로 삼았다. 왜 per protocol set을 주분석으로 선택했는지에 대한 설명은 보고서에 제시되어 있지 않지만, full analysis set을 선택하는 경우 발생하는 “bias toward the null” 현상이 더 심각한 문제라고 생각했기 때문에, per protocol set을 주분석군으로 삼은 것으로 추측된다 (강승호, 2010, 7.7절).

(f) 맹검(blindness)

확증적 임상시험에서 편의(bias) 발생을 최소화하기 위하여 이중맹검을 사용하는 것은 매우 중요한 일이다 (강승호, 2010, 3.3절). 임상시험 study XM02-02-INT에서는 시험약과 대조약의 투여되는 약물의 투여양이 달라서 완벽한 이중맹검은 가능하지 않았다. 그래서 편의 발생을 최소화하기 위하여 다음의 조치들을 취하였다.

첫째, ANC를 결정하기 위하여 혈액샘플을 뽑고 체온을 측정한 후에 약물이 투여되었다.

둘째, 맹검이 해제된 약물투여자가 정확한 약물을 투여하였다. 약물 투여자에게는 IVRS(interactive voice response system)를 통하여 각 피험자가 시험약을 투여 받는지 대조약을 투여 받는지가 알려졌고, 체중에 맞는 약물의 양이 계산되었다.

셋째, 약물투여자는 약물투여일지에 투약한 약물의 종류, 투여양, batch 번호 등을 기록하였고, 임상연구자는 이 약물투여일지를 열람할 수 없었다.

넷째, “약물투여자 일지”라고 불리는 별도의 문서에, 약물투여자는 측정한 체온, 혈액샘플 등을 기록하였다. 이 별도의 문서에는 투여된 약물의 종류(시험약인지 대조약인지 위약인지)와 투여된 양은 기록하지 않았다. 이 약물투여자 일지는 매일 임상연구자에게 제공되었고, 임상연구자는 이렇게 제공된 정보들에 대해 자신이 검토한 내용을 기록하였다.



(g) 표본크기계산

식 (3.1)에서 주어진 동등성 시험에 대하여, 유의수준  $\alpha$ , 검정력  $1 - \beta$ , 모분산  $\sigma^2$ , 동등성 마진  $\Delta$ , 대립 가설에서의 두 모평균의 차이가  $\mu_1 - \mu_2$ 인 경우 각 그룹당 표본크기 계산공식은 다음과 같음이 알려져 있다. 그러므로  $\alpha = 0.05, \beta = 0.1, \sigma = 1.7, \Delta = 1.0, \mu_1 - \mu_2 = 0.245$ 에 대하여 각 그룹당 109명의 피험자가 필요하게 된다.

$$n = \frac{2(z_{\alpha} - z_{\beta/2})^2 \sigma^2}{(\Delta - |\mu_1 - \mu_2|)^2} = \frac{2 \times (1.645 + 1.645)^2 (1.7)^2}{(1 - 0.245)^2} = 109. \quad (4.1)$$

per protocol set에 대해 20%의 탈락률을 감안하였을 때, 각 그룹당 총 140명의 피험자를 모집하였다.

(h) 안전성 변수들에 대한 분석

안전성 변수들의 통계적 분석에는 단지 기술통계량을 이용한 탐색적 분석만이 사용되었다. 이는 보통의 합성의약품 신약개발을 위하여 실시되는 3상 임상시험에서 수행되는 안전성 변수들에 대한 분석방법과 동일하다. 즉 안전성 변수들에 대한 분석은 확증적 증거를 얻으려는 것이 아니라, 시험약과 대조약 간에 상당한 차이가 날 수도 있다는 신호(signal)를 찾으려는 것이다. 안전성 변수들은 매우 많기 때문에 그 모든 안전성 변수들에 대해 확증적 증거를 얻기 위해서는 너무나 많은 표본크기가 필요하기 때문이다. 만일 안전성 변수들에 대한 탐색적 분석에서 시험약과 대조약 간에 상당한 차이가 날 것으로 우려되는 신호를 주는 중요한 안전성 변수가 있는 경우에는, 심사기관은 그 바이오시밀러의 판매승인을 기각하거나 또는 안전성에 대한 추가 정보를 얻기 위하여 안전성 평가를 주목적으로 하는 추가의 임상시험을 실시할 것을 요구할 수도 있다. 반면에 이러한 탐색적 분석에서 시험약과 대조약 간에 안전성 변수들이 큰 차이가 없는 것으로 관측되면, 일단은 잠정적으로 시험약은 안전하다고 결론내리고, 시험약의 안전성에 대한 자료는 시판 후 감독(post-marketing surveillance)을 통하여 계속 얻게 된다. 시험약의 안전성에 대한 자료는 그 시험약이 시장에 남아 있는 한 계속 수집하게 된다. 만일 시판 후 감독을 통해 얻어진 자료에서 그 바이오시밀러에 심각한 안전성 문제가 있다는 것이 밝혀지면 판매승인이 취소될 수도 있다.

(i) 동등성마진의 결정

동등성 시험에서 동등성 마진의 결정은 아마도 가장 어렵고 논란이 많은 작업일 것이다. 가이드라인에 동등성 마진이 제시되어 있는 경우에는 그 동등성 마진을 그대로 사용할 수 있으므로 동등성 마진을 정하는 일이 이슈가 되지 않는 수 있다. 예를 들어, erythropoietins에 대한 EMEA 가이드라인에는 동등성 마진 값이 다음처럼 명시되어 있다 (EMEA, 2006c).

“If haemoglobin is used as primary endpoint, an equivalence margin of 0.5g/dL is recommended.”

실사 가이드라인에 동등성 마진이 제시되어 있지 않다고 하더라도 이미 EMEA가 승인한 경험이 있는 바이오시밀러들(somatropin, erythropoietin, G-CSF, insulin)의 경우에는 EMEA의 승인에 사용된 동등성 마진을 참고할 수 있을 것이다. 하지만 우리가 정말 관심 있는 바이오시밀러는 조만간 특허가 끝나는 단일클론항체와 같은 훨씬 더 복잡한 생물의약품들이다. 이러한 생물의약품에 대해서는 아직까지 바이오시밀러를 승인한 경험이 전 세계적으로 어디에도 없다. 그러므로 참고할만한 동등성 마진이 없는 실정이다.

이처럼 동등성 마진을 정하기 어려운 경우 심사기관은 제약회사에게 임상시험계획서를 제출할 때 동등성 마진을 제안하고 그 타당성을 제시하라고 할 가능성이 높다. 사실 마진을 정하기 어려운 점은 동등성 시험만이 아니라 비열등성 임상시험에서도 발생하는데, 이 경우 심사기관은 제약회사에게 비열등성 마진을 제안하라고 하고, 일단 제안된 마진에 기초하여 심사기관과 제약회사 간의 여러 번의 논의를 통하여 최종 마진을 정하는 일이 종종있다.

현재 동등성 마진을 정하는 일정한 방법이 존재하지는 않는다. 하지만 식 (3.1)에 제시된 동등성 시험에 대한 가설은 다음처럼 두 개의 비열등성 시험에 대한 가설로 표현할 수 있다.

$$H_{01} : \mu_1 - \mu_2 \geq \Delta \quad vs. \quad H_{A1} : \mu_1 - \mu_2 < \Delta, \quad (4.2)$$

$$H_{02} : \mu_1 - \mu_2 \leq -\Delta \quad vs. \quad H_{A2} : \mu_1 - \mu_2 > -\Delta. \quad (4.3)$$

즉  $H_{A1}$ 과  $H_{A2}$ 가 모두 채택되는 것은 식 (3.1)에 있는 동등성 시험에서 대립가설  $H_A : |\mu_1 - \mu_2| < \Delta$ 가 만족되는 것과 동일하게 된다. 그러므로 비열등성 임상시험에서 비열등성 마진을 정하는데 사용한 방법을, 동등성 임상시험에서 동등성 마진의 하한( $-\Delta$ )을 정하는데 활용해볼 수도 있을 것이다. 그런 측면에서도 볼 때 2010년 미국 FDA가 발간한 비열등성 임상시험에 대한 가이드라인은 중요한 참고문헌이 될 수 있다 (FDA, 2010). 또한 미국 FDA가 발간한 비열등성 임상시험에 대한 가이드라인의 핵심 내용은 강승호 (2010)의 9.5절에 설명되어 있다.

현재 비열등성 임상시험에는 비열등성 마진을 정하는 방법으로 fixed margin method와 synthesis method라는 두 가지 방법이 많이 연구되어 있다 (Tsong, 2007; Hung 등, 2009; Kang과 Yi, 2010). 이 두 방법의 핵심아이디어는 시험약이 여전히 위약보다 효능이 우월하면서 동시에 (위약에 비해 갖는) 대조약 효능의 일정부분(예를 들면 50%)을 시험약이 여전히 갖도록 마진을 정하는 것이다. 예를 들어, 대조약의 치료율이 80%이고 위약의 치료율이 20%인 경우라면, 대조약이 위약에 비해 갖는 효능은  $80\% - 20\% = 60\%$ 이다. 시험약이 위약에 비해 우월하면서 동시에 대조약 효능의 50%를 시험약이 여전히 갖고 싶다면, 시험약은 적어도 50%의 치료율( $20\% + 60\% \times 0.5 = 50\%$ )을 갖도록 비열등성 마진을 정하게 된다. 동등성 마진의 하한을  $-\Delta$ 로 잡았다고 하여 반드시 마진의 상한을  $\Delta$ 로 잡을 필요는 없을 수도 있다. 마진의 상한 결정에는 시험약을 과다 투여 했을 때의 안전성 측면을 고려해야 하므로 임상적 판단이 더 중요한 역할을 하지 않을까 추측한다. 만일 마진의 하한과 상한을 대칭으로 잡지 않고  $(-\Delta_1, \Delta_2)$ 로 잡는다면, 식 (3.1)에 있는 동등성 시험의 대립가설은 다음과 같이 주어지게 된다.

$$H_A : -\Delta_1 < \mu_1 - \mu_2 < \Delta_2, \quad (4.4)$$

여기서  $\Delta_1$ 과  $-\Delta_2$ 는 모두 양의 상수이다.

## 5. 미국 FDA의 바이오시밀러 접근방법

미국은 EMEA보다 바이오시밀러 승인을 위한 법 제정이 늦었는데 2009년 Biologics Price Competition and Innovation Act가 승인되어 바이오시밀러가 승인될 법적인 제도를 마련하였다. Biologics Price Competition and Innovation Act는 EMEA, WHO 그리고 우리나라와는 다르게 바이오시밀러에 대해 두 개의 다른 정의를 제시하고 있다. 첫 번째는 “biosimilar”인데 그 정의는 다음과 같이 기술되어 있다. 이러한 정의는 EMEA, WHO 및 우리나라에서 사용하는 정의와 비슷하다고 판단된다.

“A biological product may be demonstrated to be “biosimilar” if data show that the product is “highly similar” to the reference product notwithstanding minor differences in clinically inactive components and there are no clinically meaningful differences between the biological product and the reference product in terms of safety, purity and potency.”

두 번째는 앞에서 정의된 “biosimilar”보다 더 엄격하게 바이오시밀러와 오리지널 생물약품의 동등성을 요구하는 “interchangeable”이다. “interchangeable” 바이오시밀러는 오리지널 생물약품과 “biosimilar”할 뿐만 아니라, 어떠한 환자에게도 오리지널 생물약품과 같은 임상적 효과를 낼 수 있어야 하며, 또한 1회 이상 반복 투여되는 생물약품의 경우에는 오리지널 생물약품에서 바이오시밀러

로 대체되는 경우 발생하는 안전성과 감소할 효능으로 인해 발생하는 위험도가 바이오시밀러 대체 없이 오리지널 생물의약품을 계속 사용하는 위험도보다 크지 않아야 한다.

“The biological product is “interchangeable” with the reference product if the biological product is biosimilar to the reference product and can be expected to produce the same clinical result as the reference product in any given patient, and for a biological product that is administered more than once to an individual, the risk in terms of safety or diminished efficacy of alternating or switching between use of the biological product and the reference product is not greater than the risk of using the reference product without such alternation or switch.”

Biologics Price Competition and Innovation Act 제정 이후 구체적으로 “biosimilar”와 “interchangeable”한 생물의약품을 어떻게 평가할 것인지에 대해서는 현재 미국 FDA가 가이드라인을 만들고 있고 아직 발표되지는 않은 상태이다. 특히 “interchangeable”한 생물의약품을 평가하기 위하여 많은 연구들이 필요하다고 생각되는데, Chow (2011)은 이러한 생물의약품을 평가하기 위하여 교차시험과 교차시험을 제안하였다.

**6. 현재 바이오시밀러에 대한 통계방법론에 대한 연구들**

바이오시밀러가 중요한 이슈로 대두된 이후에 통계전문가들 사이에서도 바이오시밀러 개발에서 발생할 수 있는 통계적 이슈에 대한 통계방법론에 대한 연구가 수행되고 있다 (Chow 등, 2010; Chow와 Liu, 2010; Chow, 2011). 이러한 연구는 현재 크게 두 가지 방향으로 이루어지고 있다.

첫 번째는 “동등성”을 평가하는 통계적 방법에 대한 연구이다. 사실 동등성을 평가하는 통계적 방법에 대한 연구는 과거 생물학적 동등성 시험에 대한 연구에서 이미 많이 수행되었다. Chow 등 (2010)은 이러한 동등성을 평가하는 기준들이 다음처럼 나누어질 수 있다고 정리하였다. 각각의 기준들은 나름대로의 장단점을 가지고 있으며, 같은 임상시험 결과도 다른 기준을 사용하면 다른 결과가 나올 수 있다.

(a) absolute change versus relative change

예를 들어 시험약과 대조약의 모평균 값을 나타내는  $\mu_T$ 와  $\mu_R$ 이 있는 경우,  $\mu_T - \mu_R$ 의 절대값이 일정한 양의 상수이내에 들어오도록 하는 경우가 absolute change를 사용하는 동등성 검정의 한 예가 된다. 반면  $\mu_T/\mu_R$ 이 1.0 근처에 있도록 하는 경우가 relative change를 사용하는 동등성 검정의 한 예가 된다.

(b) aggregated versus disaggregated

예를 들어 모평균과 모분산을 동시에 고려하여 동등성을 평가하는 경우, 모분산과 모평균의 (예를 들어) 합을 구한 후, 이 합에 대한 기준을 만들면 이런 경우가 aggregated의 한 예가 된다. 반면에 모평균에 대한 기준과 모분산에 대한 기준을 따로 만들면 이런 경우가 disaggregated의 한 예가 된다.

(c) moment-based versus probability-based

(1)에 제시된 가설에서는 동등성을 평가하는 방법으로 두 모평균의 차이가 일정한 범위 안에 들어오는 것을 사용하였다. 모평균은 적률(moment)의 한 예이므로 이러한 방법은 moment-based method라고 불린다. 반면에 probability-based method에서는 확률을 사용하여 동등성을 평가하게 되는데, 다음의 한 예이다.  $T$ 와  $R$ 이 각각 시험약과 대조약에서 관측된 주평가변수를 나타내는 확률변수라고 하면 우선 다음의 확률을 정의하고

$$p_{PB} = P\left(1 - \delta < \frac{T}{R} < 1 + \delta\right). \tag{6.1}$$

이 확률  $p_{PB}$ 를 사용하여 다음의 가설을 검정한다.

$$H_0 : p_{PB} \leq p_0 \quad \text{vs.} \quad H_A : p_{PB} > p_0, \quad (6.2)$$

여기서  $\delta$ 와  $p_0$ 는 사전에 정의된 상수인데  $\delta$ 를 채택하게 되면 동등하다고 결론 내리게 된다. Chow 등 (2010)은 moment-based method와 probability-based method를 비교하는 연구를 수행하였다.

두 번째는 바이오시밀러와 오리지널 생물의약품의 유사성을 평가하는데 사용될 수 있는 적절한 임상시험 디자인에 대한 연구이다. Chow 등 (2010)은 매우 흥미로운 두 가지 디자인을 제안하였다.

이러한 동등성을 평가하는 통계적 방법들이나 디자인이 학문적으로 연구된다고 해도, 이러한 방법들이 모두 다 미국 FDA 등과 같은 심사기관에 의하여 채택되지는 않으리라 생각한다. 그러므로 학문적 연구와 실제 심사기관이 사용하는 방법 간에는 어느 정도 차이가 있을 것이라는 점을 고려하여 학문적 연구를 평가해야 한다고 생각한다. 하지만 이러한 연구결과들이 추후에 심사기관이 가이드라인을 만들어 동등성을 평가하는 통계적 방법이나 디자인을 가이드라인에 소개할 때 영향을 줄 수도 있다. 또는 심사기관이 작성한 가이드라인에는 동등성을 평가하는 명확한 통계적 방법이나 디자인이 제시되어 있지 않고, 가이드라인에는 원칙만 제시되고, 실제로 바이오시밀러를 평가할 때는 제약회사들이 알아서 적절한 방법을 제안하라고 할 수도 있다. 사실 신약의 효능을 평가하는 경우에도, ICH E9은 통계적 방법에 대한 원칙만 제시하고 있고 구체적으로 어떤 통계적 방법을 사용해야 할지는 제약회사들이 알아서 적절한 통계방법을 선택하여 임상시험계획서에 쓰도록 되어 있다. 그러므로 바이오시밀러에서도 비슷한 일이 발생할 수도 있다.

조만간 특허가 끝나며 현재 국내 제약회사들이 만들려고 하는 바이오시밀러들(예를 들면, 단일클론항체)은 각 질병의 특징과 각 생물의약품의 특징이 크게 다르다. 그래서 과연 동일한 한 가지 통계적 방법이 제시될 수 있을지는 현재로서는 알 수 없다. 만일 심사기관이 동일한 한 가지 통계적 방법을 제시하는 것이 어렵다고 판단되면, 심사기관이 발간하는 가이드라인에는 일반적인 원칙만 제시되고 구체적으로 어떤 통계적 방법을 선택하느냐는 제약회사에게 제안하라고 할 수도 있다. 이러한 점들을 감안하면 학문적으로 연구된 바이오시밀러에 대한 다양한 통계방법과 디자인의 장단점에 대해 제약회사들은 상세하게 알아둘 필요가 생길 수도 있다. 이러한 다양한 가능성을 고려하여 현재의 통계적 방법과 디자인에 연구 결과들을 바라보아야 하며, 단지 현재 심사기관이 채택하지 않았다고 하여 그러한 연구들이 무익하다고 판단할 수는 없다.

## 7. 맺음말

앞에서 살펴본 내용들은 지금까지 판매 승인된 바이오시밀러(somatropin, erythropoietin, G-CSF, insulin)에서 발생한 통계적 이슈와 가이드라인에 대한 검토였다. 이러한 바이오시밀러들은 모두 상대적으로 비교적 간단하고 작은 단백질들이었다. 하지만 조만간 특허가 끝나며 현재 국내 제약회사들이 만들려고 하는 바이오시밀러들(예를 들면, 단일클론항체)은 훨씬 더 복잡한 구조를 가진 단백질들이다. 그러므로 지금까지 EMEA가 바이오시밀러에 대하여 만들어 놓은 가이드라인이 상당히 합리적이라는 하지만, 이러한 가이드라인들이 추후에 등장할 훨씬 더 복잡한 바이오시밀러에도 여전히 적절할지는 아직 알 수 없다. 이러한 이유들로 인하여 바이오시밀러에 대한 더 많은 학제간 연구가 필요한 실정이다.

## 참고문헌

- 강승호 (2010). <신약개발에 필요한 의학통계학>, 자유아카데미.  
 식품의약품안전청 (2009). <동등생물의약품 평가 가이드라인>.

- 식품의약품안전청 (2009). 민원설명회, <동등생물의약품(바이오시밀러) 민원설명회>, 2009.7.22.
- 식품의약품안전청 (2011). 보도자료, <식약청, 바이오시밀러 허가 지원 프로그램 본격 가동>, 2011.3.4.
- Chow, S. C. (2011). Scientific factors for assessing biosimilarity and drug interchangeability of follow-on biologics, The FDA/Industry workshop.
- Chow, S. C., Hsieh, T. C., Chi, E. and Yang, J. (2010). A comparison of moment-based and probability-based criteria for assessment of follow-on biologics, *Journal of Biopharmaceutical Statistics*, **20**, 31–45.
- Chow, S. C. and Liu, J. P. (2010). Statistical assessment of biosimilar products, *Journal of Biopharmaceutical Statistics*, **20**, 10–30.
- EMA (2008). Assessment report for Filgrastim Ratiopharm.
- European Medicines Agency (2006a). Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues.
- European Medicines Agency (2006b). Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Quality issues.
- European Medicines Agency (2006c). Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues guidance on similar medicinal products containing recombinant erythropoietins.
- European Medicines Agency (2006d). Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues guidance on similar medicinal products containing recombinant erythropoietins.
- European Medicines Agency (2006e). Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues guidance on similar medicinal products containing recombinant granulocyte-colony stimulating factor.
- European Medicines Agency (2006f). Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues guidance on similar medicinal products containing recombinant human soluble insulin.
- European Medicines Agency (2006g). Annex to guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: Non-clinical and clinical issues guidance on similar medicinal products containing recombinant human soluble somatropin.
- European Medicines Agency (2007). Draft guideline on similar medicinal products containing recombinant interferon alpha.
- European Medicines Agency (2009). Guideline on non-clinical and clinical development of similar biological medicinal products containing low molecular weight heparins.
- Fox, A. (2010). Biosimilar medicines-new challenges for a new class of medicine, *Journal of Biopharmaceutical Statistics*, **20**, 3–9
- Hung, H. M., Wang, S. J. and O'Neill, R. (2009). Challenges and regulatory experiences with non-inferiority trial design without placebo arm, *Biometrical Journal*, **51**, 324–334.
- ICH E9 (1998). Note for Guidance on Statistical Principles for Clinical Trials.
- ICH E10 (2001). Choice of Control Group and Related Issues in Clinical Trials.
- Kang, S.-H. and Yi, T. (2010). Strength of evidence of non-inferiority trials - The adjustment of the type I error rate in non-inferiority trials with the synthesis method, *Statistics in Medicine*, **29**, 1477–1487.
- Tsong, Y. (2007). The utility of active controlled noninferiority/equivalence trials in drug development, *International Journal of Pharmaceutical Medicine*, **21**, 225–233.
- US FDA (2010). Non-Inferiority Clinical Trials. Draft Guidance.
- WHO (2009). Guidelines on evaluation of similar biotherapeutic products.

# Statistical Consideration of the Development of Biosimilar Products

Seung-Ho Kang<sup>1</sup> · Jusun Nam<sup>2</sup>

<sup>1</sup>Department of Applied Statistics, Yonsei University

<sup>2</sup>Department of Statistics, Ewha Womans University, Biopharmaceutical Policy Division, KFDA

(Received December 2011; Revise January 2012; Accepted January 2012)

---

## Abstract

Recent assessments of the biosimilarity of biologic products have received considerable global attention. A clinical trial should be conducted to assess the biosimilarity of a biosimilar product and a innovator biological product. In this paper we will describe several methods for the implementation of clinical trials and statistical analysis, a real international case and related international guidelines.

Keywords: Biological products, bioequivalence, equivalence trial.

---

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No. 2010-0009224).

<sup>1</sup>Corresponding author: Professor, Department of Applied Statistics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea. E-mail: seungho@yonsei.ac.kr