



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

바이오시밀러 임상시험에서 주평가변수가
이분형인 경우 평가방법 비교
: Risk Difference와 Risk Ratio

연세대학교 대학원
의학전산통계학협동과정
의학통계학전공
길 하 영

바이오시밀러 임상시험에서 주평가변수가
이분형인 경우 평가방법 비교
: Risk Difference와 Risk Ratio

지도 남 정 모 교수

이 논문을 석사 학위논문으로 제출함

2017년 12월 일

연세대학교 대학원
의학전산통계학협동과정
의학통계학전공
길 하 영

차 례

표 차례	iii
그림 차례	iv
국문 요약	v
제 1장 서론	1
1.1. 연구 배경 및 목적	1
1.2. 연구 내용 및 방법	4
1.3. 논문의 구성	5
제 2장 이론적 배경	6
2.1. 바이오시밀러(Biosimilar)	6
2.2. 바이오시밀러에 대한 FDA와 EMA의 접근방법 비교	7
2.2.1 EMA의 접근방법	7
2.2.2 FDA의 접근방법	8
2.3. 바이오시밀러의 동등성 평가	9
2.3.1 동등성 시험(Equivalence test)	9
2.3.2 동등성 마진의 결정	10
제 3장 동등성 평가 방법	12
3.1. 동등성 마진	13
3.1.1 RD의 동등성 마진 설정	15
3.1.2 RR의 동등성 마진 설정	15
3.2. 검정통계량을 이용한 동등성 평가방법	16
3.2.1 동등성 평가기준이 RD인 경우	16
3.2.2 동등성 평가기준이 RR인 경우	17
3.3. 신뢰구간을 이용한 동등성 평가방법	19
3.4. 동등성 구간 설정방법	19

제 4장 모의실험	21
4.1. 모의실험 구축	21
4.1.1 RD와 RR의 비교를 위한 모의실험	21
4.1.2 조정된 동등성 구간을 이용한 모의실험	23
4.1.3 동등성 평가방법 간 모의실험	23
4.2. 모의실험 설계	24
4.3. 모의실험 결과	25
4.3.1 RD와 RR의 모의실험 결과 비교	26
4.3.2 제안된 방법의 모의실험 결과	31
4.3.3 동등성 평가방법 간 모의실험 결과 비교	40
제 5장 결론 및 고찰	44
참고문헌	46
영문요약	48

표 차 례

표 1. 주요 항체의약품 특허만료 일정	3
표 2. 미국 FDA의 바이오시밀러 허가 현황	3
표 3. 검정통계량 계산을 위한 2×2 분할표	12
표 4. 동등성 마진 설정을 위한 2×2 분할표	14
표 5. 모의실험을 위해 생성된 자료($\alpha = 0.025, n_1, n_2 = 100$)	23
표 6-1. 10,000번의 반복 실행 중 동등성 평가변수(RD,RR)의 95%신뢰구간이 동등성 구간에 포함된 비율 $\alpha = 0.025$, 동등성구간 : $RD = \pm 0.14, RR = (0.71, 1.40)$	27
표 6-2. 10,000번의 반복 실행 중 동등성 평가변수(RD,RR)의 95%신뢰구간이 동등성 구간에 포함된 비율 $\alpha = 0.025$, 동등성구간 : $RD = \pm 0.09, RR = (0.83, 1.19)$	28
표 7-1. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교 $\alpha = 0.025, repetition = 10,000$, 시험대상자수 : 520명	32
표 7-2. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교 $\alpha = 0.025, repetition = 10,000$, 시험대상자수 : 640명	33
표 8-1. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교 $\alpha = 0.025, repetition = 10,000$, 시험대상자수 : 1,240명	34
표 8-2. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교 $\alpha = 0.025, repetition = 10,000$, 시험대상자수 : 1,540명	35
표 9-1. \hat{p}_0 가 0.2일 때, 동등성 평가변수(RD, RR)의 신뢰구간이 동등성 구간에 포함되는 확률($\alpha = 0.025, repetition = 10,000$)	41
표 9-2. \hat{p}_0 가 0.3일 때, 동등성 평가변수(RD, RR)의 신뢰구간이 동등성 구간에 포함되는 확률($\alpha = 0.025, repetition = 10,000$)	42

그림 차례

그림 1. 바이오시밀러 시장 성장 예측	2
그림 2-1. 바이오시밀러 동등성 평가에 대한 RD와 RR의 검정력 비교 (<i>repetition</i> = 10,000, α = 0.025, <i>placebo</i> = 0.2, <i>control</i> = 0.6) . . .	29
그림 2-2. 바이오시밀러 동등성 평가에 대한 RD와 RR의 검정력 비교 (<i>repetition</i> = 10,000, α = 0.025, <i>placebo</i> = 0.3, <i>control</i> = 0.6) . . .	30
그림 3-1. 검정통계량을 이용한 바이오시밀러 동등성 평가에서 RD와 RR의 검정력 비교(<i>power</i> = 0.8, <i>placebo</i> = 0.2, <i>control</i> = 0.6) . . .	36
그림 3-2. 검정통계량을 이용한 바이오시밀러 동등성 평가에서 RD와 RR의 검정력 비교(<i>power</i> = 0.9, <i>placebo</i> = 0.2, <i>control</i> = 0.6) . . .	37
그림 4-1. 검정통계량을 이용한 바이오시밀러 동등성 평가에서 RD와 RR의 검정력 비교(<i>power</i> = 0.8, <i>placebo</i> = 0.3, <i>control</i> = 0.6) . . .	38
그림 4-2. 검정통계량을 이용한 바이오시밀러 동등성 평가에서 RD와 RR의 검정력 비교(<i>power</i> = 0.9, <i>placebo</i> = 0.3, <i>control</i> = 0.6) . . .	39
그림 5. RD와 RR의 95% 신뢰구간이 동등성 구간에 포함된 분율 비교 . . .	43

국 문 요 약

바이오시밀러 임상시험에서 주평가변수가 이분형인 경우 평가방법 비교 : Risk Difference와 Risk Ratio

바이오시밀러(biosimilar)는 기존 의약품 대비 효능효과 및 안정성에서 동등함을 입증한 복제의약품이다. 바이오시밀러는 고가의 바이오의약품을 대체하여 보다 합리적인 비용으로 환자에게 치료기회를 제공하며, 의료비용 절감효과를 통해 보다 나은 삶의 질을 보장해 줄 수 있다. 2015년 이후 블록버스터급 항체의약품들이 순차적으로 특허가 만료되면서 바이오시밀러는 다수의 제약/바이오 업체들에 의해 개발되고 있으며 글로벌 바이오시밀러 시장의 규모는 2019년에는 239억 달러까지 성장할 것으로 예견되고 있다.

바이오시밀러 산업이 성장함에 따라 바이오시밀러 개발을 위한 임상시험에서 주로 사용되는 동등성 검정의 중요성 또한 대두되고 있다. 특히, 동등성 검정에서 가장 중요하게 여겨지는 것은 두 약품이 동등하다고 인정할 수 있는 한계, 즉 동등성 마진을 설정하는 일이다.

따라서 본 연구에서는 모의실험을 통해 주평가변수가 이분형인 경우, 동등성 평가기준인 RD와 RR의 신뢰구간이 동등성 구간에 포함될 확률을 비교하였다. 모의실험을 통해 Risk Difference와 Risk Ratio의 동등성 구간의 폭에 차이가 존재하는 근본적인 문제를 발견하였고, RD의 동등성 마진에 맞춰 RR의 동등성 마진을 조정하는 새로운 동등성 마진 설정방법을 제안하였다. 제안된 방법을 적용한 모의실험 결과, RD와 RR의 신뢰구간이 동등성 구간에 포함될 확률에 큰 차이가 나지 않았다. 따라서 RD와 RR이 비슷한 가치를 지닌 평가변수이기 위해서는 두 평가변수 간 동등성 마진의 조정이 필요할 것으로 생각된다.

핵심되는 말 : 바이오시밀러, 동등성 평가, Risk Difference, Risk Ratio

제 1장 서론

1.1 연구 배경 및 목적

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 줄 목적으로 사용하는 물품 중 기구·기계 또는 장치가 아닌 것을 의미한다. 의약품은 크게 합성의약품과 바이오의약품으로 구분할 수 있는데, 합성의약품은 질량이 작은 화합물 의약품으로, 우리에게 친숙한 알약이나 캡슐과 같이 화학적으로 합성하여 만들어진 약을 말할 수 있다. 반면 바이오의약품은 미생물, 식물, 동물세포 등 생물체로부터 생산되며, 주로 유전자 재조합(재설계) 방식을 사용한다.

바이오의약품은 합성의약품에 비해 큰 분자량과 복잡한 구조로 제조과정이 복잡하고, 환경에 영향을 크게 받는 등 변화에 민감하여 관리에 유의해야 할 점이 많지만, 고유의 독성이 매우 낮아 높은 안정성을 보이며, 암이나 류마티스 등 난치성, 만성질환에 뛰어난 치료효과를 보인다.

바이오시밀러(biosimilar)는 오리지널 의약품 대비 효능효과 및 안정성에서 동등함을 입증한 복제의약품이다. 바이오시밀러는 고가의 바이오의약품을 대체하여 보다 합리적인 비용으로 환자에게 치료기회를 제공하며, 의료비용 절감효과를 통해 보다 나은 삶의 질을 보장해 줄 수 있다. 또한 의료비용 절감효과는 국가의 의료재정 부담을 낮추며 의료보험 수혜자 확대에 기여 할 수 있다. 이러한 순기능을 지닌 바이오시밀러는 2015년 이후 블록버스터급 항체 의약품들이 순차적으로 특허가 만료되면서 다수의 제약/바이오 업체들에 의해 개발되고 있으며 이에 따라 글로벌 바이오시밀러 시장의 규모는 2019년에는 239억 달러까지 성장할 것으로 예견되고 있다. (그림 1, 정보라, 정은영, 2017)

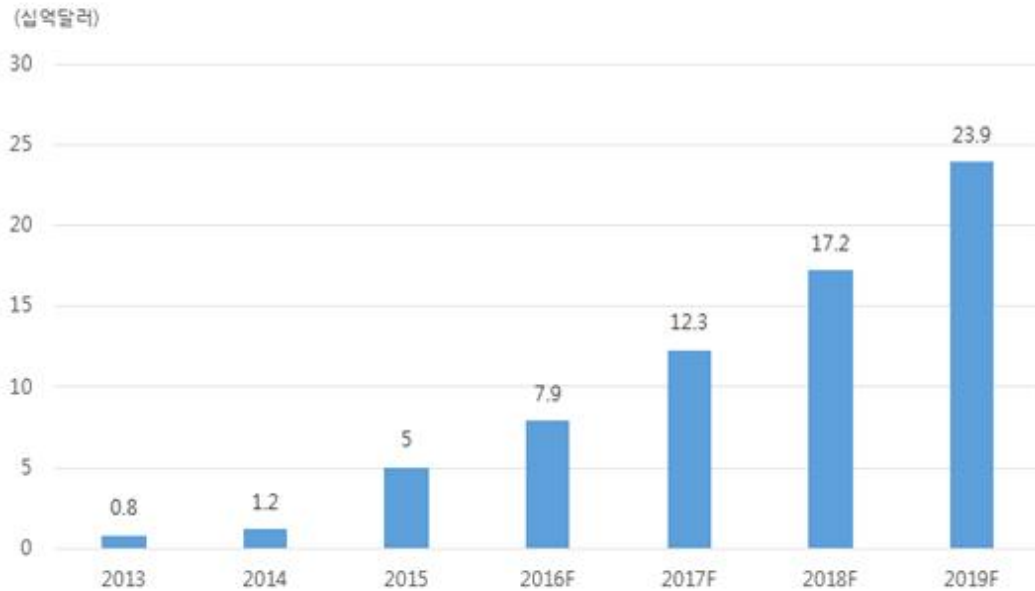


그림 1. 바이오시밀러 시장 성장 예측

2006년부터 바이오시밀러를 적극적으로 도입해 온 유럽과 다르게 자국 의 약산업 보호를 위해 바이오시밀러 도입에 보수적이었던 미국도 2017년 1월 말, 오리지널 의약품을 바이오시밀러로 대체조제 할 수 있는 가이드라인 초안을 발표하는 등 바이오시밀러 점유율 확대에 긍정적이다.(표 2) 또한 바이오시 밀러 글로벌 시장이 본격적인 성장을 시작하는 시점에서 삼성바이오로직스의 KOSPI 상장(2016년 11월 10일), 셀트리온 헬스케어의 KOSDAQ 상장승인이 맞물리면서 바이오시밀러가 제약 업종 내 새로운 대안으로 떠오르고 있다.

표 1. 주요 항체의약품 특허만료 일정 (단위 : 백만 달러)

제품명	개발 및 판매 회사	16년 매출	미국 특허만료	EU 특허만료
휴미라	애브비, 에자이	16,480	16년 12월	18년 04월
엔브렐	암젠, 다케다, 화이자	9,322	28년 11월	15년 02월
레미케이드	J&J, 머크, 미즈비시 타나베	8,790	19년 09월	15년 02월
리툭산	로슈	7,591	18년 07월	13년 11월
아비스틴	로슈	7,053	19년 03월	19년 12월
허셉틴	로슈	7,052	19년 06월	14년 07월
루센티스	노바티스, 로슈	3,297	19년 03월	22년 01월
타이사브리	바이오젠 아이텍	1,964	17년 04월	20년 01월

표 2. 미국 FDA의 바이오시밀러 허가 현황

제품명	개발사	판매사	오리지널제품명 (원개발사)	미국 허가시기
작시오	산도즈	산도즈	뉴포젠 (알젠)	2015.03.06
인플렉트라 /램시마	셀트리온	화이자	레미케이드 (얀센)	2016.04.05
에렐지	산도즈	산도즈	엔브렐(암젠, 화이자)	2016.08.30
암제비타	암젠	암젠	휴미라 (애브비)	2016.09.23
베이사글라	릴리/ 베링거	릴리 /베링거	란투스 (사노피)	2016.12.16

바이오시밀러 산업이 성장함에 따라 바이오시밀러 개발을 위한 임상시험에서 주로 사용되는 동등성 검정의 중요성 또한 대두되고 있다. 특히, 동등성 검정에서 가장 중요하게 여겨지는 것은 두 약품이 동등하다고 인정할 수 있는 한계, 즉 마진을 설정하는 일이다. RD와 RR을 이용한 동등성 마진의 설정은 동등성 평가 결과에 직접적인 영향을 끼치기 때문이다.

2017년 8월, FDA는 류머티스 관절염, 궤양성 대장염 등에 적응증을 갖는 의약품 Humira의 바이오시밀러인 Boehringer사의 Cyltezo의 판매를 승인하였고, EMA는 2017년 11월 Cyltezo의 판매를 승인하였다. 또한, 2013년 9월 EMA는 셀트리온의 Remsima를 승인, FDA는 2016년 4월 Remsima의 판매를 승인하였다.

이렇듯 동일한 바이오시밀러에 대해 두 기관의 판매 승인결과는 크게 다르지 않았다. 따라서 본 연구에서는 동일한 조건(sample size, power, alpha)하에 동등성 평가기준(RD, RR)만 다르게 두고 수행되는 바이오시밀러 동등성평가의 모의실험 결과를 통해 FDA와 EMA에서 주로 사용되는 평가기준(RR(Relative Risk), RD(Risk Difference))이 실제로 동등성 여부 결정에 비슷한 결과를 보이는지 비교하고자 한다.

1.2 연구 내용 및 방법

본 연구에서는 바이오시밀러 동등성 평가에 대한 EMA와 FDA의 가이드라인을 소개하고, 차이를 논하고자 한다. 또한 마진 설정 기준에 따라 마진을 적용한 모형에 대한 모의실험을 통해 바이오시밀러 동등성 평가의 결과를 비교한다. 또한 검정통계량 z 값을 이용한 두 가지 동등성 평가 방법(Miettinen and Nurminen(1985), Farrington and Manning(1990))에서 RD와 RR의 동등성 평가 결과에 차이가 존재하는지 비교하고, 새로운 동등성 마진 설정 방법을 제안한다.

1.3 논문의 구성

제 1장에서는 연구의 배경 및 목적과 연구 내용 및 방법을 소개한다. 제 2장에서는 본 연구의 이론적 배경이 되는 바이오시밀러의 정의와 바이오시밀러의 동등성 평가에 대해 소개하고, 바이오시밀러에 대한 FDA와 EMA의 접근 방법을 비교한다. 제 3장에서는 여러 가지 동등성 평가 방법에 대해 소개한다. 제 4장에서는 다양한 가정 하에 동등성 평가변수 RD와 RR의 신뢰구간을 이용한 모의실험을 진행하고, 실험 결과를 비교·평가한다. 또한 동등성 평가를 위한 검정통계량의 검정력에 관한 모의실험을 수행, 결과를 비교하고, 새로운 동등성 마진 설정 방법을 제안한다. 또한 제안된 방법을 이용하여 RR의 동등성 마진을 조정하고 모의실험을 수행, 결과를 앞선 실험 결과와 비교한다. 제 5장에서는 결론 및 고찰을 제시한다.

제 2장 이론적 배경

2.1에서 FDA, EMA, 한국의 바이오시밀러의 정의에 대해 소개하고, 2.2에서는 바이오시밀러에 대한 FDA와 EMA의 접근 방법을 비교한다. 2.3에서는 바이오시밀러의 동등성 평가에 대해 간략히 정리한다.

2.1 바이오시밀러(Biosimilar)

바이오시밀러는 특허가 만료된 생물약품에 대한 복제약을 말하며, 바이오복제약, 바이오제네릭(biogeneric)이라 부르기도 한다. 바이오시밀러의 경우 기존 약과 동일한 공정으로 제조되지는 않지만 임상실험을 하여 혈액검사 결과가 생물학적으로 거의 동일한 효과를 낼 경우 동등성을 인증받아 바이오시밀러로 명명된다. 동일하지는 않지만 비슷하다는 의미에서 바이오시밀러(biosimilar)라고 한다.

국가별 바이오시밀러의 정의는 다음과 같다.

미국 : “A biosimilar is a biological product that is highly similar and has no clinically meaningful differences from an existing FDA-approved reference product.”(FDA, 2016)

EU : “A biosimilar is a biological medicinal product that contains a version of the active substance of an already authorised original biological medicinal product (reference medicinal product) in the European Economic Area (EEA). Similarity to the reference medicinal product in terms of quality characteristics, biological activity, safety and efficacy based on a

comprehensive comparability exercise needs to be established.”(EMA, 2014)

한국 : “동등성 의약품, 또는 미국의 경우 follow-on biologics로 칭하는 바이오시밀러의 개념은 ‘이미 제조판매, 수입품목 허가를 받은 품목과 품질 및 비임상, 임상적 비교동등성(comparability)이 입증된 생물의약품’이라 정의됨.” (식품의약품안전처, 2014)

2.2 바이오시밀러에 대한 FDA와 EMA의 접근방법 비교

2.2.1 EMA의 접근방법

EMA는 바이오의약품의 복제약인 바이오시밀러를 합성의약품의 복제약을 평가하는데 사용하는 생물학적 동등성 시험에 그대로 적용이 가능한가에 대한 과제를 가장 먼저 해결하고자 시도한 심사기관이다. 해당 과제에 대한 결론은 적용이 가능하지 않다는 것이었다(Fox, 2010). 그 이유는 합성의약품은 일반화학구조를 이용하여 합성하는 의약품인데 반해, 바이오의약품은 살아있는 생물체(세포)를 이용하여 제조되어 환경에 민감하며 바이오의약품 복용 이후 상당한 시간이 지난 후에야 평가할 수 있는 면역원성(immunogenicity)과 같은 항목들이 존재하기 때문이다. 이와 같은 이유로 EMA는 바이오의약품 복제약은 기존의 바이오의약품과 동일할 수 없으며, 단지 비슷(similar)할 수만 있다고 결론지었다.

또한 EMA는 바이오의약품의 복제약, 즉 바이오시밀러의 승인을 위한 새로운 제도를 마련하였다(EMA, 2006a, 2006b). 이 제도의 핵심 내용은 바이오시밀러 비임상시험과 임상시험을 실시하여, 바이오시밀러와 기존 바이오의약품의 효능과 안전성이 비슷하다는 것을 보일 것을 요구하고 있다. 바이오의약품은 각기 고유의 특성이 매우 달라서, EMA는 바이오시밀러의 품질, 비임상시

험 및 임상시험에 대한 전반적인 가이드라인을 제시한 후, 제품별로 별도의 가이드라인을 제시하였다.

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. 따라서 EMA는 각각의 바이오시밀러 제품별로 가이드라인을 만들었다 (Fox, 2010).

EMA는 바이오시밀러와 오리지널 생물의약품의 유사성을 보이는 방법으로 동등성 시험을 채택하였다. 즉 바이오시밀러는 오리지널 생물의약품에 비하여 열등해도 안 되지만, 더 우월해서도 안 된다는 것이다. 이는 우리나라가 채택한 방법과 동일하다 (식품의약품안전청, 2009). EMA가 채택한 바이오시밀러 평가방법은 캐나다, 일본, 호주, 스위스, WHO 등에 의해서도 비슷하게 채택되었다.

2.2.2 FDA의 접근방법

미국은 EMA보다 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 EMA, WHO 그리고 우리나라와 다르게 바이오시밀러에 대해 두 개의 다른 정의를 제시하고 있다.

첫째는 2.1절에서 언급되었던 것과 동일한 정의인 “biosimilar”이며, 두 번째는 “biosimilar”보다 더 엄격한 기준을 두어 바이오시밀러와 기존 바이오의약품의 동등성을 요구하는 “interchangeable”이다. FDA는 바이오시밀러의 interchangeable에 대해 다음과 같이 설명하고 있다.

“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.”(FDA, 2015)

Biologics Price Competition and Innovation Act 제정 이후 구체적으로 “biosimilar”와 “interchangeable”한 생물약품을 어떻게 평가할 것인지에 대해서는 현재 미국 FDA가 가이드라인을 만들고 있으며, 2017년 9월 가이드라인의 초안만 발표된 상태이다.

발표된 초안에 따르면, 바이오시밀러 평가를 위한 세 가지 방법(Equivalence test, Quality Range Approach, Visual Displays)을 제시하고 있지만, 본 논문에서는 equivalence test에 대해서만 논하고자 한다.

2.3 바이오시밀러의 동등성 평가

2.3.1 동등성 시험(Equivalence test)

일차 평가 변수(primary endpoint)가 이분형인 경우, 동등성 평가를 위해 사용되는 평가변수로 RD와 RR이 있다. 동등성 시험의 목적은 두 치료제(treatment) 사이에 임상적으로 의미 있는 차이가 없음을 보이는 것이다. 가설 검정의 특성상 정확히 동등함을 보이는 것은 불가능하므로 주로 다음과 같은 가설을 설정하여 동등성 검정을 시행한다.

먼저, 평가변수가 RD인 경우, 다음과 같은 가설을 설정한다. 이때 p_T 는 바이오시밀러의 event probability이고, p_C 는 기존 의약품의 event probability이며, δ_0 는 지정된 RD의 동등성 마진이다.

$$H_N : |p_T - p_C| \geq \delta_0 \quad \text{vs.} \quad H_A : |p_T - p_C| < \delta_0 \quad (2.1)$$

다음으로 평가변수가 RR인 경우, 가설은 다음과 같으며, ϕ_0 는 지정된 RR의 동등성 마진이다.

$$H_N : \frac{P_T}{P_C} \geq \phi_0 \quad \text{vs.} \quad H_A : \frac{P_T}{P_C} < \phi_0 \quad (2.2)$$

동등성 검정에서 두 치료제의 효과가 완전히 같다는 것을 보이기 힘들기 때문에 동등성 마진을 설정하여 두 치료제의 효과의 차이 여부를 검정한다. 본 연구에서는 두 치료제의 효과의 차(RD) 혹은 치료제 효과의 비(RR)의 신뢰구간이 동등성 구간 내에 포함될 때, 두 치료제 효과 간 통계적으로 의미 있는 차이가 없다고 검정하였다.

2.3.2 동등성 마진의 결정

동등성 시험에서 동등성 마진의 결정은 어렵고 논란이 많은 부분이다. 가이드라인에 동등성 마진이 제시되어 있는 경우에는 그 동등성 마진을 그대로 사용할 수 있으므로 동등성 마진을 정하는 일이 어렵지 않다. 예를 들어, erythropoietins에 대한 EMA 가이드라인에는 동등성 마진 값이 다음과 같이 명시되어 있다.

“If haemoglobin is used as primary endpoint, an equivalence margin of 0.5g/dL is recommended.”(EMA, 2006c)

만약 가이드라인에 동등성 마진이 제시되어 있지 않다고 하더라도 이미 EMA가 승인한 경험이 있는 바이오시밀러(somatropin, erythropoietin, G-CSF, insulin)의 경우에는 EMA의 승인에 사용된 동등성 마진을 참고할 수 있다. 하지만 단일클론항체와 같은 복잡한 생물의약품에 대해서는 아직까지 바이오시밀러를 승인한 경험이 많지 않으므로 참고할만한 동등성 마진이 드물다. 이처럼 동등성 마진을 정하기 어려운 경우, 심사기관은 제약회사에게 임상 시험계획서를 제출할 때 동등성 마진을 제안하고 그 타당성을 제시하라고 할 가능성이 높다.

마진을 정하기 어려운 점은 동등성 시험만이 아니라 비열등성 임상시험에서도 발생하는데, 이 경우 심사기관은 제약회사에게 비열등성 마진을 제안하라고 하고, 제안된 마진에 기초하여 심사기관과 제약회사 간의 여러 번의 논의를 통하여 최종 마진을 결정하기도 한다.

한편, 비열등성 임상시험에서 비열등성 마진을 설정하는데 사용되는 방법은 동등성 임상시험에서 동등성 마진의 하한을 설정하는데 사용되기도 하는데, 이는 아래 제시된 가설에서 H_{A1} 과 H_{A2} 가 모두 채택되는 것은 식 (2.1)에 있는 동등성 시험에서 대립가설 $H_A : |p_T - p_C| < \delta_0$ 가 만족되는 것과 동일한 의미를 가지기 때문이다.

$$\begin{aligned}
 H_{N1} : p_T - p_C \geq \delta_0 & \quad vs. \quad H_{A1} : p_T - p_C < \delta_0 \\
 H_{N2} : p_T - p_C \leq -\delta_0 & \quad vs. \quad H_{A2} : p_T - p_C > -\delta_0
 \end{aligned}$$

제 3장 동등성 평가 방법

이 장에서는 일차 평가 변수(primary endpoint)가 이분형인 경우, 동등성 평가 변수로 사용되는 RR과 RD를 이용한 동등성 평가 방법에 대해 논한다.

3.1절에서 RR과 RD의 동등성 한계 설정 방법에 대해 설명하고, 3.2절에서 검정통계량을 이용한 동등성 평가 방법(Miettinen and Nurminen(1985), Farrington and Manning(1990))에 대해 간략히 정리한다. 또한 3.3절에서 RR과 RD의 신뢰구간을 이용한 동등성 평가방법에 대해 논하고, 3.4절에서 본 연구에서 제안하는 동등성 구간 설정 방법을 소개한다.

먼저, <표 3>은 총 N명 중 두 치료군(실험군, 대조군)의 약물에 대한 성공적인 반응여부에 관한 분포표이다.

표 3. 검정통계량 계산을 위한 2x2 분할표

대상자 군 \ 약물에 대한 반응	성공	실패	총합
	실험군	x_{11}	x_{12}
대조군	x_{21}	x_{22}	n_2
총합	m_1	m_2	N

x_{11} 은 실험군이 성공적인 반응을 보인 대상자 수, x_{21} 은 대조군이 성공적인 반응을 보인 대상자 수이며, x_{12} , x_{22} 는 실험군과 대조군 각각의 실패인 경우를 나타낸 변수이다. m_1 은 약물 투여 시 성공적인 반응을 보인 대상자 수의 합이며, m_2 는 실패인 대상자의 합이다. n_1 과 n_2 는 각각 실험군과 대조군의 실

험대상자 수를 나타낸다. 따라서 각 치료군 별 약물에 대한 성공적인 반응 확률과 통계량에 대한 추정치는 다음과 같다.

$$\hat{p}_1 = \frac{x_{11}}{n_1}, \quad \hat{p}_2 = \frac{x_{21}}{n_2}, \quad RD = \hat{p}_2 - \hat{p}_1, \quad RR = \frac{\hat{p}_2}{\hat{p}_1}$$

\hat{p}_1 은 실험군 중 성공적인 반응을 보인 대상자 수를 실험군 전체 대상자 수인 n_1 으로 나눈 것이며 이는 실험군의 event probability이다. \hat{p}_2 은 대조군의 event probability를 나타낸 것이다. 동등성 평가 기준인 RD는 \hat{p}_2 에서 \hat{p}_1 를 뺀 값으로 계산되며, \hat{p}_2 를 \hat{p}_1 으로 나눈 값이 RR의 추정치가 된다.

3.1 동등성 마진

바이오시밀러의 동등성 평가를 위한 동등성 마진을 구하는 방법은 다양한 방법이 존재하지만, 본 연구에서는 보수적인 방법인 RR과 RD의 95% 신뢰구간의 하한을 이용한 방법을 통해 동등성 마진을 구하였다. 이 때, RD와 RR은 과거 임상시험에서 특정 약물에 대한 위약군과 대조군의 성공적인 반응확률을 통해 구해지는 값이다. 해당 통계량을 구하기 위한 성공적인 반응여부의 분포는 아래 <표 4>와 같다.

표 4. 동등성 마진 설정을 위한 2x2 분할표

대상자 군 \ 약물에 대한 반응	성공	실패	총합
	위약군	x_{11}^0	x_{12}^0
대조군	x_{21}^0	x_{22}^0	n_2^0
총합	m_1^0	m_2^0	N^0

\hat{p}_0^0 는 과거 임상시험에서 특정 약물에 대해 위약군 중 성공적인 반응을 보인 대상자 수를 위약군 전체 대상자 수인 n_1^0 로 나눈 것이며, 이는 위약군의 event probability(\hat{p}_0^0)이다. 마찬가지로 \hat{p}_2^0 는 과거 임상시험에서 특정 약물에 대해 대조군 중 성공적인 반응을 보인 대상자 수를 대조군의 전체 대상자 수인 n_2^0 로 나눈 것이며, 이는 대조군의 event probability(\hat{p}_2^0)를 의미한다.

$$\hat{p}_0^0 = \frac{x_{11}^0}{n_1^0}, \quad \hat{p}_2^0 = \frac{x_{21}^0}{n_2^0}, \quad \widehat{RD} = \hat{p}_2^0 - \hat{p}_0^0, \quad \widehat{RR} = \frac{\hat{p}_2^0}{\hat{p}_0^0}$$

이 때, RD는 \hat{p}_2^0 에서 \hat{p}_0^0 를 뺀 값으로 계산되며, \hat{p}_2^0 를 \hat{p}_0^0 으로 나눈 값이 RR이 된다. 이와 같이 과거 임상시험에서 특정 약물에 대해 성공적인 반응을 보인 위약군과 대조군의 분포를 통해 추정치를 얻었다. 이를 통해 RD와 RR의 동등성 마진을 구하는 데에 필요한 신뢰구간을 구할 수 있다.

3.1.1 RD의 동등성 마진 설정

RD의 동등성 한계를 설정하기 위해 아래와 같은 식을 통해 RD의 신뢰구간을 구한다.

$$\text{RD의 신뢰구간} : (\hat{p}_2^0 - \hat{p}_0^0) \pm z_{\alpha/2} \sqrt{\left(\frac{\hat{p}_0^0 (1 - \hat{p}_0^0)}{n_1^0} + \frac{\hat{p}_2^0 (1 - \hat{p}_2^0)}{n_2^0} \right)}$$

구해진 신뢰구간의 하한을 통계적 마진 M_1 으로 가정하고, 통계적 마진 M_1 에 0.5를 곱하여 임상적 마진인 M_2 를 구한다. 본 논문에서는 RD의 임상적 마진 M_2 를 동등성 마진으로 설정하고, δ_0 로 표현한다. 구해진 δ_0 를 이용하여 동등성 구간 $(-\delta_0, \delta_0)$ 을 설정한다.

3.1.2 RR의 동등성 마진 설정

RR의 동등성 한계를 설정하기 위해 아래와 같은 식을 통해 RR의 신뢰구간을 구한다.

$$\text{RR의 신뢰구간} : \exp\left(\log(RR) \pm z_{\alpha/2} \sqrt{\frac{1}{x_{11}^0} + \frac{1}{x_{21}^0} - \frac{1}{n_1^0} - \frac{1}{n_2^0}}\right)$$

구해진 신뢰구간의 하한을 통계적 마진 M_1 으로 가정하고, 임상적 마진 M_2 을 $\exp\left\{\frac{1}{2} \times \left(\log(RR) - z_{\alpha/2} \sqrt{\frac{1}{x_{11}^0} + \frac{1}{x_{21}^0} - \frac{1}{n_1^0} - \frac{1}{n_2^0}}\right)\right\}$ 식을 통해 구한다.

본 논문에서는 RR의 임상적 마진 M_2 를 동등성 마진으로 설정하고, ϕ_0 로 표현한다. 구해진 ϕ_0 를 이용하여 동등성 구간 $(\frac{1}{\phi_0}, \phi_0)$ 을 설정한다.

3.2 검정통계량을 이용한 동등성 평가방법

3.2.1 동등성 평가기준이 RD인 경우

먼저, Miettinen and Nurminen 방법에 대한 설명이다. Miettinen and Nurminen은 두 확률 간 $\delta_0 (\neq 0)$ 만큼 차이가 존재하는지 여부를 검정하기 위한 검정통계량을 제시하였다. 검정통계량 z_{MND} 를 구하기 위해 필요한 변수들은 아래와 같이 정의된다.

$$\hat{p}_1 = \frac{x_{11}}{n_1}, \quad \hat{p}_2 = \frac{x_{21}}{n_2}$$

$$L_0 = x_{21}\delta_0(1-\delta_0), \quad L_1 = [N_2\delta_0 - N - 2x_{21}]\delta_0 + m_1$$

$$L_2 = (N+n_2)\delta_0 - N - m_1, \quad L_3 = N$$

$$A = \frac{1}{3} \left[\pi + \cos^{-1} \left(\frac{C}{B^3} \right) \right], \quad B = \sqrt{\frac{L_2^2}{(3L_3)^2} - \frac{L_1}{3L_3}}, \quad C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} - \frac{L_0}{2L_3}$$

또한, 제안된 검정통계량 z_{MND} 는 아래와 같은 식으로 나타낼 수 있다.

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}} \quad (3.1)$$

여기서, $\hat{\sigma}_{MND} = \sqrt{\left(\frac{\tilde{p}_1 \tilde{q}_1}{n_1} + \frac{\tilde{p}_2 \tilde{q}_2}{n_2} \right) \left(\frac{N}{N-1} \right)}$, $\tilde{p}_1 = \tilde{p}_2 + \delta_0$, $\tilde{p}_2 = 2B \cos(A) - \frac{L_2}{3L_3}$

$\hat{\sigma}_{MND}$ 에 포함되어있는 수정계수 $\frac{N}{N-1}$ 은 분산의 편향을 보정해주는 역할을

하며, 통계량의 유의수준은 스코어 통계량의 근사적 정규성에 기초한다.

Miettinen and Nurminen이 제안한 검정통계량 z_{MND} 의 절댓값이 $z_{\alpha/2}$ 보다 클 때, 귀무가설(H_0 : 두 치료제 사이에 임상적으로 의미 있는 차이가 있다.)을 기각한다.

다음은 Farrington and Manning(1990) 방법에 대한 설명이다. Farrington and Manning은 두 비율 간 특정 값 δ_0 만큼 차이가 존재하는지 여부를 검정하기 위한 검정통계량 z_{FMD} 를 제안하였다. 검정통계량 z_{FMD} 는 아래와 같은 식으로 나타낸다.

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \quad (3.2)$$

Miettinen and Nurminen이 제안한 검정통계량과 마찬가지로 z_{FMD} 의 절댓값이 $z_{\alpha/2}$ 보다 클 때, 귀무가설(H_0 : 두 치료제 사이에 임상적으로 의미 있는 차이가 있다.)을 기각한다.

3.2.2 동등성 평가기준이 RR인 경우

먼저, Miettinen and Nurminen이 제안한 검정통계량에 대한 설명이다. Miettinen and Nurminen은 RR이 지정된 값 ϕ_0 과 같은지를 검정하기 위한 검정통계량을 제안하였다. 분산의 편향을 보정하기 위한 보정 계수로

$\frac{N}{N-1}$ 가 사용되었으며 이 검정통계량의 유의 수준은 스코어 통계량의 점근적

정규성에 기초된다.

Miettinen and Nurminen이 제안한 검정통계량을 계산하기 위해 필요한 변수들은 아래와 같이 정의되며

$$\tilde{p}_1 = \tilde{p}_2 \phi_0, \quad \tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0, \quad B = -[n_1\phi_0 + x_{11} + n_2 + x_{21}\phi_0], \quad C = m_1$$

검정통계량 z_{MNR} 은 아래와 같은 식으로 나타낼 수 있다.

$$z_{MNR} = \frac{\hat{p}_1/\hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2 \frac{\tilde{p}_2\tilde{q}_2}{n_2}\right) \left(\frac{N}{N-1}\right)}} \quad (3.3)$$

다음은 Farrington and Manning이 제안한 방법에 대한 설명이다.

z_{FMR} 은 Miettinen and Nurminen(1985)의 z_{MNR} 에서 보정계수 $\sqrt{\frac{N-1}{N}}$ 가 제외된 식이며, \tilde{p}_1, \tilde{p}_2 은 Miettinen and Nurminen이 제안한 통계량과 동일한 방식으로 계산된다.

Farrington and Manning이 제안한 검정통계량은 아래와 같다.

$$z_{FMR} = \frac{\hat{p}_1/\hat{p}_2 - \phi_0}{\sqrt{\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2 \frac{\tilde{p}_2\tilde{q}_2}{n_2}}} \quad (3.4)$$

RD의 경우와 마찬가지로 RR의 검정통계량 z_{MNR} , z_{FMR} 의 절댓값이 $z_{\alpha/2}$ 보다 클 때, 귀무가설(H_0 : 두 치료제 사이에 임상적으로 의미 있는 차이가 없다.)을 기각한다.

3.3 신뢰구간을 이용한 동등성 평가방법

신뢰구간을 이용한 동등성 평가는 본 논문의 모의실험에서 사용된 방법으로, 95% 신뢰구간을 사용하였다.

RD의 경우, 3.1.1에서 동등성 마진을 구하기 위해 사용되었던 RD의 신뢰구간 계산 공식을 이용하여 RD의 신뢰구간을 구한다. 이 때, 산출된 신뢰구간이 RD의 동등성 구간 $(-\delta_0, \delta_0)$ 내에 포함되면 두 치료제 간 임상적으로 유의한 차이가 없다(귀무가설 기각)고 판단한다.

RR의 경우, 3.1.2에서 동등성 마진을 구하기 위해 사용되었던 RR의 신뢰구간 계산 공식을 이용하여 RR의 신뢰구간을 구한다. 이 때, 산출된 신뢰구간이 RR의 동등성 구간 $(\frac{\phi_0}{\phi_0}, \phi_0)$ 내에 포함되면 두 치료제 간 임상적으로 유의한 차이가 없다고 검정할 수 있다.

3.4 동등성 마진 설정방법 제안

동등성 평가기준 RD와 RR에 동일한 마진 설정 기준을 적용하였을 때, 산출되는 동등성 구간은 RD에 비해 RR이 상대적으로 넓다. 동등성 구간 폭의 차이는 동등성 평가 결과에도 큰 영향을 미칠 수 있으므로 본 연구에서는 RD의 동등성 마진에 맞춰 RR의 동등성 마진을 조정하는 방법을 제안하였다.

본 연구에서 제안하는 동등성 마진 설정방법에 대한 설명이다. 대조군의 event probability(p_2)에서 앞서 정의된 RD의 동등성 마진(δ_0)을 뺀 값을 p_2 으로 나눈 값을 RR의 동등성 구간의 하한으로, p_2 에 δ_0 를 더한 값을 RR의 동등성 구간의 상한으로 하여 RR의 동등성 구간을 조정하는 방식이다.

$$\text{조정된 RR의 동등성 구간} : \left(\frac{p_2 - \delta_0}{p_2}, \frac{p_2 + \delta_0}{p_2} \right)$$

제안된 방식으로 RR의 동등성 구간을 조정하면, RD와 RR간 동등성 평가 결과에 발생하는 차이가 줄어들 것으로 생각하여 위와 같은 식에 따라 조정된 RR의 동등성 구간을 적용하여 모의실험을 진행하였다. 해당 모의실험에 대한 내용은 4장에 자세히 기술하였다.

제 4장 모의실험

제 4장에서는 2장에서 언급되었던 바이오시밀러의 동등성 평가 시 사용되는 평가변수인 RR과 RD의 신뢰구간이 동등성 구간에 포함될 확률을 비교하고자 모의실험을 수행, 결과를 비교한다. 다음으로 검정통계량을 이용한 동등성 평가 방법에서 RD와 RR의 검정력 비교를 위한 모의실험을 수행, RD와 RR의 결과를 비교한다. 마지막으로 본 논문에서 제안한 방법인 RD와 RR의 동등성 마진을 조정하여 첫 번째와 동일한 방식의 모의실험을 수행하고, 앞선 모의실험의 결과와 비교한다.

4.1 모의실험 구축

4.1.1 RD와 RR의 비교를 위한 모의실험

RD와 RR의 비교를 위한 모의실험은 다음과 같은 단계로 구축하였다.

① 동등성 마진을 설정하기 위해 과거 임상시험의 위약대비 대조군 시험 결과를 다음과 같이 가정하였다. 시험 대상자 수는 각 군당 100명씩 배정하였고, RD와 RR을 산출하기 위해 대조약의 probability를 0.6, 위약의 event probability는 0.2, 0.3으로 설정하였다.

② 앞서 가정된 상황에 맞는 RD와 RR의 95% 신뢰구간을 구하고, 구해진 신뢰구간을 토대로 RD와 RR의 동등성 마진을 설정하였다.

RD의 경우, 위약의 event probability(\hat{p}_0)가 0.2일 때, 신뢰구간은

(0.28, 0.52)이 되며, 이에 대한 통계적 마진(M_1)은 0.28, 임상적 마진(M_2)은 0.14, 동등성 구간은 $(-0.14, 0.14)$ 이다. 동일한 방식으로, 위약의 event probability(\hat{p}_0^0)가 0.3인 경우 신뢰구간은 $(0.17, 0.43)$ 이 되며, M_1 은 0.17, M_2 는 0.09, 동등성 구간은 $(-0.09, 0.09)$ 가 된다.

RR의 경우, 위약의 event probability(\hat{p}_0^0)가 0.2인 경우 신뢰구간은 $(1.96, 4.58)$ 이 되며, M_1 은 1.96, M_2 는 1.40, 동등성 구간은 $(0.71, 1.40)$ 이 된다. 또한 위약의 event probability(\hat{p}_0^0)가 0.3인 경우 신뢰구간은 $(1.42, 2.81)$ 이 되며, M_1 은 1.42, M_2 은 1.19, 동등성 구간은 $(0.83, 1.19)$ 이 된다.

③ 본 연구의 바이오시밀러 동등성 시험에 필요한 대상자 수를 계산한다. 이 때 RD를 이용한 동등성 시험에 필요한 대상자 수를 먼저 계산하여, RR에도 동일하게 적용하였고, 동등성 시험의 power는 80%인 경우, 90%인 경우로 총 두 가지 상황을 고려하였다.

시험 대상자수는 실험설계에 따라 PASS2008 프로그램을 통해 산출 하였고, 산출된 대상자 수는 다음과 같다.

위약의 event probability(\hat{p}_0^0)가 0.2인 경우, power 80%이면 각 군당 260명의 시험대상자가 필요하고, power 90%이면 각 군당 620명의 시험대상자가 필요하다. 또한 위약의 event probability(\hat{p}_0^0)가 0.3인 경우, power 80%이면 각 군당 320명의 시험대상자가 필요하고, 90%이면 각 군당 770명의 시험대상자가 필요하다.

이상의 설계에 따라 결정된 동등성 마진과 대상자 수를 정리하면 다음과 같다.

표 5. 모의실험을 위해 생성된 자료 ($\alpha=0.025, n_1, n_2=100$)

평가변수	\hat{p}_0^0	\hat{p}_2^0	95% 신뢰구간	M_1	M_2	필요한 대상자 수
RD	0.2	0.6	(0.28, 0.52)	0.28	0.14	260명
	0.3	0.6	(0.17, 0.43)	1.96	1.40	620명
RR	0.2	0.6	(1.96, 4.58)	0.17	0.09	320명
	0.3	0.6	(1.42, 2.81)	0.83	1.19	770명

4.1.2 검정통계량을 이용한 모의실험

검정통계량을 이용한 모의실험에서는 두 가지 동등성 평가방법 (Miettinen and Nurminen(1985), Farrington and Manning(1990))에서 RD와 RR의 검정력을 비교하는 모의실험을 수행하였다.

모의실험은 총 10,000번 반복실험 하였으며 3.2절에서 정의된 검정통계량에 따라 검정통계량을 산출하여, 해당 검정통계량의 절댓값이 $z_{\alpha/2}$ 인 1.96보다 클 때, 두 치료제가 임상적으로 의미 있는 차이가 있다는 귀무가설을 기각할 확률을 계산하였다.

4.1.3 제안된 동등성 마진을 이용한 모의실험

본 논문의 3.4절에서 제안했던 방법인 RR의 동등성 구간을 RD의 동등성 구간에 맞춰 조정하는 방법에 대한 모의실험이다. 4.1.1에서 구해진 RD의 동등성 구간에 맞춰 RR의 동등성 구간을 조정하였고, 조정된 RR의 동등성 구간은 δ_0 가 0.14일 때, (0.77, 1.23)이고, δ_0 가 0.09일 때, (0.85, 1.15)이다.

4.1.1에서 진행된 모의실험과 동일한 방식으로 진행하되 RR의 동등성 구간만 변경하여 RD와 RR의 신뢰구간이 동등성 구간에 포함될 확률을 알아보는 모의실험을 진행하였다.

4.2 모의실험 설계

본 연구의 모의실험에서는 생성된 자료를 이용하여 동등성 시험의 의사결정 과정에서 평가변수로 RD를 사용하였을 때와 RR 사용하였을 때 바이오시밀러와 기존 약의 동등성 평가 시 검정력에 차이가 존재하는지 알아보고자 한다. 다음으로 Miettinen and Nurminen, Farrington and Manning에 의해 제안된 검정통계량을 이용하여 기존약과 바이오시밀러의 동등성 평가를 수행할 때, 평가변수(RD와 RR)에 따른 검정력 차이가 존재하는지 비교한다. 마지막으로 본 논문에서 제안한 방법에 따라 조정된 RR의 동등성 마진을 이용하여 바이오시밀러와 기존 약의 동등성 평가 시 RD와 RR의 검정력을 비교한다.

첫째, 본 논문에서는 대조약이 위약보다 우월함을 검정(Superiority test)하기 위해 필요한 sample size를 100명으로 가정한다. 기존 약의 event proportion을 0.6으로 고정하고, 바이오시밀러의 event proportion(\hat{p}_1)을 0.5에서 0.6까지 0.01씩 증가시키며 모의실험을 진행한다.

RD와 RR, 두 가지 기준을 이용해 구해진 신뢰구간이 10,000번의 반복실험 중 동등성 구간에 총 몇 번 포함되는지 동등성 시험 결과를 비교한다. 또한 각 방법에서 power, 위약과 대조약의 event probability의 차이에 따라 검정 결과가 달라지는지 확인한다.

둘째, 바이오시밀러의 동등성 평가를 위해 제안된 두 가지 검정통계량을 이용하여 동등성 평가 변수(RD, RR)에 따라 바이오시밀러 동등성 평가 검정력

에 차이가 발생하는지 비교한다. 방법마다 총 10,000번씩 반복 실행하여 귀무가설을 기각하는 횟수를 비교한다.

셋째, 본 논문에서 제안한 동등성 구간의 조정이 RD와 RR의 신뢰구간이 동등성 구간에 포함될 확률에 미치는 영향을 알아보는 모의실험을 진행하고, 동등성 구간 조정 전후의 모의실험 결과를 비교한다.

모든 모의실험에서 $\alpha = 0.025$ 로 설정한다. 모의실험에서 사용한 모든 분석은 R 프로그램 3.3.3 버전을 사용하여 분석하였으며, 임상시험에 필요한 대상자 수의 산출은 PASS(Power Analysis and Sample Size) 2008 버전을 사용하였다.

4.3 모의실험 결과

모의실험은 위약의 event probability(\hat{p}_0)가 0.2, 0.3인 경우를 power가 0.8, 0.9일 때로 구분하여 수행하였으며 4.3.1에서 결과를 비교하여 제시하였다. 4.3.2에서는 검정통계량 z 를 이용한 평가방법(Miettinen and Nurminen(1985), Farrington and Manning(1990))간 모의실험을 진행하여 평가변수(RD, RR)에 따라 동등성 평가의 검정력에 차이가 존재하는지에 대한 모의실험 수행결과를 제시하였다. 마지막으로 4.3.3에서는 본 논문에서 제안한 마진 설정방법과 기존의 마진 설정방법 간 검정력을 비교한다.

4.3.1 신뢰구간을 이용한 모의실험 결과

표 6-1은 위약군의 event probability가 0.2인 경우 RR과 RD의 신뢰구간이 동등성 구간에 포함될 확률을 나타낸 표이며 이를 그래프로 나타낸 것이 그림 2-1이다. RD를 기준으로 산출된 시험 대상자 수를 RR에도 동일하게 적용하여 모의실험을 수행한 결과 RD에 비해 RR의 신뢰구간이 동등성 구간에 포함되는 비율이 더 높았다. 이는 위약군의 event probability가 0.3인 경우(표 6-2, 그림 2-2)에도 비슷한 결과를 보였다.

표 6-1. 총 10,000번의 반복 실행 중 동등성 평가변수(RD, RR)의
95% 신뢰구간이 동등성 구간에 포함된 분율
($\alpha = 0.025$, 동등성 구간 : $RD = \pm 0.14$, $RR = (0.71, 1.40)$)

		power		RD	RR
과거 임상시험	현재 모의 실험				
\hat{p}_0^0	\hat{p}_2^0	p_1	p_2	N=520	
		0.5		0.1529	0.5144
		0.51		0.2210	0.6251
		0.52		0.2935	0.7184
		0.53		0.3749	0.8015
		0.54		0.4586	0.8702
0.2	0.6	0.55	0.6	0.8	0.9206
		0.56			0.9525
		0.57			0.9722
		0.58			0.9840
		0.59			0.9918
		0.60			0.9935
\hat{p}_0^0	\hat{p}_2^0	p_1	p_2	N=640	
		0.5		0.1755	0.5993
		0.51		0.2532	0.7133
		0.52		0.3410	0.8028
		0.53		0.4369	0.8749
		0.54		0.5392	0.9296
0.2	0.6	0.55	0.6	0.9	0.9627
		0.56			0.9807
		0.57			0.9913
		0.58			0.9960
		0.59			0.9979
		0.60			0.9988

표 6-2. 총 10,000번의 반복 실행 중 동등성 평가변수(RD, RR)의
95% 신뢰구간이 동등성 구간에 포함된 분율
($\alpha = 0.025$, 동등성 구간 : $RD = \pm 0.09$, $RR = (0.83, 1.19)$)

				power	RD	RR
과거임상시험	현재모의실험					
\hat{p}_0^0	\hat{p}_2^0	p_1	p_2	N=1,240		
0.3	0.6		0.5	0.8	0.0112	0.0306
			0.51		0.0238	0.0668
			0.52		0.052	0.1366
			0.53		0.1036	0.2393
			0.54		0.1828	0.3726
			0.55		0.2905	0.5276
			0.56		0.4188	0.6767
			0.57		0.5607	0.8025
			0.58		0.6808	0.8845
			0.59		0.7700	0.9317
	0.60	0.8009	0.9428			
\hat{p}_0^0	\hat{p}_2^0	p_1	p_2	N=1,540		
0.3	0.6		0.5	0.9	0.0098	0.0314
			0.51		0.0267	0.0777
			0.52		0.0602	0.1616
			0.53		0.1258	0.2853
			0.54		0.2243	0.4481
			0.55		0.3602	0.6182
			0.56		0.5100	0.7673
			0.57		0.6658	0.8795
			0.58		0.7911	0.9430
			0.59		0.8717	0.9750
	0.60	0.8975	0.9796			

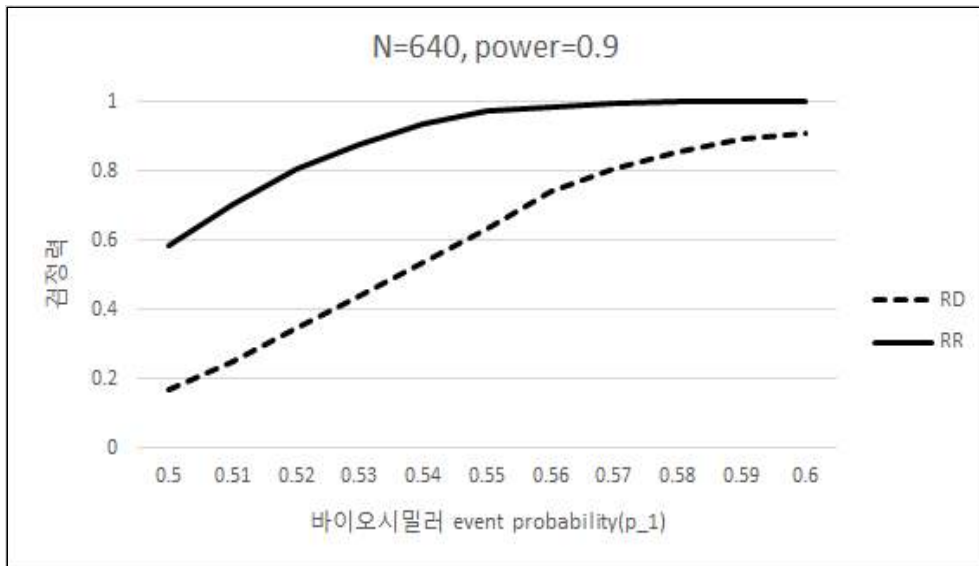
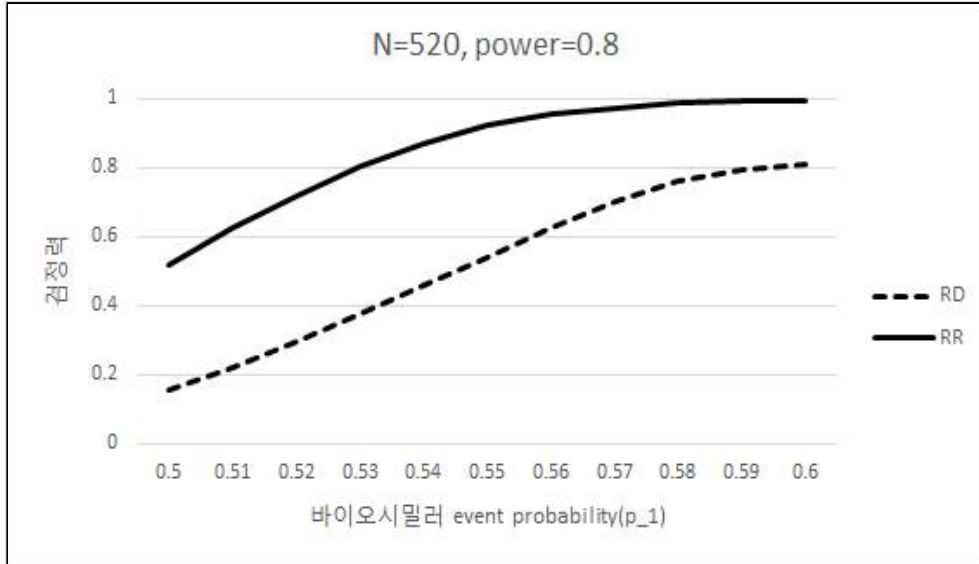


그림 2-1. 바이오시밀러 동등성 평가에 대한 RD와 RR의 검정력 비교
 (repetition = 10,000, $\alpha = 0.025$, placebo = 0.2, control = 0.6)

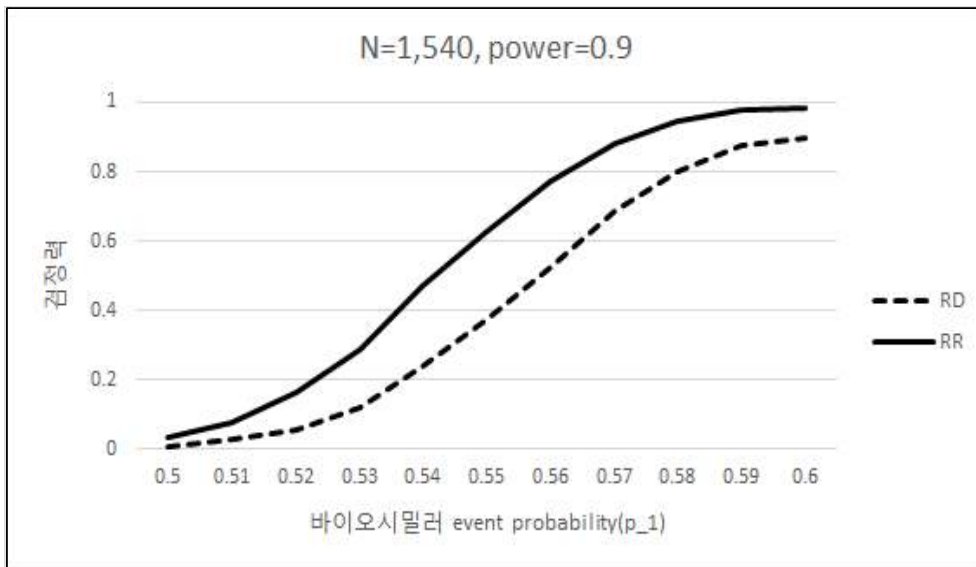
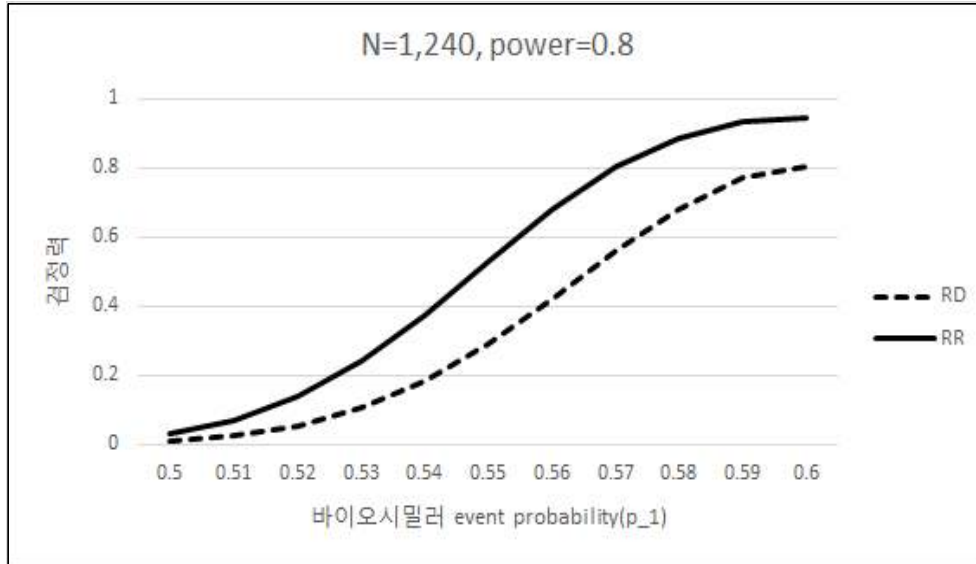


그림 2-2. 바이오시밀러 동등성 평가에 대한 RD와 RR의 검정력 비교
 (repetition = 10,000, $\alpha = 0.025$, placebo = 0.3, control = 0.6)

4.3.2 검정통계량을 이용한 모의실험 결과

4.3.2에서는 두 가지 검정통계량에서 RD와 RR의 동등성 평가 모의실험 결과를 비교하였다. 두 가지 검정통계량은 Miettinen and Nurminen과 Farrington and Manning이 제안한 검정통계량들이다. 시험 대상자 수는 RD와 RR의 신뢰구간을 이용한 동등성 평가 방법과 동일하게 RD를 기준으로 산출된 대상자 수를 RR에도 적용하였다.

위약군의 event probability가 0.2일 때(RD=0.4, RR=3) 평가방법 간 결과 비교는 표 7-1, 7-2, 그림 3-1, 3-2이며 위약군의 event probability가 0.3일 때(RD=0.3, RR=2) 평가방법 간 결과 비교는 표 8-1, 8-2, 그림 4-1, 4-2이다.

모의실험 결과, 위약군의 event probability가 0.2인 경우, RR의 검정력은 RD에 비해 눈에 띄게 높았고, 이는 위약군의 event probability가 0.3인 경우에도 비슷한 경향을 보였다. 이와 같은 결과는 RD와 RR의 신뢰구간을 이용한 동등성 평가의 모의실험 결과와 비슷한 결과이다.

표 7-1. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교

($\alpha = 0.025$, $repetition = 10,000$, 시험 대상자 수 : 520명)

과거임상시험		power			Miettinen and Nurminen*		Farrington and Manning**	
\hat{p}_0	\hat{p}_2	p_1	p_2		RD	RR	RD	RR
		0.5			0.1529	0.5054	0.1529	0.5054
		0.51			0.2210	0.6150	0.2210	0.6150
		0.52			0.2935	0.7099	0.2935	0.7099
		0.53			0.3749	0.7929	0.3749	0.7929
		0.54			0.4587	0.8651	0.4587	0.8651
0.2	0.6	0.55	0.6	0.8	0.5405	0.9176	0.5405	0.9176
		0.56			0.6224	0.9491	0.6225	0.9491
		0.57			0.7013	0.9707	0.7014	0.9707
		0.58			0.7608	0.9832	0.7611	0.9834
		0.59			0.7942	0.9913	0.7944	0.9913
		0.60			0.8082	0.9931	0.8084	0.9933

* Miettinen and Nurminen(1985) 검정통계량의 검정력

** Farrington and Manning(1990) 검정통계량의 검정력

표 7-2. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교

($\alpha = 0.025$, $repetition = 10,000$, 시험 대상자 수 : 640명)

과거임상시험		현재모의실험		power	Miettinen and Nurminen*		Farrington and Manning**	
\hat{p}_0	\hat{p}_2	p_1	p_2		RD	RR	RD	RR
		0.5			0.1755	0.5896	0.1755	0.5898
		0.51			0.2532	0.7037	0.2532	0.7038
		0.52			0.3410	0.7953	0.3410	0.7956
		0.53			0.4369	0.8691	0.4369	0.8691
		0.54			0.5392	0.9254	0.5392	0.9254
0.2	0.6	0.55	0.6	0.9	0.6361	0.9606	0.6361	0.9606
		0.56			0.7272	0.9795	0.7272	0.9795
		0.57			0.7997	0.9905	0.7999	0.9905
		0.58			0.8566	0.9957	0.8570	0.9957
		0.59			0.8907	0.9977	0.8908	0.9977
		0.60			0.9036	0.9986	0.9040	0.9986

* Miettinen and Nurminen(1985) 검정통계량의 검정력

** Farrington and Manning(1990) 검정통계량의 검정력

표 8-1. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교

($\alpha = 0.025$, $repetition = 10,000$, 시험 대상자 수 : 1,240명)

과거임상시험		power			Miettinen and Nurminen*		Farrington and Manning**	
\hat{p}_0	\hat{p}_2	현재모의실험		RD	RR	RD	RR	
		p_1	p_2					
		0.5			0.0113	0.0297	0.0113	0.0297
		0.51			0.0238	0.0663	0.0238	0.0663
		0.52			0.0521	0.1341	0.0521	0.1341
		0.53			0.1036	0.2354	0.1036	0.2354
		0.54			0.1834	0.3692	0.1834	0.3692
0.3	0.6	0.55	0.6	0.8	0.2916	0.5231	0.2923	0.5231
		0.56			0.4198	0.6738	0.4206	0.6738
		0.57			0.5633	0.7990	0.5644	0.7990
		0.58			0.6843	0.8825	0.6851	0.8825
		0.59			0.7727	0.9298	0.7731	0.9298
		0.60			0.8029	0.9417	0.8033	0.9417

* Miettinen and Nurminen(1985) 검정통계량의 검정력

** Farrington and Manning(1990) 검정통계량의 검정력

표 8-2. 검정통계량을 이용한 동등성 평가에서 RD와 RR의 검정력 비교

($\alpha = 0.025$, $repetition = 10,000$, 시험 대상자 수 : 1,540명)

과거임상시험		power			Miettinen and Nurminen*		Farrington and Manning**	
\hat{p}_0	\hat{p}_2	현재모의실험		RD	RR	RD	RR	
		p_1	p_2					
		0.5			0.0100	0.0310	0.0100	0.0310
		0.51			0.0269	0.0770	0.0269	0.0770
		0.52			0.0602	0.1599	0.0602	0.1599
		0.53			0.1263	0.2828	0.1263	0.2830
		0.54			0.2244	0.4448	0.2244	0.4450
0.3	0.6	0.55	0.6	0.9	0.3603	0.6155	0.3603	0.6158
		0.56			0.5100	0.7635	0.5101	0.7640
		0.57			0.6659	0.8775	0.6659	0.8782
		0.58			0.7913	0.9420	0.7915	0.9421
		0.59			0.8721	0.9741	0.8721	0.9741
		0.60			0.8985	0.9790	0.8988	0.9791

* Miettinen and Nurminen(1985) 검정통계량의 검정력

** Farrington and Manning(1990) 검정통계량의 검정력

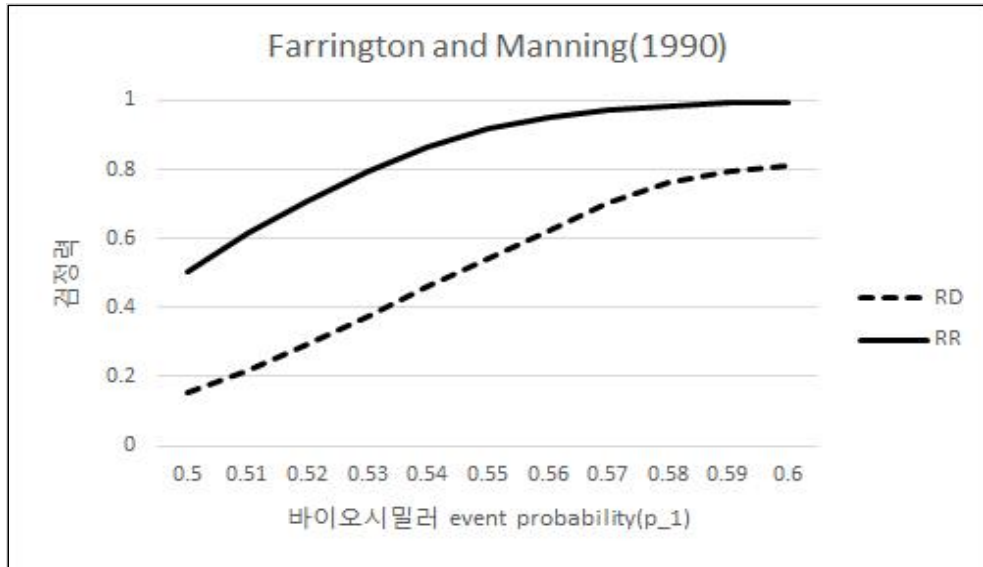
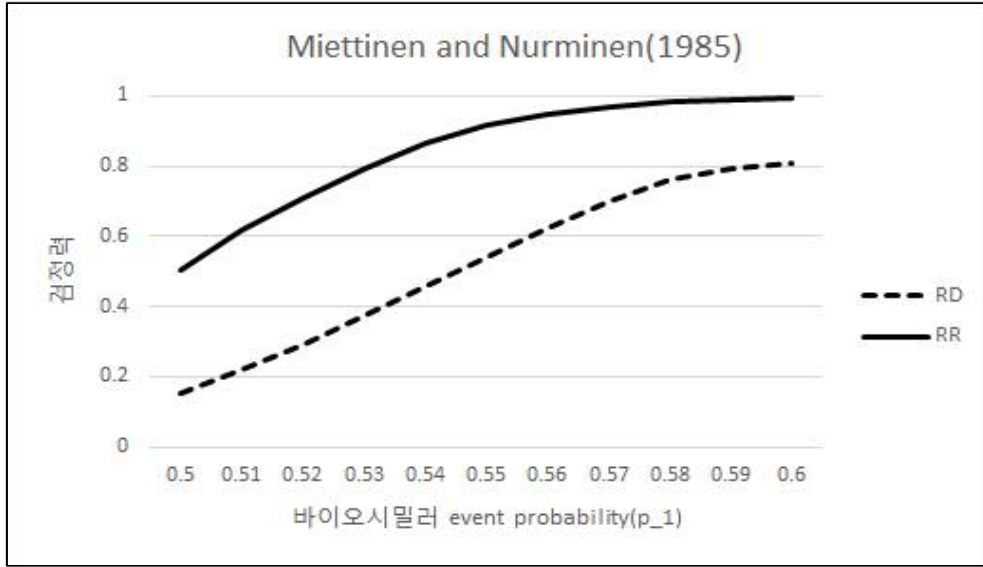


그림 3-1. 검정통계량을 이용한 바이오시밀러의 동등성 평가에서
 RD와 RR의 검정력 비교
 (power = 0.8, repetition = 10,000, α = 0.025, placebo = 0.2, control = 0.6)

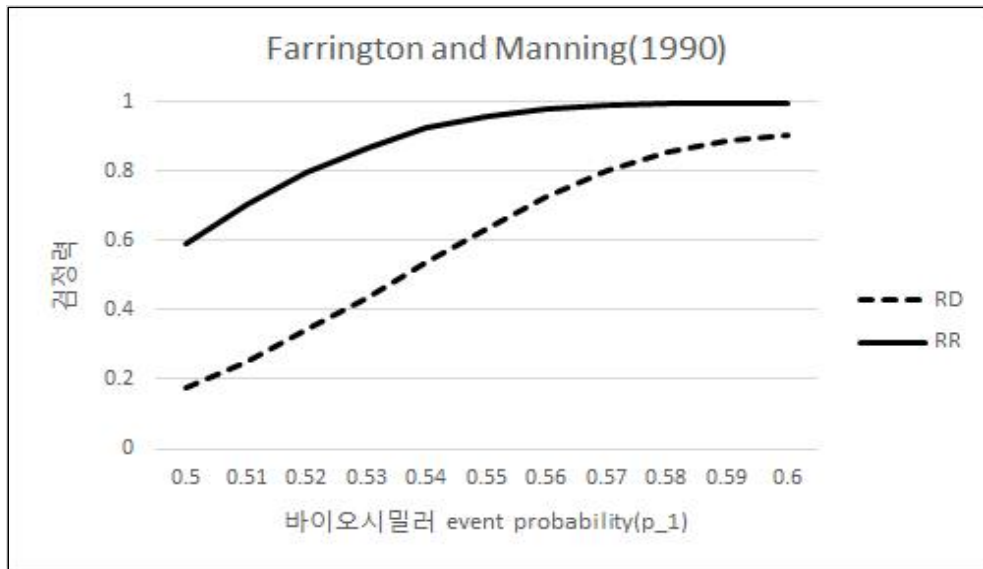
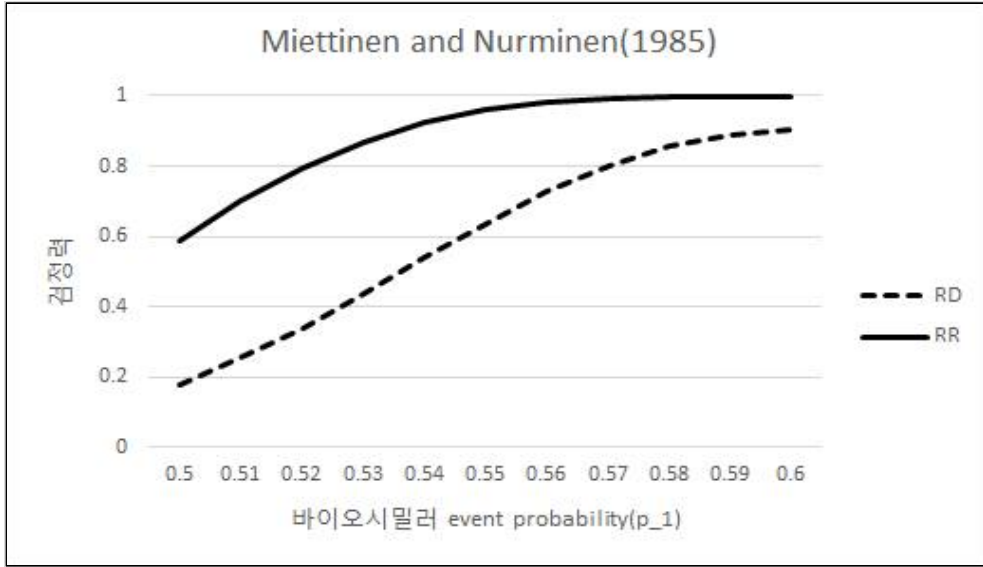


그림 3-2. 검정통계량을 이용한 바이오시밀러의 동등성 평가에서
 RD와 RR의 검정력 비교
 (power = 0.9, repetition = 10,000, α = 0.025, placebo = 0.2, control = 0.6)

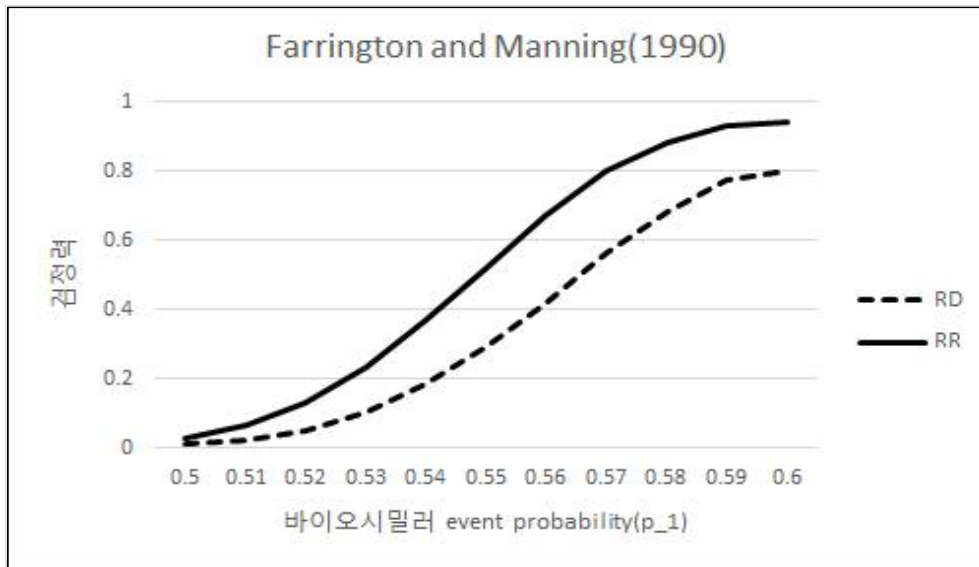
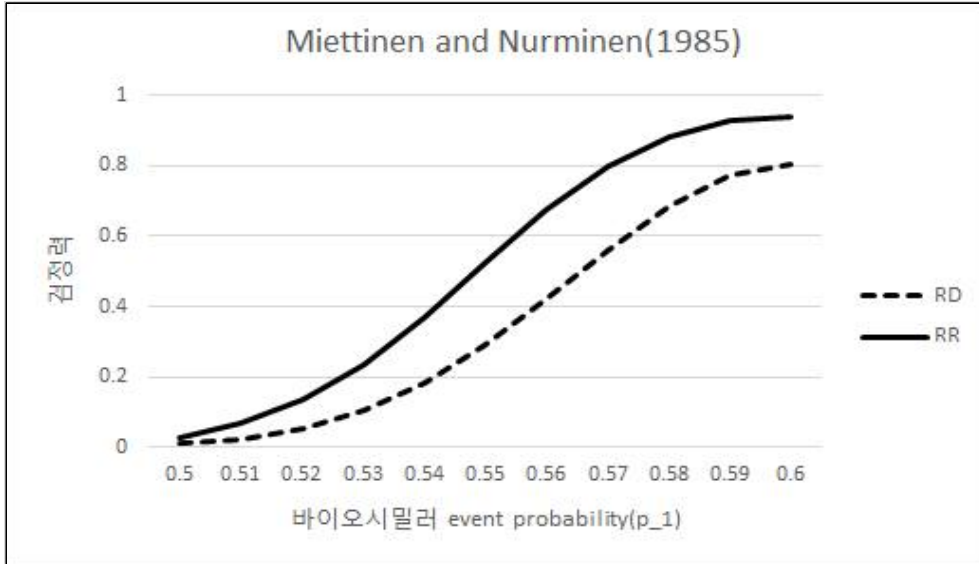


그림 4-1. 검정통계량을 이용한 바이오시밀러의 동등성 평가에서
 RD와 RR의 검정력 비교
 (power = 0.8, repetition = 10,000, α = 0.025, placebo = 0.3, control = 0.6)

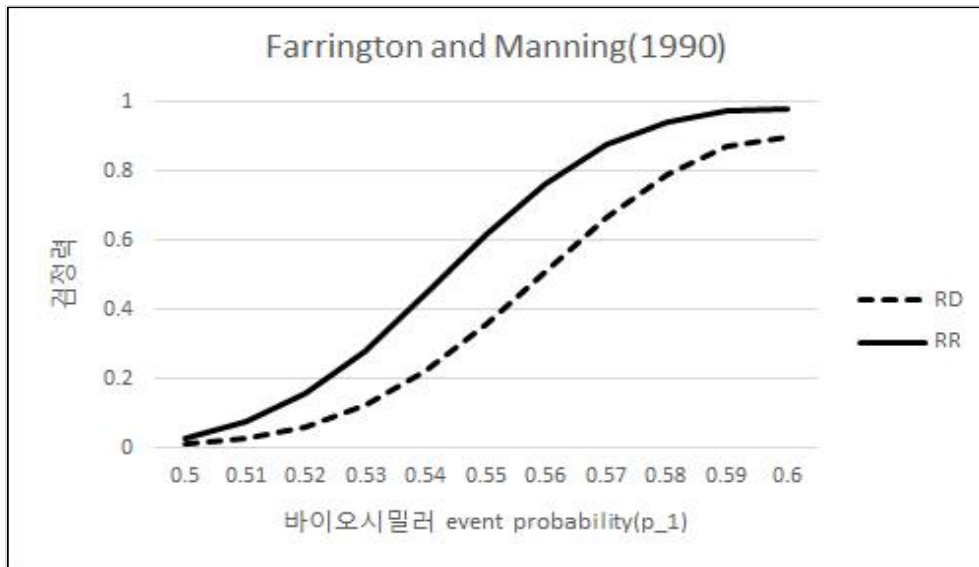
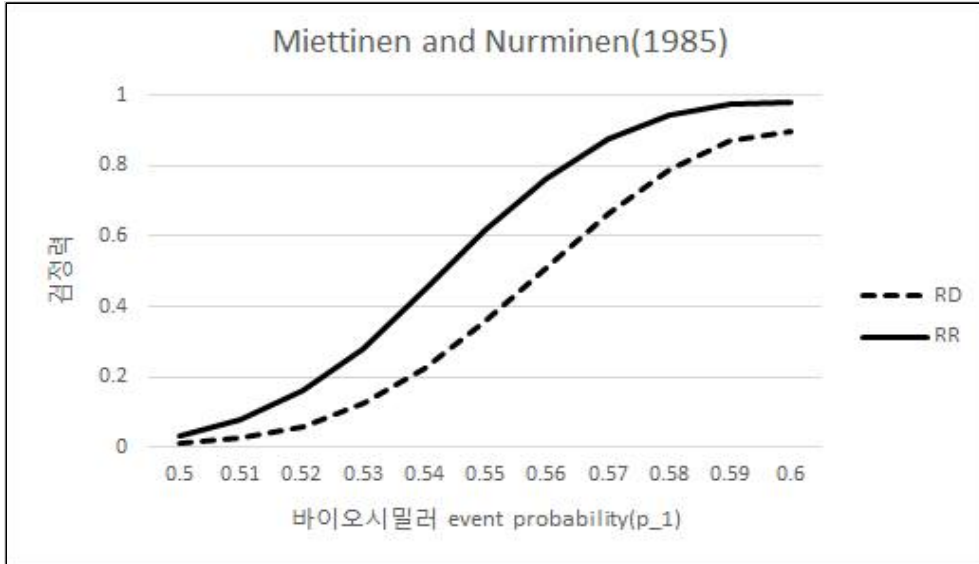


그림 4-2. 검정통계량을 이용한 바이오시밀러의 동등성 평가에서
 RD와 RR의 검정력 비교
 (power = 0.9, repetition = 10,000, α = 0.025, placebo = 0.3, control = 0.6)

4.3.3 제안된 방법의 모의실험 결과

4.3.1과 4.3.2에서 제시된 모의실험 결과에 따르면 바이오시밀러의 동등성 평가 기준인 RD와 RR의 검정력에는 분명한 차이가 존재한다. 검정력에 차이가 존재하는 이유는 RD의 신뢰구간에 비해 RR의 신뢰구간이 넓게 설정되었기 때문이라고 생각하였고, 두 평가변수의 검정력 차이를 줄이기 위해 RD의 동등성구간에 맞추어 RR의 동등성 구간을 조정하여 바이오시밀러 동등성평가 모의실험을 수행하였다.

모의실험 결과, 동등성 마진을 조정한 두 평가변수 간 검정력에는 앞선 모의실험 결과와 달리 큰 차이가 존재하지 않는 것을 확인할 수 있었다. (표 9-1, 9-2, 그림 5) 따라서 두 가지 동등성 평가변수(RR, RD)를 비슷한 가치를 지니는 평가 기준이라 칭하기 위해서는 평가변수(RD, RR)의 동등성 마진 조정이 필요할 것으로 생각된다.

표 9-1. \hat{p}_0 가 0.2일 때, 동등성 평가변수(RD,RR)의 신뢰구간이 동등성 구간에 포함되는 확률 ($\alpha = 0.025, rep. = 10,000$)

과거임상시험		현재모의실험		power		RD	RR
\hat{p}_0	\hat{p}_2	p_1	p_2	동등성구간		N=520	
				RD	RR		
		0.5				0.1529	0.1655
		0.51				0.2210	0.2460
		0.52				0.2935	0.3271
		0.53				0.3749	0.4244
		0.54				0.4586	0.5149
0.2	0.6	0.55	0.6	0.8	± 0.14	(0.77, 1.23)	0.6038
		0.56				0.6222	0.6861
		0.57				0.7012	0.7438
		0.58				0.7607	0.7798
		0.59				0.7939	0.7850
		0.60				0.8071	0.7735

표 9-2. \hat{p}_0 가 0.3일 때, 동등성 평가변수(RD,RR)의 신뢰구간이 동등성 구간에 포함되는 확률 ($\alpha = 0.025, rep. = 10,000$)

과거임상시험		현재모의실험		power		동등성구간		RD	RR
\hat{p}_0	\hat{p}_2	p_1	p_2			RD	RR	N=1,240	
		0.5						0.0112	0.0110
		0.51						0.0238	0.0254
		0.52						0.0520	0.0572
		0.53						0.1036	0.1164
		0.54						0.1828	0.2100
0.3	0.6	0.55	0.6	0.8		± 0.09	(0.85, 1.15)	0.2905	0.3369
		0.56						0.4188	0.4790
		0.57						0.5607	0.6216
		0.58						0.6808	0.7336
		0.59						0.7700	0.7889
		0.60						0.8009	0.7921

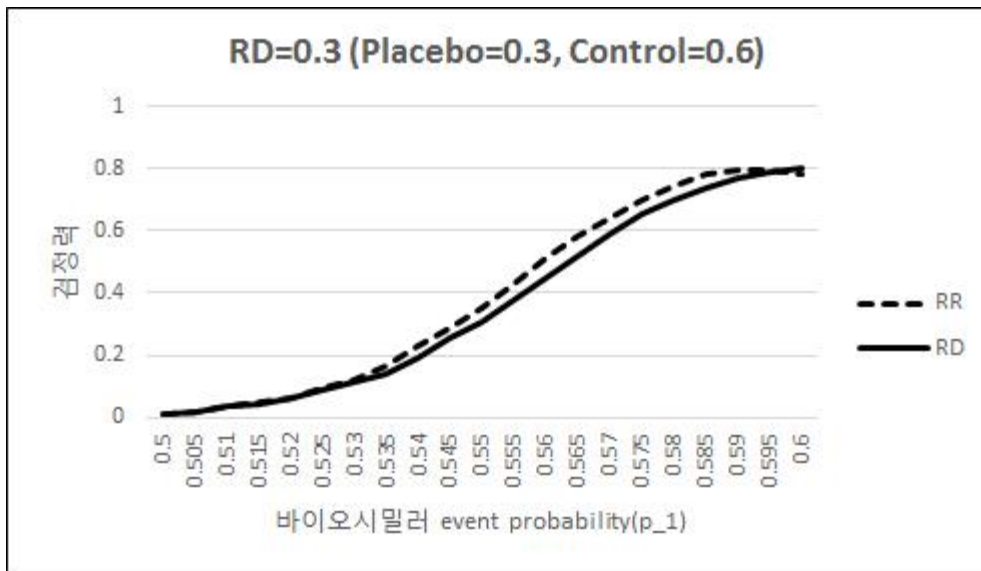
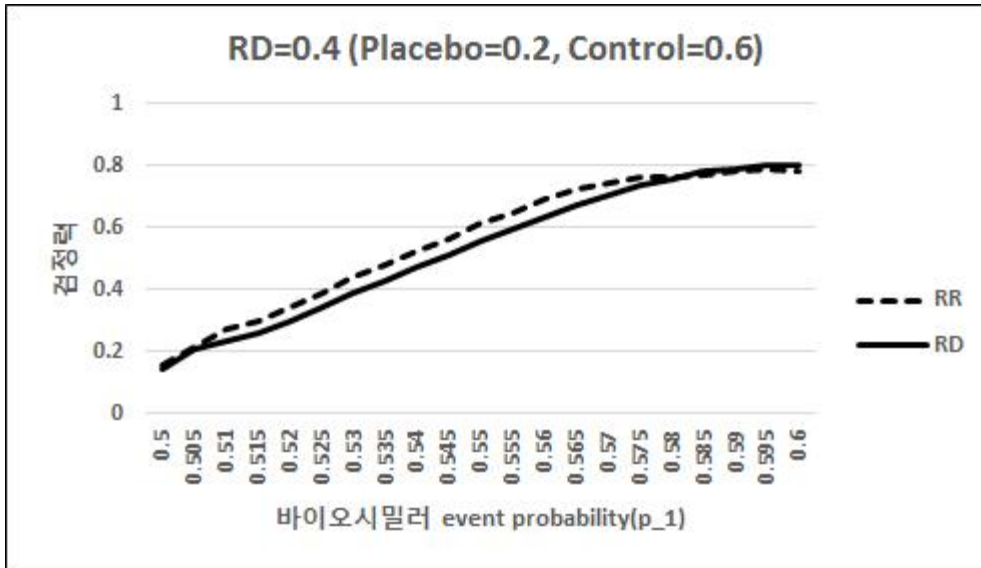


그림 5. RD와 RR의 95% 신뢰구간이 동등성 구간에 포함될 확률 비교

제 5장 결론 및 고찰

바이오시밀러(biosimilar)는 오리지널 의약품 대비 효능효과 및 안정성에서 동등함을 입증한 복제의약품으로, 고가의 바이오의약품을 대체하여 보다 합리적인 비용으로 환자에게 치료기회를 제공하며, 의료비용 절감효과를 통해 보다 나은 삶의 질을 보장해 줄 수 있다. 이러한 순기능을 가진 바이오시밀러는 2015년 이후 블록버스터급 항체의약품들이 순차적으로 특허가 만료되면서 다수의 제약/바이오 업체들에 의해 개발되고 있으며 따라서 글로벌 바이오시밀러 시장의 규모 또한 크게 성장할 것으로 예견되고 있다.

이러한 시기에 바이오시밀러가 기존 바이오의약품과 동등한지 여부를 결정하는 동등성 평가에 사용되는 평가기준은 더욱 중요할 수밖에 없다. 따라서 본 논문에서는 바이오시밀러의 동등성 평가기준으로써 FDA의 RR과 EMA의 RD를 모의실험을 통해 비교하고자 하였다.

모의실험 결과, RD와 RR의 신뢰구간이 동등성 구간 내에 포함될 확률(검정력)은 RR이 RD에 비해 매우 높았으며, 이러한 결과는 power나, 위약군의 event probability에 관계없이 RR의 검정력이 RD의 검정력보다 더 높은 경향을 보였다. 이러한 결과의 원인은 RR을 평가기준으로 하는 동등성 시험에서는 RD에 비해 상대적으로 적은 수의 시험 대상자 수가 필요한데 본 연구의 모의실험에서는 RD를 기준으로 산출된 시험 대상자 수를 RR에도 동일하게 적용하였기 때문에 RR이 의도치 않게 over-sampling되어 RR의 검정력이 RD에 비해 상대적으로 높게 나온 것으로 생각된다.

검정통계량을 이용한 RD와 RR의 검정력 비교를 위한 모의실험에서는

Miettinen and Nurminen(1985)이 제안한 검정통계량과 Farrington and Manning(1990)이 제안한 검정통계량을 이용하여 RD와 RR의 검정력을 비교하였다. 모의실험 결과, 위약군의 event probability가 0.2인 경우, RR의 검정력이 RD에 비해 눈에 띄게 높았으며, 이러한 결과는 두 가지 검정통계량 모

두에서 비슷한 경향을 보였다. 또한 위약군의 event probability가 0.3일 때에도 RR의 검정력이 RD에 비해 월등히 높았다. 이는 앞서 수행되었던 신뢰구간을 이용한 방법의 모의실험에서와 비슷한 결과였다.

동일한 동등성 마진 설정 기준을 적용한 평가기준 RD와 RR의 검정력에 차이가 발생하는 이유는 앞서 언급되었듯 RR이 RD에 비해 적은 수의 sample이 필요하기 때문에 RR의 over-sampling으로 인한 결과라고 생각할 수 있다. 그렇다면 RR이 over-sampling된 원인은 애초에 RD와 RR의 동등성 구간의 폭이 달랐기 때문이라는 가정 하에, 평가기준의 동등성 마진을 조정하였다. 동등성 마진은 RD의 마진에 맞춰 RR의 동등성 마진을 조정하였고, 조정된 동등성 마진을 적용하여 RD와 RR의 신뢰구간이 동등성 구간에 들어갈 확률을 확인하는 모의실험을 수행하였다. 조정된 RR의 동등성 구간을 이용한 모의실험 수행 결과, 두 평가기준(RD, RR)의 신뢰구간이 동등성 구간에 포함될 확률은 비슷한 값을 가졌다.

위와 같은 모의실험을 통해 동일한 동등성 마진 설정기준을 적용한 RD와 RR의 동등성 구간의 폭에는 차이가 존재함을 알 수 있었다. 따라서 평가기준으로 RD와 RR를 사용하여 바이오시밀러의 동등성 평가를 수행할 때, 두 평가기준이 서로 상응하는 가치를 지니는 평가기준이 되기 위해서는 두 평가기준 간 동등성 마진의 조정이 필요할 것으로 생각된다.

참고문헌

Farrington, C. P., & Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* 9(12), 1447-1454.

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 (2014). Guideline on similar biological medicinal products containing biotechnology-derived proteins as active substance: non-clinical and clinical issues.

Food, U. and D. Administration (2016). "Non-inferiority clinical trials to establish effectiveness. Guidance for industry."

Food, U. and D. Administration (2015). "FDA's Overview of the Regulatory Guidance for the Development and Approval of Biosimilar Products in the US."

Food, U. and D. Administration (2016). "Biological Product Definitions."

Fox, A. (2010). "Biosimilar medicines new challenges for a new class of medicine." *Journal of biopharmaceutical statistics*, 3-9.

Group, I. E. W. (2001). "Choice of control group and related issues in clinical trials."

Kang, S.-H. and J.-S. Nam (2012). "Statistical Consideration of the Development of Biosimilar Products." *Korean Journal of Applied Statistics*, 125-138.

Miettinen, O. and M. Nurminen (1985). "Comparative analysis of two rates." *Statistics in medicine* 4(2), 213-226.

강승호 (2010). *신약개발에 필요한 의학통계학*, 자유아카데미.

식품의약품안전처 (2014). "동등생물의약품 평가 가이드라인".

정보라, 정은영 (2017). "2017년 다시 주목받는 바이오시밀러", 한국투자증권.

정지용 (2014). "동등성 검정에서 표본크기 및 검정력 산출", 연세대학교 대학원 석사학위 논문.

Abstract

Comparison of equivalence tests in biosimilar clinical trials for binary primary endpoint : Risk Difference and Risk Ratio

Gil, Ha Yeong

Dept. of Biostatistics and Computing

The Graduate School

Yonsei University

Biosimilars are replicative drugs that demonstrate equivalence in efficacy and safety compared to original drugs. Biosimilars can replace expensive biopharmaceuticals to provide more affordable treatment opportunities for patients, and can help ensure better quality of life through cost savings. Biosimilars are expected to grow to \$23.9 billion by 2019, as biosimilars are being developed by a number of pharmaceutical and biotech companies, with the sequential patent expiration of blockbuster antibody drugs since 2015.

As the biosimilars industry grows, the importance of equivalence test, which is often used in clinical trials to develop biosimilars, is also emerging. In particular, what is considered to be most important in equivalence test is the setting of margins, the limits that two drugs can regard as equal.

Therefore, in this study, we compare the probability that the confidence interval of RD and RR, which are the equivalence criteria, will be included in the equivalence interval through simulation. Through this simulation, we found that there is a difference in the width of equivalence between RD and RR. So We proposed a new method to adjust the equivalence interval of RR according to the RD equivalence interval.

As a result of the simulation using the proposed method, there is no significant difference in the probability that the confidence interval of RD and RR will be included in the equivalence interval. Therefore, the adjustment of equivalence interval for RD and RR is needed to RD and RR be more reliable evaluation variables.

__Key words : Biosimilar, Equivalence test, Risk Difference, Risk Ratio