

변이가 큰 약물의 생물학적 동등성 평가에 관한 연구

정규진¹ · 박상규²

¹한남대학교 비즈니스통계학과, ²중앙대학교 응용통계학과

(2011년 9월 접수, 2011년 11월 채택)

요약

본 논문에서는 변이가 큰 약물(HVD)의 생동성 평가 문제와 이를 해결하기 위하여 지금까지 수행된 연구 결과들을 소개한다. 우리나라를 비롯한 세계 여러 나라에서 현재 사용하는 전통적 생동성 평가규정 중에서 HVD에 적용 가능한 방법들과 HVD 문제 해결을 주된 목적으로 고안된 방법들로 나누어 살펴본다. 특히 scaled average bioequivalence(SABE)에 대한 연구 결과들을 통계학적인 관점에서 검토하며 SABE를 이용하여 수행하는 생동성 평가 방법에 대해 알아보고 추후 연구주제 등에 대해서도 논의한다.

주요어: 변이가 큰 약물, 생물학적 동등성, scaled 평균생동성.

1. 서론

오리지널 의약품에 대해 복제 의약품을 만들려면 복제 의약품의 효능이 오리지널 의약품과 동등함을 입증하여야한다. 생물학적 동등성(bioequivalence, 이하 생동성과 BE) 시험은 두 의약품의 약효가 동등함을 입증하는 방법으로 세계 각국에서 복제 의약품 허가에 널리 이용되고 있다. 미국에서는 1977년부터 연방규정(21CFR part 320, US Government, 2009)에 의해 복제 의약품의 허가 조건으로 생동성 시험을 의무화하였으며, 우리나라의 경우에도 1989년에 복제 의약품의 생동성 입증을 의무화하고, 1998년에 식품의약품안전청(이하 식약청과 KFDA)에서 생동성 시험기준(고지 제1998-51호)을 제정한 이후 그 기준이 계속 개정되어 현재는 고지 제2010-43호(식약청, 2010)가 사용되고 있다.

우리나라의 생동성 시험기준에 따르면 생동성 시험이란 “두 제제가 생체이용률에 있어서 통계학적으로 동등하다는 것을 입증하기 위해 실시하는 시험”로 정의된다. 그리고 생체이용률(bioavailability)이란 의약품의 효능을 나타내는 척도의 하나로 “주성분 또는 그 활성대사체가 제제로부터 전신순환혈로 흡수되는 속도와 양의 비율”로 정의 된다. 생체이용률 지표로는 혈중농도-시간곡선하면적(area under the plasma concentration-time curve; AUC_t)과 최고혈중농도(the peak concentration; C_{max})가 주로 사용된다.

전통적인 생동성 평가에서는 시험약(Test drug/formulation; T)과 대조약(Reference drug/formulation; R)의 AUC_t 와 C_{max} 모두에 대하여 평균값의 비(Geometric Mean Ratio; GMR)가 0.8과 1.25 사이에 있는지 판단한다. 이 기준은 두 제제의 평균값의 차이를 $\pm 20\%$ 허용하는 것과 같은 의미를 갖는다. 보통 지표를 로그 변환하여 사용하기 때문에 로그 변환한 지표의 평균값의 차이인 $\mu_T - \mu_R$ 이 $\ln 0.8$ 에서 $\ln 1.25$ (또는 $\pm \ln 1.25$) 내에 속하면 두 제제의 생동성을 인정한다. 통계학적으로 생동성의 평가는 주로 Schuirmann (1987)의 검정법인 TOST(two one-sided test)에 기초하지만, $\mu_T - \mu_R$ 의 신

²교신저자: (156-756) 서울시 동작구 흑석동 221, 중앙대학교 응용통계학과, 교수. E-mail: spark@cau.ac.kr

퇴구간이 생동성 판정한계(BE limit) $\pm \ln 1.25$ 내에 속하면 시험약과 대조약이 동등한 것으로 판정하는 신뢰구간 방법도 많이 활용된다. 우리나라와 미국 등 많은 나라에서 신뢰구간 방법을 권장하고 있으며 거의 대부분의 의약품에 90%의 신뢰수준이 적용된다 (식약청, 2010; US FDA, 2001).

잘 알려져 있듯이 평균차이에 대한 신뢰구간의 폭은 피험자내 변동(within-subject variability)에 비례하고 표본 크기에 반비례하기 때문에 Benet (1993)이나 Boddy 등 (1995)의 우려처럼 피험자내 변동이 큰 약물(highly variable drugs; 이하 HVD)의 경우 실제 평균에 차이가 없을 때에도 두 제제의 생동성을 입증하려면 상당히 많은 표본이 필요하여 피험자의 안전, 경비, 시간 등에 심각한 문제를 야기하게 된다. 이 문제에 대해 Benet (2006)은 HVD들은 보통 therapeutic window가 넓어서 안정성과 약효가 문제가 되는 경우가 별로 없는 것으로 믿어지기 때문에 생동성을 입증하기 위하여 많은 피험자를 요하는 전통적인 생동성 평가방법을 굳이 사용할 필요가 없다는 점을 지적하였다.

최근 20년 동안 US FDA나 Health Canada, 유럽의 EMEA 등 각국의 국가기관을 중심으로 이 문제에 대한 연구가 진행되어왔으며 이후에도 많은 연구가 수행될 것으로 생각된다. 본 논문에서는 생동성 시험에서 나타나는 HVD의 현황과 HVD가 야기하는 문제점, HVD의 생동성 평가 문제와 이를 해결하기 위하여 지금까지 수행된 연구 결과들을 2장에서 소개한다. 또한 HVD 문제에 대하여 제시된 해결 방안들을 우리나라를 비롯한 세계 여러 나라에서 현재 사용하는 전통적 생동성 평가규정 중에서 HVD에 적용 가능한 방법들과 HVD 문제 해결을 주된 목적으로 고안된 새로운 방법들을 3장에서 살펴본다. 특히 4장에서는 US FDA의 제안과 EMEA의 규정을 중심으로, scaled average bioequivalence(이하 SABE) 방법을 이용하여 수행되는 생동성 평가 방법에 대해 SABE에 대한 연구 결과들을 통계학적인 관점에서 검토하며 추후 연구주제 등에 대해서도 논의한다.

2. HVD 문제

2.1. 평균 생동성(Average Bioequivalence) 평가

우리나라 생동성 시험기준 (식약청, 2010)에 따르면 생동성 시험은 2×2 교차설계법(cross-over design)을 기본으로 하고 피험자 수는 적절한 통계처리를 위하여 최소 군당 12명, 총 피험자 수 24명 이상이어야 한다. 만약 다른 시험디자인을 사용하고자 하는 경우에는 시험의 타당성 있는 근거를 제시하여야 한다고 규정되어있다. 또 이 기준에는 두 제제의 생동성을 인정받기 위해서는 “비교평가항목치를 로그 변환하여 통계 처리하였을 때, 로그변환한 평균치 차의 90% 신뢰구간이 $\ln 0.8$ 에서 $\ln 1.25$ 이내이어야 한다.”고 명시되어있다. 미국을 비롯한 세계 각국의 규정도 이와 크게 다르지 않다. 다만 US FDA, EMEA, Health Canada 등에서는 로그변환하기 전 오리지널 지표의 평균값의 비(Geometric means ratio; 이하 GMR)에 대한 신뢰구간이 0.8과 1.25 사이에 있으면 두 제제의 생동성을 인정한다. 그러나 이들도 역시 지표를 로그변환하여 사용할 것을 권고하고 있다. 이 접근법은 두 제제의 평균을 비교한다 하여 평균 생동성(average bioequivalence; 이하 ABE)이라 불린다.

ABE 문제에서 관심 가설은 식 (2.1)로 표현된다.

$$-\theta_A \leq \mu_T - \mu_R \leq \theta_A. \quad (2.1)$$

일반적인 가설검정법의 구조에서 가설 (2.1)은 대립가설이 되고 이와 반대되는 가설이 귀무가설이 된다. 식 (2.1)에서 θ_A 는 ABE의 허용한계(acceptance limit, 보통 $\ln 1.25$)를 나타내며 이 가설이 만족되면 두 제제의 평균 생동성을 인정한다. 이 검정에서 제1종 오류는 두 제제의 평균이 같지 않은 데도 생동성을 인정하기 때문에 소비자 위험(consumer's risk)이라 부른다. 제2종 오류는 두 제제의 평균이 같지만 검정 결과 생동성이 입증되지 않는 경우이기 때문에 생산자 위험(producer's risk)이라 부르며 검정력의

비교는 많은 경우 $GMR = 1$ 경우에 이루어진다. 생동성 연구에서 가설 (2.1)의 채택 여부는 모형에 의존하는데, 주어진 모형 하에서 $\mu_T - \mu_R$ 의 90% 신뢰구간이 $\pm\theta_A$ 이내에 속하면 생동성 가설 (2.1)을 채택하게 된다.

각국의 생동성 시험 기준에서 권장하고 있는 2×2 교차설계법의 일반적인 모형은 다음과 같다:

$$y_{ijk} = \mu + G_k + S_{ik} + P_j + F_{(j,k)} + \epsilon_{ijk} \tag{2.2}$$

$$S_{ik} \sim iid N(0, \sigma_s^2), \quad \epsilon_{ijk} \sim iid N(0, \sigma_e^2), \quad i = 1, \dots, n_k, \quad j, k = 1, 2.$$

y_{ijk} 는 AUC 혹은 C_{max} 등 로그 변환된 생체이용률 값을 의미하며, μ 는 전체평균을 나타내고, 순서효과(G), 기간효과(P), 제제효과(F)는 다음의 조건을 만족하는 생동성 모수들이다.

$$\sum_{k=1}^2 G_k = 0, \quad \sum_{j=1}^2 P_j = 0, \quad \sum_{j=1}^2 \sum_{k=1}^2 F_{(j,k)} = 0, \quad F_{(j,k)} = \begin{cases} F_R, & (j, k) = (1, 1), (2, 2), \\ F_T, & (j, k) = (1, 2), (2, 1). \end{cases}$$

S_{ij} 는 k -번째 순서집단에 투입된 i -번째 피험자에게 관측되는 피험자의 변동성을 의미하는 확률변수이고, ϵ_{ijk} 는 k -번째 순서에 j -번째 기간에 i -번째 피험자에게 관측되는 제제의 변동성을 의미하는 확률변수이다. S_{ik} 와 ϵ_{ijk} 는 통계적으로 서로 독립을 가정하고, S_{ik} 의 분산 σ_s^2 는 피험자에 기인한 변동, ϵ_{ijk} 의 분산 σ_e^2 는 제제에 기인한 변동을 의미한다. 식 (2.2)와 같은 통계모형에서 두 제제의 모평균은 $\mu_T = \mu + F_T$ 와 $\mu_R = \mu + F_R$ 로 정의된다. 따라서 ABE 모수 $\mu_T - \mu_R$ 은 제제효과의 차이 $F_T - F_R$ 과 같아진다.

두 집단의 크기가 동일하다고 하고 모형 (2.2)에서 생동성 가설 (2.1)의 채택영역을 구하면 다음과 같다.

$$-\theta_A \leq \bar{y}_T - \bar{y}_R \pm t(N - 2; 0.05) \sqrt{\frac{2}{N} \hat{\sigma}_e^2} \leq \theta_A. \tag{2.3}$$

식 (2.3)에서 \bar{y}_T 와 \bar{y}_R 는 두 제제의 표본평균으로 모평균 μ_T 와 μ_R 의 추정량이고, N 은 두 집단 피험자의 합, $\hat{\sigma}_e^2$ 는 모형 (2.2)로 전개한 분산분석표의 잔차제곱평균(residual mean square)이다 (Chow와 Liu, 2008). 2×2 교차설계법에서는 시험제제와 대조제제의 피험자내 분산과 피험자-제제 교호작용 분산(subject-formulation interaction variance, 2.3절 식 (2.8))을 모두 포함한다. 만약 이들을 분리하고 싶으면 2.3절에 기술되는 반복설계법(replicate design)이 필요하다. 2×2 교차설계법에서는 $\hat{\sigma}_e^2$ 을 피험자 내 변동(within-subject variability)의 추정량으로 간주한다.

식 (2.3)의 가운데 항은 $\mu_T - \mu_R$ 의 90% 신뢰구간을 나타내는데 신뢰구간의 폭은 피험자 내 변동 $\hat{\sigma}_e$ 에 비례하고 총 피험자수의 제곱근에 반비례한다. 따라서 피험자 내 변동이 큰 경우 생동성 검정을 통과하려면 상당히 큰 수의 피험자가 필요하다. 표 2.1은 90%의 검정력을 얻기 위하여 필요한 표본크기를 나타내는데, 이는 Boddy 등 (1995)에서 발췌하였다. 표 2.1을 보면 두 제제의 모평균이 같은 $GMR = 1.00$ 의 경우 상당히 큰 표본이 필요할 뿐만 아니라 $GMR = 1.15$ 인 경우에도 현실적으로 불가능할 정도의 표본이 요구된다. 이와 같이 피험자 내 변동이 커지면 표본크기가 기하급수적으로 커지는 현상은 US FDA Guidance (2001)나 Patterson 등 (2001)에도 지적되고 있다. 따라서 피험자 내 변동이 큰 약물의 경우 적절한 크기의 피험자수로는 식 (2.3)의 생동성 기준을 통과하기가 어렵고 이에 맞는 특별한 생동성 기준이 필요한 것으로 여겨지고 있다.

참고로 생동성 연구에서는 피험자 내 변동을 원 데이터의 피험자 내 변동계수(within-subject coefficient of variation) CV_W 를 이용하여 표현하기도 하는데 모형 (2.2)에서 로그 변환된 데이터의 피험자 내 변동 σ_e 와 원 데이터의 변동계수 CV_W 의 관계는 다음과 같다.

$$\sigma_e^2 = \ln(CV_W^2) + 1. \tag{2.4}$$

표 2.1. 검정력 90%를 얻는데 필요한 표본크기

CV _W (%)	σ_e	GMR			
		1.00	1.05	1.10	1.15
25	0.246	30	36	66	152
30	0.294	40	52	92	214
35	0.340	54	68	124	288
40	0.385	68	86	158	368
45	0.429	84	106	196	456
50	0.472	100	128	236	552

식 (2.4)를 이용하여 계산하면 CV_W 값 25, 30, 40, 50%에 대응되는 σ_e 값은 각각 0.246, 0.294, 0.385, 0.472이다.

2.2. HVD의 생동성 평가 문제

Benet (2006)의 지적처럼 단지 변동이 크다는 이유로 안정성에 문제가 없는 약에 대해서조차 많은 건 강한 피험자를 약물에 노출시키는 것은 윤리적으로 문제가 있어 보인다. 미국 연방규정에는 생동성 연구에 대한 가이드라인을 제시하고 있는데 그중 첫 번째 원칙으로 불필요한 인체 연구를 금지하고 있으며(21CFR part 320 Sec. 320.25 (a)), 이에 따라 US FDA는 생동성 시험에서 피험자수를 줄이는 문제에 많은 관심을 기울여 왔다. Boddy 등 (1995)는 이 문제가 생동성 연구에서 상당히 오래된 주제임을 말해준다. 피험자수가 많아지면 윤리적인 문제와 함께 경제적인 문제도 야기하는데 Tothfalusi 등 (2009)는 필요한 표본크기가 커지면 비용증가로 인하여 중소기업의 제약산업 진입이 어려워져 결국 약값이 올라가는 요인으로 작용할 수 있음을 지적하였다. 이들은 또 피험자내 변동이 크면 신뢰구간의 폭이 넓어지기 때문에 생동성 기준을 통과하기가 어려워져 어떤 약이 그 자신과의 생동성 검정조차 통과하지 못하는 경우가 생길 수 있음을 언급하였다.

이와 같은 HVD 문제를 해결하기 위하여 많은 연구들이 이루어져 왔지만 아직 합의에 이른 내용은 별로 많지 않은 것 같다. 다만 국제적 컨퍼런스인 Bio-international 92 이후로 거의 모든 연구에서 HVD란 AUC_t와 C_{max}의 CV_W가 30% (피험자내 변동이 0.294) 이상인 약물로 정의 되고 있다.

HVD에 대한 실태 연구는 주로 미국에서 평가된 사례를 중심으로 이루어지고 있다. Tanguay 등 (2002)이 800개 연구사례에 대하여 CV_W의 크기에 따라 생동성 검정을 통과하지 못한 비율을 조사한 결과 CV_W가 < 10%, 10-20%, 20-30%일 때 각각 6%, 10%, 26%인 반면 CV_W가 > 30%인 경우 이 비율이 62%에 이르렀고, CV_W ≥ 35%인 경우에는 무려 85%까지 증가하였다. 이들은 또 표본크기에 따른 생동성검정의 실패 비율도 살펴보았는데 표본크기 보다는 CV_W의 크기가 실패 비율에 더 직접적인 관련이 있음을 발견하였다. 이 연구 결과는 Midha (2006)와 Blume (2010)에도 인용되어있다.

Davit 등 (2008)은 2003년부터 2005년 사이에 US FDA의 OGD(Office of Generic Drugs)에서 평가 되어 승인을 얻은 180개의 서로 다른 약물의 524개의 상품에 대한 1,010개의 연구사례를 분석하였다. CV_W에 대한 추정치로는 ANOVA RMSE(root mean square error)를 이용하여 이 값이 0.3 이상이면 HVD로 간주하였다. HVD의 비율은 약물 기준으로 32% (57/180), 상품 기준으로 19% (101/524), 연구사례 기준으로 11% (111/1010)로 나타났는데 승인을 얻지 못한 연구까지 포함한다면 이 비율은 훨씬 높을 것으로 생각된다. 이들의 연구에서 HVD로 분류된 연구사례의 RMSE 값은 AUC_t와 C_{max}에 대하여 각각 0.30-0.55, 0.30-0.60 사이에 놓여있으며, 각 연구사례의 피험자수는 AUC_t에 대하여 변동이 작은 경우(RMSE < 0.3)와 큰 경우(RMSE ≥ 0.3) 각각 12-113명 (평균 32명), 24-134명 (평균 55명)

이고, C_{max} 에 대하여는 변동이 작은 경우와 큰 경우 각각 12-113명 (평균 31명), 18-134명 (평균 46명)으로 나타났다. 그 밖에 약물을 기준으로 할 때의 HVD 비율이 연구사례 기준으로 하는 경우와 다른 이 유나 HVD의 원인 등에 대한 연구 결과는 이들의 논문에 자세히 설명되어 있다.

Davit 등 (2008)의 연구 결과를 보면 CV_W 는 대부분 30-40%의 값을 갖으며 가장 큰 경우에도 60%를 넘지 않는다. 또 CV_W 나 피험자수에서 AUC_t 와 C_{max} 의 차이는 별로 없는 것으로 보인다. 그러나 이 연구는 FDA로부터 승인을 얻은 연구사례만을 대상으로 하기 때문에 이런 결과가 얻어진 것으로 보이며, 이와는 상당히 다른 연구결과들도 찾아볼 수 있다.

CV_W 가 상당히 큰 경우로 DiLiberti (2007)는 71명의 피험자를 대상으로 한 시험에서 AUC_t 와 C_{max} 의 CV_W 가 각각 157%, 173%나 되는 경우를 언급하였다. 이 경우 80% 검정력에서 ABE 검정을 통과하려면 598명의 피험자가 필요하게 된다. 일반적으로 C_{max} 가 AUC_t 보다 변동이 큰 것으로 알려져 있는데 한 예로 Davit (2004)은 US FDA의 DBE(Division of Bioequivalence)에서 2003년에 평가한 212건의 연구사례를 분석하여 RMSE가 0.3 이상인 경우 33건을 발견하였는데 이 중에서 C_{max} 만 HVD인 경우가 28건인데 반해 AUC_t 만 HVD인 경우는 한 건도 없었음을 보고하고 있다.

AUC_t 보다는 변동이 큰 C_{max} 에 의하여 HVD로 판정되는 경우가 많기 때문에 C_{max} 에 대한 처리방법이 오랫동안 관심을 끌어왔다. Endrenyi 등 (1991)는 AUC_t 는 제제가 전신순환혈로 흡수되는 양에만 관계가 있는 반면 C_{max} 는 양과 속도 모두와 연관되어진다는 점을 지적하고 이를 해결하기 위하여 C_{max}/AUC_t 를 새로운 생체이용률 지표로 제안하였다. DiLiberti (2007)는 앞서 언급한 사례에 이 지표를 적용하여 CV_W 가 43%로 낮아짐을 보이고, C_{max}/AUC_t 가 개별 지표에 비하여 변동이 훨씬 적기 때문에 HVD문제에서 이 지표를 이용하면 많은 경우 CV_W 가 30% 아래로 떨어진다는 점을 언급하였다. 그러나 이 지표를 포함하여 새로운 지표의 이용에 대한 특별한 합의에 이르지 못한 것으로 보인다. 다만 C_{max} 에 대한 처리방법으로 대조제제의 CV_W 가 커짐에 따라서 생동성 검정에 사용되는 신뢰구간의 폭을 넓혀 피험자수를 늘리지 않고 검정을 통과하는 비율을 높이는 방안이 지난 20년 동안 집중적으로 연구되었는데 다음에 살펴볼 개인 생동성(individual bioequivalence; IBE)과 SABE가 이런 생각에 기초하고 있다.

2.3. 개인 생동성(IBE) 평가

평균 생동성은 두 제제의 평균만을 고려하여 평균의 차이가 $\pm 20\%$ 범위 내에 있으면 생동성을 인정하는 방법이다. 이 방법은 분산을 고려하지 않기 때문에 평균이 같은 제제의 생동성이 기각되거나 반대로 평균이 확실히 다른 제제의 생동성도 인정되는 문제가 발생하게 된다. 생동성 평가에서 분산을 고려한 척도로 Scheiner (1992)는 $E(y_T - y_R)^2 / E(y_R - y'_R)^2$ 를, Schall과 Luus (1993)는

$$E(y_T - y_R)^2 - E(y_R - y'_R)^2 \quad (2.5)$$

를 제안하였는데 두 척도는 동일한 것으로 인식되며, Schall과 Luus (1993)의 척도를 보통 두 번째 항인 대조제제의 분산으로 나누어(scoring) 사용한다. 모형과 scaling 방법 등에 따라 조금씩 다른 수많은 척도가 고안되었는데 US FDA에서는 집단 생동성(population bioequivalence; PBE)과 개인 생동성(IBE)을 규정에 포함하여 운영하고 있다. 집단 생동성과 개인 생동성은 집단 내에서 혹은 개인에 대하여 복제의약품이 오리지널 의약품을 대체할 수 있는가를 평가하는 개념으로 전자를 약의 대체조제 가능성(prescribability), 후자를 교환 가능성(switchability)이라 부르기도 한다. PBE와 IBE에 대한 보다 자세한 내용은 박상규 등 (2005), 정규진 등 (2010), Chow와 Liu (2008)를 참조할 수 있다.

US FDA는 PBE와 IBE의 평가를 위해 시험기간(period)이 3이상인 2×3 나 2×4 교차설계법의 사용을 권하고 있는데 이런 모형을 반복설계법(replicate design)이라 한다. y_{ijk} 를 k 번째 순서(sequence)의

i 번째 피험자(subject)로부터 j 번째 기간(period)에 얻어진 반응값 ($i = 1, \dots, n_i; j = 1, 2, 3, \text{ or } 4, k = 1, 2$)이라 할 때 그 모형은 다음과 같이 표현된다.

$$y_{ijk} = \mu + F_l + W_{ljk} + S_{ikl} + e_{ijk}, \quad (2.6)$$

여기서 첨자 $l = T \text{ or } R$ 이고, W_{ljk} 는 모형 (2.2)와 다르게 순서효과, 기간효과, 교호작용효과, 또는 모형에 따라 잔류효과를 포함하는 고정효과(fixed effects)로 $\sum_k \overline{W_{lk}} = 0$ ($\overline{W_{lk}}$ 는 고정된 (l, k) 에 대한 W_{ljk} 의 평균)이다. 피험자의 변동성을 나타내는 랜덤효과 S_{ikl} 은 (S_{ikT}, S_{ikR}) 가 서로 독립이고 동일하게 평균 0, 분산 공분산 행렬이

$$\begin{bmatrix} \sigma_{BT}^2 & \rho\sigma_{BT}\sigma_{BR} \\ & \sigma_{BR}^2 \end{bmatrix}$$

인 이변량 정규분포를 따르는 확률 벡터이다. 피험자내 제제의 변동성을 의미하는 오차항 e_{ijk} 는 서로 독립이며 동일하게 평균 0, 분산 σ_{Wl}^2 ($l = T \text{ or } R$)인 정규분포를 따르고, S_{ikl} 와 e_{ijk} 는 서로 독립이라 가정한다.

모형 (2.2)에서는 T 와 R 에 상관없이 랜덤효과의 분산과 오차항의 분산이 같다고 가정하기 때문에 ($\sigma_{BT}^2 = \sigma_{BR}^2 = \sigma_s^2, \sigma_{WT}^2 = \sigma_{WR}^2 = \sigma_e^2$) 반응값의 분산이 동일($\text{Var}(y) = \sigma^2 = \sigma_s^2 + \sigma_e^2$)하였으나, 모형 (2.6)에서는 T 와 R 에 따라 랜덤효과의 분산과 오차항의 분산이 다르다고 가정하고 있기 때문에 반응값의 분산도 다음과 같이 서로 다르게 표현된다.

$$\text{Var}(y_T) = \sigma_{TT}^2 = \sigma_{BT}^2 + \sigma_{WT}^2, \quad \text{Var}(y_R) = \sigma_{TR}^2 = \sigma_{BR}^2 + \sigma_{WR}^2. \quad (2.7)$$

또, 모형 (2.6)으로부터 피험자-제제 교호작용 분산(subject-formulation interaction variance) σ_D^2 을 얻게 된다.

$$\sigma_D^2 = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR}. \quad (2.8)$$

모형 (2.6)에서 Schall과 Luus (1993)의 척도인 식 (2.5)를 계산하면, 식 (2.7)과 (2.8)로부터

$$E(y_T - y_R)^2 - E(y_R - y'_R)^2 = (\mu_T - \mu_R)^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2 \quad (2.9)$$

이 얻어진다.

IBE와 PBE는 식 (2.9)를 scaling함으로써 얻어지는데, FDA에서 제시하는 기준은 각각 다음과 같다.

$$\frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2)}{\max\{\sigma_{WR}^2, \sigma_{w0}^2\}} < \theta_I, \quad (2.10)$$

$$\frac{(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)}{\max\{\sigma_{TR}^2, \sigma_{T0}^2\}} < \theta_P, \quad (2.11)$$

여기서 θ_I 와 θ_P 는 각각 IBE와 PBE의 허용한계를 나타낸다. IBE와 PBE는 의미와 성질이 비슷하기 때문에 이후의 논의에서는 IBE만을 고려하기로 한다.

IBE 기준식 (2.10)에서 σ_{w0}^2 은 US FDA에서 정한 상수로 $\sigma_{WR} > \sigma_{w0}$ 면 기준식의 분모가 σ_{WR}^2 이 되고 이 방법을 reference-scaled method라 부른다. 만약 $\sigma_{WR} \leq \sigma_{w0}$ 면 분모가 상수 σ_{w0}^2 이 되며 이를 constant-scaled method, 두 방법을 합친 식 (2.10)의 방법을 mixed scaled method라 부른다. IBE에 대하여 US FDA가 제시하는 상수값은 $\sigma_{w0} = 0.2$ 이고 한계값은

$$\theta_I = \frac{(\ln 1.25)^2 + 0.03 + 0.02}{0.2^2} = 2.4948$$

인데 여기서 분자의 0.03과 0.02는 각각 σ_D^2 과 $\sigma_{WT}^2 - \sigma_{WR}^2$ 의 허용한계를 나타낸다. σ_{W0} 처럼 관계당국에서 규정하는 상수를 규정상수(regulatory constant)라 부른다.

식 (2.10)의 왼쪽 부분은 분자가 평균의 차이, 피험자-제제 교호작용 분산, 그리고 피험자내 분산의 차이를 나타내는 세 개의 항으로 구성되어 있기 때문에 이를 나누어 검정하는 방법을 생각할 수 있는데 이를 분리형(disaggregate) 방법이라 부르고 이와는 반대로 식 (2.10)과 같이 전체를 하나로 하여 검정하는 방법을 일체형(aggregate) 방법이라 한다. 분리형 검정은 검정력이 떨어지기 때문에 보통 일체형을 선호하게 된다.

만약 식 (2.10)에서 피험자-제제 교호작용이 없고 두 제제의 피험자내 분산이 같다면 다음과 같이 평균의 차이에 대한 부분만 남게 된다.

$$\frac{(\mu_T - \mu_R)^2}{\max\{\sigma_{WR}^2, \sigma_{W0}^2\}} < \frac{(\ln 1.25)^2}{\sigma_{W0}^2}. \tag{2.12}$$

양변에서 분모를 제거하면 식 (2.12)는 ABE의 기준과 동일하고, 분모를 함께 보면 좌변의 분모가 우변의 분모보다 크기 때문에 이 구간은 ABE 구간보다 더 넓다는 것을 알 수 있다. 즉, scaling이 BE 한계를 넓히는 역할을 하는데 US FDA Guidance (2001)는 이 방법이 피험자내 변동이 큰 경우에 대처하기 위한 것이라는 점을 밝히고 있다. 다음에 살펴볼 SABE 역시 이와 같은 방법으로 BE 구간을 넓히는 생각에 기초한다.

US FDA Guidance (2001)에 따르면 IBE나 PBE 기준을 통과하려면 식 (2.10)과 (2.11)을 만족하는 것 외에, GMR의 점추정치가 80~125% 이내에 속할 것을 요구하고 있다. 어떤 연구에서는 이 조건을 point estimate constraint라 부르는데, 이후 우리는 이 조건을 GMR constraint라 부르기로 한다. 만약 생체이용률 지표의 로그 변환한 값이 평균 μ , 분산 σ^2 인 정규분포를 따른다면 이 지표의 기댓값은 $\mu + \sigma^2/2$ 이므로 시험제제와 대조제제의 분산이 같지 않다면 GMR은 일반적으로 $\exp(\mu_T - \mu_R)$ 와는 다른 값이 된다. 다만 ABE의 기본이 되는 2x2 교차설계 모형 (2.2)에서는 두 제제의 분산이 같기 때문에 $GMR = \exp(\mu_T - \mu_R)$ 이 된다.

2.4. Scaled ABE(SABE) 평가

피험자내 분산이 큰 경우에는 두 제제의 평균이 별로 차이가 나지 않는 경우에도 ABE로는 생동성 한계를 만족시키지 못할 수가 있다. 이에 대한 해결 방법으로 앞 절에서 설명한 IBE를 사용할 수도 있지만 보다 간단하게 피험자내 표준편차 σ_{WR} 에 비례하여 생동성 한계를 넓혀주는 방법을 생각할 수 있다. 이런 방법으로 Boddy 등 (1995)는

$$\exp(\pm k\sigma_{WR}) \tag{2.13}$$

을 GMR에 대한 생동성 허용한계로 제안하였다. 여기서 k 는 표준정규분포의 백분위수로 2x2 모형에서 그들은 $k = 1.0$ 을 제안하였다. 피험자내 표준편차 σ_{WR} 은 ANOVA의 잔차분산으로부터 얻어지는 추정값을 사용한다. 식 (2.13)은 시험제제와 대조제제의 분산이 동일하다면 로그변환 하에서

$$-k\sigma_{WR} \leq \mu_T - \mu_R \leq k\sigma_{WR} \tag{2.14}$$

이 되어 ABE의 허용한계를 σ_{WR} 에 비례하여 넓혀준 것이 된다.

식 (2.14)를 σ_{WR} 로 나누면(scaling) 다음과 같은 SABE의 허용한계가 얻어진다.

$$-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{WR}} \leq \theta_S. \tag{2.15}$$

식 (2.15)로 표현되는 SABE는 기본적으로 ABE에서 사용하는 평균의 차이 $\mu_T - \mu_R$ 를 표준편차 σ_{WR} 로 scaling함으로써 얻어진다. scaling은 σ_{WR} 이 생동성 검정에 미치는 영향을 줄여 주기 때문에 피험자내 분산이 큰 경우에도 적은 수의 피험자만으로 생동성 검정을 통과할 수 있게 된다.

SABE 한계값 θ_S 는

$$\theta_S = \frac{\ln(1.25)}{\sigma_{W0}} \quad (2.16)$$

의 모양으로 주어진다. IBE와 유사하게 SABE에서도 mixed strategy 개념을 적용한다. 보통 CV_W 가 30%보다 작으면 식 (2.1)의 ABE를 사용하고 CV_W 가 30%보다 크면 식 (2.15)나 (2.16)의 SABE 허용한계를 사용한다. σ_{W0} 는 관계당국에서 지정하는 규정상수로 SABE의 허용한계에 직접 영향을 미친다. 보통 0.250($CV_W = 25.4\%$)이나 0.294($CV_W = 30\%$)에 대한 연구가 많이 이루어지고 있으며 이때 θ_S 는 0.890과 0.760이 된다. σ_{W0} 의 의미를 포함하여 이 값들의 비교는 3장에 주어진다.

SABE에 대하여 많은 연구가 이루어졌는데 최근의 연구로는 Tothfalusi와 Endrenyi (2003), Karalis 등 (2004), Haidar 등 (2008a), Haidar 등 (2008b), Tothfalusi 등 (2009) 등을 들 수 있다. 이 중 Tothfalusi 등 (2009)은 HVD 문제와 SABE에 대해 지금까지 이루어진 연구결과들을 잘 요약하고 있다.

2.5. HVD에 대한 처리방법 및 각국의 규정

HVD에 대한 대처 방법은 크게 생체이용률 지표의 변동을 줄이는 방법과 생동성 허용구간을 확장하는 방법으로 나누어 볼 수 있다. 변동을 줄이는 방법 중에서 여러 나라에서 규정으로 채택하고 있는 것으로 반복투약(multiple dose)과 추가시험(add-on test)이 있다.

반복투약은 우리나라를 비롯하여 미국, EU, 일본 등에서 채택하고 있다. 우리나라 규정에는 개인에 따라 생체이용률의 차이가 큰 경우 등에 약제를 반복 투약하여 정상상태(steady state) 농도에 도달시켜 시험할 수 있는 것으로 되어있다. 반복투약은 일회투약에 비해 보통 변이가 줄어드는 것으로 알려져 있으나 Tothfalusi 등 (2009)의 지적처럼 반복 투약한 경우 C_{max} 의 변동이 일회투약에 비해 더 큰 경우가 존재하는 등, 한계를 지닌 것으로 평가 받고 있다.

추가시험은 처음에 계획된 생동성 시험으로 생동성 입증되지 않는 경우에 행하는 시험으로 우리나라, 캐나다, 일본 등에서 이를 사용하고 있다. 우리나라의 규정을 보면 1회에 한하여 추가시험을 실시할 수 있으며, 그 결과는 선행된 결과와 총괄해서 최종 분석되어야 한다고 되어있다 (식약청, 2010; 박상규 등, 2010). 결국 이 방법은 일차 시험으로 생동성이 입증되지 않을 때 피험자수를 늘리는 방법이기 때문에 HVD의 근본 문제인 표본크기를 전혀 해결하지 못하는 방법이다.

피험자수를 늘리지 않고 표본 수를 늘이는 방법으로 2×2 교차설계법 대신 3개 이상의 기간(period)을 갖는 고차원의 교차설계법을 사용하면 된다. 2×3 이상의 설계법에서는 R 과 T 가 반복하여 시험이 이루어지기 때문에 반복설계법(replicate design)이라 부른다. 이 설계법은 우리나라, 미국을 포함하여 대부분의 나라에서 사용을 권장하고 있다. 반복설계법은 피험자수를 늘리지 않으면서 높은 검정력을 얻을 수 있는 장점이 있으나 2×2 설계법에 비해 피험자들을 장시간 약물에 노출시켜야 하는 단점도 갖는다.

변이를 줄일 수 있는 다른 방법으로 2.2절에서 언급한 대로 C_{max} 대신 변동이 작은 C_{max}/AUC_t 를 새로운 생체이용률 지표로 사용하는 방법과 정규진 등 (2011)에 제안된 것으로 생체이용률 지표에 영향을 주는 공변량(covariate)이 존재할 때 이를 고려하여 생동성 검정을 하는 방법 등이 있으나 아직 일반화되지는 못하고 있다.

HVD 문제에 대처하는 최근의 추세는 변동을 줄이는 방법보다는 생동성 허용구간을 확장하는 방법으로 흐르고 있는 것으로 보인다. 가장 극단적인 방법은 변동이 큰 C_{max} 를 생동성 검정에서 제외하는 것인

데 캐나다가 이 방법을 사용하고 있다. Health Canada (2003)는 HVD에 대한 해결 방안을 다각도로 검토하였으나 아직 결론에 이르지 못한 것 같고, 다만 Health Canada (1992)의 규정에는 90% 신뢰 구간을 이용하는 생동성 검정은 상대적으로 변동이 작은 AUC_t 에만 적용하고 C_{max} 에 대해서는 GMR constraint만 만족하면 되는 것으로 되어 있다.

C_{max} 에 대한 BE 허용한계를 확장하는 방법도 있는데, EMEA (2008)는 2010년에 새 규정이 나오기 전까지 C_{max} 에 한하여 BE 허용한계를 80~125%에서 75~133%로 확장하여 적용할 수 있도록 규정하고 있었다. 캐나다와 EU의 방법이 통계적 근거가 약한 방법이라면 US FDA Guidance (2001)는 확실한 통계적 방법으로 PBE나 IBE를 GMR constraint와 함께 사용할 수 있도록 규정하고 있다.

IBE는 피험자내 변동에 비례하여 BE 허용한계를 확장함으로써 HVD 문제를 해결하는 기능 이외에 Benet (2006)의 언급처럼 switchability 개념을 포함하고 있으며, 제제의 분산의 크기가 고려되고 있고, 피험자-제제 교호작용까지 담고 있다. 이와 같은 이유로 1990년대와 2000년대 초반까지 IBE는 생동성 연구의 많은 문제를 해결해 줄 것으로 기대되었고 실제로 이 기간에 수많은 연구가 이루어졌다. 그러나 Tothfalusi와 Endrenyi (2003)나 Benet (2006)의 지적처럼 IBE의 장점으로 인식되던 내용들은 시간이 흐르면서 가치를 잃게 되었다. switchability라는 개념이 과연 필요한 것인지 회의하게 되고, 피험자-제제 교호작용은 이해하기 어려운 모수로 여겨지고, 피험자내 변동으로 scaling하는 것은 ABE만으로도 충분하다는 인식이 점차 힘을 얻게 되었다. 이와 같은 이유로 미국 FDA조차 IBE에서 SABE로 연구의 중심을 옮기고 있는 것으로 생각된다.

SABE에 대한 규정은 EMEA (2010)에서 찾을 수 있는데, 여기에는 C_{max} 에 한하여 식 (2.13)에 주어진 Boddy 등 (1995)의 한계를 69.84%~143.19% 이내에서 SABE의 BE 한계로 이용할 수 있도록 규정하고 있으며, GMR constraint도 함께 사용하도록 요구하고 있다.

SABE는 피험자내 변동에 비례하여 생동성 허용구간을 확장하는 방법을 대표하는 것으로 HVD에 대한 해결방법으로 최근 미국, 캐나다, 유럽에서 많은 주목을 받고 있다. Haidar 등 (2008a)와 Haidar 등 (2008b)의 연구 결과에서 알 수 있듯이 특히 US FDA를 중심으로 최근 10여 년 동안 많은 연구가 이루어졌으나 아직 특정한 기준을 제시하지는 못하고 있는 실정이다. 우리나라나 일본은 아직 허용구간을 확장하는 방법을 채택하지 않고 있으며 연구도 미진한 편이라 생각된다. SABE에 대한 EMEA의 규정과 US FDA의 권고사항에 대한 내용은 다음 장에서 보다 상세히 살펴본다.

3. SABE에 의한 생동성 평가

3.1. SABE의 해석

IBE의 특수형태

SABE의 정의식 (2.15)를 제공하면 아래의 식이 얻어진다.

$$\frac{(\mu_T - \mu_R)^2}{\sigma_{WR}^2} \leq \theta_S^2. \quad (3.1)$$

이 식을 IBE의 허용한계인 식 (2.10)과 비교해 보면 식 (2.10)의 분자를 구성하는 세 개 항 중에서 첫 번째 항만 떼어낸 형태와 동일하다. 즉, SABE는 IBE의 세 분리형 검정 중에서 첫 번째 검정과 동일한 검정방법이다. 또, 식 (3.1)의 mixed strategy를 적용한 형태를 고려하면 식 (2.12)가 얻어짐을 알 수 있고, 이 식은 IBE에서 피험자-제제 교호작용이 없고 두 제제의 피험자내 분산이 같을 때 얻어진다. 따라서 SABE란 IBE의 특수한 형태로 해석할 수 있다.

두 제제를 비교하는 척도로써의 의미

$(\mu_T - \mu_R)/\sigma_{WR}$ 은 두 제제의 평균의 차이를 표준화한 값으로 해석할 수 있는데, 이를 Cohen (1988)의 아이디어에 따라 ‘표준화된 효과의 크기(standardized effect size)’라 부른다. Tothfalusi와 Endrenyi (2003)은 생체이용률 지표가 로그정규분포를 따른다면 이 양은 두 제제를 비교할 때 충분히 의미 있는 도구로 사용할 수 있다고 주장하였다. 또, Arieta (2006)는 표준화된 효과의 크기를 제공한 양을 기댓값을 이용하여 식 (3.2)의 형태로 표현하였다.

$$\frac{(\mu_T - \mu_R)^2}{\sigma_{WR}^2} = \frac{[E(y_T) - E(y_R)]^2}{E[(y_R) - E(y_R)]^2}. \quad (3.2)$$

이 식은 IBE를 유도하는데 사용된 Scheiner (1992)나 Schall과 Luus (1993)의 식 (2.5), (2.9)와 유사하다.

Boddy 등의 방법과 비교

SABE에 대해 조금 더 살펴보기 위하여 앞 장에서 정의된 ABE, Boddy 등 (1995), SABE의 식 (2.1), (2.14), (2.15)를 함께 모아 비교해 보자.

$$\begin{aligned} -\theta_A &\leq \mu_T - \mu_R \leq \theta_A. \\ -k\sigma_{WR} &\leq \mu_T - \mu_R \leq k\sigma_{WR}. \\ -\theta_S &\leq \frac{\mu_T - \mu_R}{\sigma_{WR}} \leq \theta_S. \end{aligned}$$

ABE의 허용한계는 상수인 반면 Boddy 등의 허용한계는 σ_{WR} 에 비례한다. 한편, SABE는 식 (2.15)의 각 변에 σ_{WR} 를 곱하면 Boddy 등과 같이 σ_{WR} 에 비례하는 허용한계를 갖는다. 언뜻 보면 Boddy 등과 SABE가 동일한 방법으로 보이지만 Boddy 등의 방법은 $\mu_T - \mu_R$ 의 신뢰구간을 구하여 이 구간이 허용한계 $\pm k\sigma_{WR}$ 사이에 포함되는지를 판단한다. 따라서 이 방법은 허용한계 $\pm k\sigma_{WR}$ 를 구하기 위한 σ_{WR} 의 추정값이 필요하다. 이와는 다르게 SABE는 $(\mu_T - \mu_R)/\sigma_{WR}$ 의 신뢰구간을 구하여 생동성을 평가한다. Tothfalusi와 Endrenyi (2003), Arieta (2006)은 두 방법이 대표본에서 특히 비슷한 성질을 보이지만, Boddy 등의 방법은 추정값이 포함되어 허용한계가 상수가 아닌 변수라 SABE가 더 선호될 수 있는 방법임을 지적하였다.

규정상수 σ_{W0} 의 의미와 종류

SABE의 정의식을 식 (2.14)의 형태로 고쳐 써보면 허용한계가

$$\theta_S \sigma_{WR} = \frac{\ln(1.25)}{\sigma_{W0}} \sigma_{WR} \quad (3.3)$$

가 된다. 이는 피험자내 표준편차 σ_{WR} 에 대한 1차식으로 비례상수인 기울기 값은 $\theta_S = \ln(1.25)/\sigma_{W0}$ 이 된다. 규정상수 값 σ_{W0} 는 기울기에 관여를 하는데 σ_{W0} 의 값이 작을수록 기울기는 급해지고 허용한계는 넓어진다. Endrenyi와 Tothfalusi (2009)의 그림 1은 $\sigma_{W0} = 0.246, 0.294(CV_{W0}, 30\%)$ 일 때 mixed SABE의 허용한계를 잘 보여주고 있다.

규정상수값으로는 FDA에서 권장하는 IBE의 규정상수값인 $\sigma_{W0} = 0.200$, Boddy 등의 방법에서 그들이 제안한 값 $k = 1$ 에 대응되는 값 $\sigma_{W0} = \ln 1.25 = 0.223$, US FDA에서 SABE의 규정상수로 권장하는 값 $\sigma_{W0} = 0.250$, HVD로 판단되는 기준값인 $CV_{W0} = 30\%$ 에 대응되는 값 $\sigma_{W0} = 0.294$ 등을 고려할 수 있다. 표 3.1은 몇 가지 규정상수 σ_{W0} 와 피험자내 표준편차 σ_{WR} 의 값에 따라 형성되는 SABE의 GMR에 대한 허용한계를 보여준다.

표 3.1. σ_{W0} 와 σ_{WR} 에 따른 SABE의 허용한계 (%)

CV _W (%)	σ_{WR}	σ_{W0}			
		0.200	0.223	0.250	0.294
30	0.294	72.0-138.8	74.5-134.2	76.9-130.0	80.0-125.0
35	0.340	68.4-146.1	71.2-140.5	73.8-135.5	77.3-129.4
40	0.385	65.1-153.7	68.0-147.0	70.9-141.0	74.7-133.9
45	0.429	62.0-161.4	65.1-153.6	68.2-146.7	72.2-138.5
50	0.472	59.1-169.3	62.4-160.4	65.6-152.4	69.9-143.1

3.2. SABE의 생동성 평가 방법

SABE를 이용하여 생동성을 평가하기 위해서는 SABE의 식 (2.15)에 나타나는 $(\mu_T - \mu_R)/\sigma_{WR}$ 의 신뢰구간이 필요하다. 신뢰구간 계산에 3가지 방법이 사용될 수 있다. Tothfalusi와 Endrenyi (2003)와 Wellek (2003)이 각각 제안한 비중심(noncentral) t -분포를 이용하는 정확한 계산법, Howe (1974)의 아이디어를 생동성 시험에 적용한 Hyslop 등 (2000)의 선형화 근사 계산법, Boddy 등 (1995)의 방법이 그것이다. 2×2 교차설계 모형에서는 세 방법이 모두 가능하나, 고차의 반복모형에서는 비중심 t -분포를 이용하는 방법을 제외한 나머지 두 방법을 이용할 수 있다.

3.2.1. 2×2 교차설계법의 신뢰구간 신뢰구간을 계산하려면 통계 모형이 필요하다. 이번 절에서 가정하는 2×2 교차설계법 모형은 식 (2.2)에서 정의하였다. 이 모형에서는 $\sigma_{WT}^2 = \sigma_{WR}^2 = \sigma_e^2$ 을 가정하고 있기 때문에 피험자내 분산 σ_{WR}^2 의 추정량은 분산분석의 잔차제곱평균 $\widehat{\sigma_e^2}$ 이 됨은 앞서 밝힌 바 있다. 편의상 σ_e 의 표현 대신 σ_{WR} 을 사용하기로 한다. 모형 (2.2)에서 각 추정량의 분포를 구하면

$$\begin{aligned} \bar{y}_T - \bar{y}_R &\sim N\left(\mu_T - \mu_R, \frac{2}{N}\sigma_{WR}^2\right), \\ (N-2)\frac{\widehat{\sigma_{WR}^2}}{\sigma_{WR}^2} &\sim \chi^2(N-2) \end{aligned} \tag{3.4}$$

이 되고 두 분포는 서로 독립이다.

비중심 t -분포 이용법

식 (3.4)에 주어진 추정량의 분포로부터 $\theta = (\mu_T - \mu_R)/\sigma_{WR}$ 의 신뢰구간 계산에 필요한 비중심 t -분포

$$\sqrt{(N-2)}\hat{\theta} = \sqrt{(N-2)}\frac{(\bar{y}_T - \bar{y}_R)}{\widehat{\sigma_{WR}}} \sim t'\left(\sqrt{\frac{N}{2}}\theta, N-2\right) \tag{3.5}$$

를 얻는다. 여기서 $t'(\sqrt{N/2}\theta, N-2)$ 은 자유도 $N-2$, 비중심 모수 $\sqrt{(N/2)}\theta$ 인 비중심 t -분포를 의미한다.

θ 의 $100(1-2\alpha)\%$ 신뢰구간은 Johnson과 Kotz (1970, p.203)에 설명된 것처럼 θ 에 대한 방정식

$$t'\left(\sqrt{\frac{N}{2}}\theta, N-2; \alpha\right) = \sqrt{\frac{N}{2}}\hat{\theta} \tag{3.6}$$

을 θ 에 대해 풀면 얻어진다. 여기서 $t'(\sqrt{N/2}\theta, N-2; \alpha)$ 는 비중심 t -분포의 $100\alpha\%$ $t'(\sqrt{N/2}\theta, N-2; \alpha)$ 백분위수를 나타낸다. 식 (3.6)의 해를 $t'(\sqrt{N/2}\hat{\theta}, N-2; \alpha)$ 으로 표현한다.

Tothfalusi와 Endrenyi (2003)는 $\sqrt{N/2}\theta$ 의 90% 신뢰구간

$$\left[t' \left(\sqrt{\frac{N}{2}} \hat{\theta}, N-2; 0.05 \right), t' \left(\sqrt{\frac{N}{2}} \hat{\theta}, N-2; 0.95 \right) \right].$$

이 SABE의 허용한계 $\pm\sqrt{N/2}\theta_S$ 사이에 속하면 생동성을 인정할 것을 제안하였다. Wellek (2003)은 방정식 (3.6)의 우변을 허용한계 $\pm\sqrt{N/2}\theta_S$ 로 바꾸어 풀면 얻어지는 백분위수 중에서 작은 쪽 구간의 상한과 큰 쪽 구간의 하한으로부터 생성되는 구간

$$\left(t' \left(\sqrt{\frac{N}{2}} \theta_S, N-2; 0.95 \right), t' \left(\sqrt{\frac{N}{2}} \theta_S, N-2; 0.05 \right) \right)$$

에 추정치 $\sqrt{N/2}\hat{\theta}$ 이 포함되면 생동성을 인정하는 방법을 제안하였다. Tothfalusi 등 (2009)은 이 두 방법이 완전히 같지는 않다는 점을 지적하였다.

선형화 근사 신뢰구간

SABE의 식 (2.15)를 제공한 식 (3.1)을 이용하면 선형화 된 식

$$\lambda = (\mu_T - \mu_R)^2 - \theta_S^2 \sigma_{WR}^2 \leq 0 \quad (3.7)$$

을 얻는다. λ 의 95% 신뢰상한이 0보다 작으면 생동성을 인정한다. 좌변의 λ 는 두 개의 모수가 선형으로 결합된 형태이고 이처럼 여러 개의 모수가 선형으로 결합된 경우 신뢰구간을 구하는 방법으로 Hyslop 등 (2000)의 선형화 근사 신뢰구간이 유용하다. 많은 연구에서 이 방법을 이용했는데, Chow 등 (2002), Tothfalusi와 Endrenyi (2003), 정규진 등 (2010)에서 예를 찾아볼 수 있다.

Chow 등 (2002)은 식 (3.7)보다 복잡하고 일반적인 경우에 적용 가능한 표현을 제공한다. 관심 모수가 $\lambda = (\lambda_1 + \dots + \lambda_k) - (\lambda_{k+1} + \dots + \lambda_m)$, $\lambda_i > 0$ 로 표현되고 서로 독립인 λ_i 의 추정량 $\hat{\lambda}_i$ 을 구할 수 있다면 λ 의 95% 근사 신뢰상한은

$$\lambda = \left(\hat{\lambda}_1 + \dots + \hat{\lambda}_k \right) - \left(\hat{\lambda}_{k+1} + \dots + \hat{\lambda}_m \right) + \left[\left(u_1 - \hat{\lambda}_1 \right)^2 + \dots + \left(\hat{\lambda}_m - l_m \right)^2 \right]^{\frac{1}{2}} \quad (3.8)$$

로 주어진다. u_i 는 λ_i ($i = 1, \dots, k$)의 95% 신뢰상한을, l_j 는 λ_j ($j = 1, \dots, m$)의 95% 신뢰하한을 표시한다.

따라서 식 (3.8)에 대한 신뢰상한은

$$\begin{aligned} & (\bar{y}_T - \bar{y}_R)^2 - \theta_S^2 \widehat{\sigma}_{WR}^2 + \sqrt{d_1^2 + d_2^2}, \\ d_1 &= \left(|\bar{y}_T - \bar{y}_R| + t(N-2; 0.05) \sqrt{\frac{N}{2}} \widehat{\sigma}_{WR} \right)^2 - (\bar{y}_T - \bar{y}_R)^2, \\ d_2 &= \theta_S^2 (N-2) \frac{\widehat{\sigma}_{WR}^2}{\chi^2} (N-2; 0.05) - \theta_S^2 \widehat{\sigma}_{WR}^2 \end{aligned} \quad (3.9)$$

로 계산된다.

Boddy 등의 방법

식 (2.3)에 주어진 ABE 신뢰구간이 식 (2.14)의 허용한계 내에 속하면 생동성을 인정한다. 이 때 σ_{WR} 은 추정량 $\widehat{\sigma}_{WR}$ 을 사용한다.

3.2.2. 고차 반복모형의 신뢰구간 고차 모형으로 모형 (2.6)을 고려한다. 이 모형은 2×3 와 2×4 교차설계법 뿐만 아니라 순서가 3이상인 경우로도 확장된다. 모형 (2.6)은 반복 설계법이기에 때문에 피험자내 분산 σ_{WR}^2 과 σ_{WT}^2 이 같다고 가정할 필요가 없으며 분리 추정이 가능하다. scaling에 이용되는 σ_{WR} 이 중요한 모수인데 이를 분리 추정할 수 있는 모형이 고차 모형으로서 자격이 있다 하겠다. 이런 관점에서 T 와 R 의 횟수가 동일한 전통적인 2×3 , 2×4 교차설계법 외에 대조제제를 각 순서에서 2회씩 반복 시험하고 시험제제는 1회만 시험하는 2×3 나 3×3 교차설계법을 생각할 수 있다. Chow 등 (2002)은 이런 형태의 모형을 잉여대조(extra reference) 교차설계법이라 명하였고, Haidar 등 (2008a)에는 대조-반복(reference-replicated) 교차설계법이라 표현되어있다.

2×2 모형에 적용된 방법 중에서 비중심 t -분포를 이용하는 방법은 일반적으로 고차모형에 적용할 수 없다. 식 (2.9)의 계산에서 보여 지는 것처럼 $\bar{y}_T - \bar{y}_R$ 의 분산에 σ_{WR}^2 , σ_{WT}^2 , σ_D^2 이 포함되어 식 (3.4)와는 전혀 다르기 때문이다. 고차모형에 비중심 t -분포를 이용하는 방법을 적용할 수 있는 경우는 $\sigma_D^2 = 0$ 이고 $\sigma_{WT}^2 = \sigma_{WR}^2$ 일 때이다. sxp 교차설계법에서 이 조건들이 만족된다면 앞 절에 기술된 내용 중에서 자유도는 $N - 2$ 를 $N - s$ 로, $\sqrt{N/2}$ 는 $\sqrt{N/p}$ 로 바꾸어 사용하면 된다.

고차 모형에서는 2.3절의 IBE 부분에서 논의된 것처럼 추정량의 분포가 복잡하기 때문에 선형화 작업이 신뢰구간 계산에 매우 효율적인 수단이다. 따라서 Hyslop 등 (2000)의 선형화 근사 신뢰구간은 고차 모형에 매우 유용한 방법이 된다. 한 가지 유의할 점은 식 (3.8)의 신뢰상한을 계산하려면 추정량의 분포가 독립이어야 한다. Chow 등 (2002)은 잉여대조 설계법을 포함하여 2×3 교차설계 모형에서, Hyslop 등 (2000)은 2×4 교차설계법에서 $\bar{y}_T - \bar{y}_R$ 와 $\hat{\sigma}_{WR}^2$ 이 독립임을 보였다.

Boddy 등의 방법은 고차 모형에 그대로 적용 가능하다. 다만 이 방법은 엄밀한 의미로 SABE의 신뢰구간과는 차이가 있다. 두 방법에 대한 세밀한 비교가 필요해 보인다.

3.3. SABE에 대한 US FDA의 제안과 EMEA의 규정

US FDA의 제안

US FDA는 HVD 문제를 해결하기 위하여 많은 연구를 수행해 왔다. 2004년에는 그동안 연구된 여러 가지 해결방법들을 비교하여 그 결과를 ACPS(Advisory committee for Pharmaceutical Science) 회의에서 논의 하였고, 후속 작업으로 대규모 모의실험을 실시하여 그 결과를 2006년 ACPS회의에 보고하였다. 이렇게 얻어진 US FDA의 제안은 2007년 AAPS/FDA workshop을 거쳐 Haidar 등 (2008a)와 Haidar 등 (2008b) 두 편의 논문으로 출판되었다.

US FDA의 제안에는 $CV_{WR} \geq 30\%$ 인 경우를 HVD로 정의하고, 이 때 생동성 시험에 필요한 최소 피험자수는 24명으로 명시되어 있다. 모형으로는 각 순서에서 R 이 두 번씩 반복되는 (TRR, RTR, RRT)의 3×3 교차설계법을 선호하고 있다. 4기간 모형은 연구 기간이 길어 피험자들의 중도 탈락률이 높고, 3×3 모형이 효율성도 높기 때문에 3×3 교차설계법의 사용을 권장하고 있다. 그렇지만 4기간 모형도 여전히 사용할 수 있다.

여러 가지 방법 중 식 (3.7)과 (3.8)으로 표현된 SABE 방법과 Hyslop 등 (2000)의 선형화 근사 신뢰구간이 추천되었다. 신뢰도는 95%, 규정상수로는 $\sigma_{W0} = 0.250$ ($\theta_S = 0.893$)이 제안되었다. EMEA와는 다르게 US FDA의 제안에 따르면 AUC와 C_{max} 가 모두 허용구간을 만족해야 한다. 허용구간 80~125%의 GMR constraint도 함께 사용된다.

EMEA 규정

EMEA (2010)는 HVD 문제를 HVD 상품 (HVD products = HVDP)을 기준으로 기술하고 있다.

$CV_{WR} > 30\%$ 인 경우 HVDP로 정의하고 있으며, 어떤 상품이 HVDP로 의심되는 경우 3기간이나 4기간의 반복 교차설계법을 사용할 수 있다고 규정하고 있다.

HVDP에 대한 EMEA 규정의 핵심은 어떤 상품이 비록 HVDP라 하더라도 AUC에 대한 생동성 평가는 ABE로 수행하고 오직 C_{max} 의 GMR에 대한 허용구간만 최대 69.84%~143.19%까지 확장을 허용하는 점이다. 허용구간의 확장 방법으로는 Boddy 등 (1995)의 방법을 이용하되 GMR의 허용구간으로

$$\exp[\pm k\sigma_{WR}] \quad (3.10)$$

을 사용한다. 여기서 규정상수 k 는 SABE의 규정상수 θ_S 와 동일한 것으로 $k = 0.760$ ($\sigma_{W0} = 0.294$)을 사용하도록 규정되어 있다. 최대 허용구간은 $CV_{WR} = 50\%$ 에 대응되는 구간으로 더 이상의 확장은 허용하지 않는다. 허용구간 80~125%의 GMR constraint도 함께 사용된다.

확장된 허용한계 (3.10)을 사용하려면 이 사실을 미리 시험계획서에 명시하여야 하며, C_{max} 의 허용구간 확장이 의학적 문제로 문제가 없음을 입증하여야 한다. 또, 반복 설계법을 이용하여 C_{max} 의 변동이 $CV_{WR} > 30\%$ 임을 보여야 하고, 이 값이 신뢰할만한 추정치로 이상치의 영향을 받지 않았음을 증명하여야 한다.

US FDA 제안과 EMEA 규정에는 같은 점도 많지만 서로 상이한 내용도 적지 않음을 알 수 있다. 이 중에는 지금까지 살펴본 것도 있지만, 다음 장에서 그동안 연구된 모의시험의 결과들을 중심으로 하여 규정상수 값의 결정이나 GMR constraint의 기능, 이상치 문제 등에 대하여 살펴보고자 한다.

4. SABE의 평가

SABE의 성능에 영향을 미치는 요인은 매우 다양하다. 여기서는 소비자 위험과 생산자 위험의 비교 방법, 생동성 시험의 모형, mixed와 unmixed strategy, HVD를 가르는 σ_{WR} 또는 CV_{WR} 의 기준 값(switching variation σ_S or CV_S), 규정상수 σ_{W0} 의 값, GMR constraint, 이상치, σ_{WR} 와 σ_{WT} 의 값 등을 고려한다.

4.1. 소비자 위험과 생산자 위험

ABE, IBE, SABE 등 생동성 평가 방법은 일종의 검정법이기 때문에 이들의 성능을 비교하려면 통계적 가설 검정 구조에서 제 1종 오류와 제 2종 오류 (또는 검정력)를 비교하는 방법이 일반적이다. 2장의 식 (2.1)의 설명에서 언급한대로 제 1종 오류는 소비자 위험, 제 2종 오류는 생산자 위험이라 부른다. 검정력은 $1 -$ 생산자 위험이므로 검정력이 큰 평가 방법은 생산자 위험이 작은 방법이란 의미가 된다.

GMR의 허용한계가 80~125%로 주어지는 ABE에서는 대칭성 때문에 소비자 위험은 $GMR = 1.25$ 일 때 계산하고, 생산자위험이나 검정력은 GMR 값이 1.00에서 1.25 사이에 속하는 구간에서 비교하면 된다. 그러나 SABE는 식 (2.14)나 (2.15) 모두 허용한계가 σ_{WR} 에 비례하여 커지는 구조를 갖고 있어서 소비자 위험과 생산자 위험을 계산하는 GMR 값이 σ_{WR} 에 의존하게 된다. 이런 이유로 SABE의 검정력 비교는 주로 $GMR = 1.00$ 에서 이루어지지만 한 점에서의 비교를 전체 검정력의 비교로 보기 어렵고 소비자 위험의 비교는 기준을 정하는 일조차 매우 어렵게 된다.

본 논문에서는 ABE를 생동성 평가의 기본으로 보고 허용한계가 80~125%를 벗어나면 소비자 위험이 증가하는 것으로 간주하여, 소비자 위험은 $GMR = 1.25$, 검정력은 GMR 값이 1.00에서 1.25 사이 일 때를 위주로 비교하려고 한다. 물론 이 영역에서 SABE는 σ_{WR} 이 클수록 허용한계가 커지기 때문에 소비자 위험과 검정력이 큰 값을 갖게 되고, 이를 감안하여 비교에 임해야한다는 점을 밝혀둔다.

4.2. 모형의 비교

SABE는 IBE에서처럼 σ_{WR} 로 scaling을 하기 때문에 이에 대한 신뢰성 있는 추정량이 필수적이다. 따라서 생동성 평가를 책임지는 국가 기관으로써는 2×2 모형 보다는 σ_{WR} 을 분리해서 추정할 수 있는 고차의 모형을 선호하게 된다. EMEA는 3이나 4 기간 모형의 사용을 규정하고 있고, Haidar 등 (2008a)에 기술된 US FDA의 제안에는 3×3 잉여대조 교차설계 모형을 적극 권하고 있다.

실제로 US FDA는 이 모형을 이용하여 SABE의 특성을 평가하는 모의실험을 행하여 그 결과를 Haidar 등 (2008b)에 발표하였다. 이들의 시험에는 각 기간에 12명이 표본으로 배치되어 전체 표본크기는 $N = 36$ 명이었다. US FDA에서 제안된 사항들은 Endrenyi와 Tothfalusi (2009)에 의해 모의실험을 통하여 조목조목 평가되었는데 이들 역시 US FDA에서 권장한 표본크기 36인 3×3 잉여대조 모형을 사용하였다. 3×3 잉여대조 모형과 비교할만한 것으로 전통적인(TRT, RTR)의 2×3 교차설계법을 이용한 Tothfalusi 등 (2009)의 표본크기 36인 모의실험을 들 수 있다.

이런 모의실험 결과들은 비록 실험 조건이 똑같지는 않지만 같거나 비슷한 조건을 중심으로 비교가 가능해 보인다. Haidar 등 (2008b)의 그림 1~4, Endrenyi와 Tothfalusi (2009)의 그림 2, Tothfalusi 등 (2009)의 그림 3을 비교해 보면 3×3 잉여대조 모형과 전통적인 2×3 모형의 검정력이 거의 같은 것을 알 수 있다. 특히 $GMR = 1.00$ 일 때의 검정력을 보면 모형이나 σ_{WR} 의 값에 관계없이 $\sigma_{W0} = 0.294$ 일 경우 검정력은 약 90%, $\sigma_{W0} = 0.250$ 일 때 검정력은 약 95% 정도의 값을 갖는 것으로 보인다. 이런 현상은 Tothfalusi 등 (2009)이 지적한대로 $GMR = 1.00$ 일 때 SABE (Boddy 등 (1995)의 방법)의 검정력이 σ_{WR} 의 값과는 무관하고 표본크기에만 의존하기 때문으로 이해된다.

그러나 Chow 등 (2002)은 IBE에 2×3 잉여대조 교차설계법 모형을 적용하여 전통적인 2×3 모형보다 높은 검정력이 얻어짐을 보였다. 이들은 이론적으로 σ_{WR} 의 추정량이 전통적인 모형에 비해 잉여대조 모형에서 2배 많은 자유도를 갖기 때문에 높은 검정력을 갖는다고 설명하였다. 같은 이유로 SABE에서도 잉여대조 모형의 검정력이 높을 것으로 기대하였으나 앞 절에서 살펴본 모의실험 결과는 그렇지 못했다. 원인의 하나로 모의실험의 조건을 들 수 있다. Chow 등 (2002)은 $(\sigma_D, \sigma_{WR}, \sigma_{WT})$ 의 값을 매우 다양하게 변화시켜가면서 모의실험을 수행한 반면에 SABE에 대한 모의실험에서는 $\sigma_D = 0$ 이고 $\sigma_{WR} = \sigma_{WT}$ 의 가정 하에서 수행되었다. SABE에 적용된 조건에서는 검정력이 σ_{WR} 와 무관하게 되어 여러 모형을 비교할 수 없을 것으로 생각된다. 일반적인 조건에서는 2×3 나 3×3 잉여대조 모형이 검정력이 높을 것으로 예상 되지만 이를 확인할 보다 정밀한 모의실험이 필요해 보인다.

4.3. switching variation(σ_S or CV_S)과 mixed or unmixed strategy

SABE에는 mixed strategy가 함께 사용되는데 mixed SABE에서 ABE와 SABE의 경계를 가르는 'switching variation'은 HVD와 그렇지 않은 약을 구별하는 피험자내 변동의 기준값과 동일하고 이는 보통 $\sigma_S = 0.294$ (or $CV_S = 30\%$)가 널리 받아들여지고 있다는 점은 전술한 바와 같다. Endrenyi와 Tothfalusi (2009)는 $CV_S = 30\%$, $CV_{W0} = 25\%$ 일 때와 $CV_S = CV_{W0} = 25\%$ 일 때 ABE와 SABE 등의 검정력을 비교하여 두 경우 거의 동일한 그림을 얻었고, 이는 switching variation이 이런 검정 방법들 사이의 검정력 차이에 영향을 주지 않는다는 점을 나타낸다고 해석하였다. 따라서 $CV_S = 30\%$ 는 적절한 값이라 여겨진다.

이제 mixed SABE의 대안이 될 수 있는 ABE와 unmixed SABE의 성능을 비교한다. 이들의 성능을 나타내는 소비자 위험과 검정력은 이들이 만들어 내는 허용한계의 크기에 절대적인 영향을 받는다. 허용한계의 크기를 직관적으로 생각해 보면 $\sigma_{WR} < \sigma_S$ 인 영역의 경우 ABE와 mixed SABE는 같고 unmixed SABE는 이들보다 작고, $\sigma_{WR} \geq \sigma_S$ 영역에서는 mixed와 unmixed SABE가 같고 ABE는

더 작은 허용한계를 갖는다. 물론 식 (3.3)에서 언급한대로 허용한계는 규정상수 σ_{w0} 의 영향도 받아서 이 값이 작을수록 커진다. 이들의 허용한계와 상대적 크기, 상호 관계 등을 잘 관찰할 수 있는 그림이 Endrenyi와 Tothfalusi (2009)의 그림 1에 주어져있다.

Tothfalusi와 Endrenyi (2003)의 그림 3은 $\sigma_{w0} = 0.20$, $CV_S = 10, 20, 30, 40\%$ 일 때 세 방법의 성능을 보여주고 있다. switching variation과 규정상수가 비슷한 경우 ($\sigma_{w0} = 0.20$, $CV_S = 20\%$) 세 방법의 검정력이 비슷하고, CV_S 가 큰 경우에는 두 가지 SABE의 검정력은 비슷한데 반해 ABE의 검정력은 현저히 떨어지고, CV_S 가 작은 경우에는 ABE와 mixed SABE의 검정력이 비슷하고 unmixed SABE의 검정력이 상당히 낮아짐을 알 수 있다. 어떤 경우이든 mixed SABE가 좋은 검정력을 보여준다. Endrenyi와 Tothfalusi (2009)의 표 1은 $CV_S = CV_{w0} = 30\%$ 일 때 소비자 위험을 보여준다. 이 예에서는 ABE가 5%의 소비자 위험을 잘 맞추는데 비해 mixed SABE는 5.56%로 약간 크지만 허용할 만한 크기이고 unmixed SABE는 6.98%로 좀 더 큰 값을 갖는다. 이상을 종합하면 mixed SABE가 좋은 방법으로 판단된다.

4.4. 규정상수 σ_{w0} 의 값

규정상수 σ_{w0} 의 값으로 US FDA는 0.250을 권장하고 EMEA는 0.294의 사용을 규정하고 있다. Tothfalusi와 Endrenyi (2003)은 이 값에 IBE 규정상수 값인 0.200과 Boddy 등 (1995)의 권장 값인 0.223을 포함하여 검정력을 비교하였다. 보다 규모가 큰 모의실험은 Haidar 등 (2008b)나 Endrenyi와 Tothfalusi (2009)에서 찾아볼 수 있다.

FDA의 제안 내용의 근거가 되는 Haidar 등 (2008b)은 σ_{w0} 가 0.20, 0.250, 0.294일 때 ABE와 SABE를 포함하는 여러 가지 검정 방법을 비교하였다. 이들은 IBE 상수 0.200은 소비자 위험이 너무 커지고, 0.294는 $CV = 30\%$ 에서 ABE보다 SABE가 오히려 power가 떨어지는 경우까지 생겨서 바람직하지 않은 것으로 판단하여 0.250이 규정상수로 적당한 값이라는 결론을 내리고 있다.

Endrenyi와 Tothfalusi (2009)는 σ_{w0} 가 0.250, 0.294일 때를 비교하였다. 이들이 주목한 것은 switching variation과 규정상수의 값이 다르면 $CV_{WR} = 30\%$ 에서 SABE의 허용한계에 점프가 일어난다는 점이다. 이 때문에 소비자 위험이 커질 뿐만 아니라 CV_{WR} 이 불연속점에 매우 가까운 값일 때 이 값이 불연속점보다 단지 큰 영역에 속하는지 작은 영역에 속하는지에 따라 거의 같은 값에 대한 생동성 평가의 결과가 확연히 갈라지는 문제를 야기하게 된다. σ_{w0} 가 0.250일 때 $GMR = 1.25$ 에서 소비자 위험을 보면 ABE를 제외한 SABE의 여러 가지 방법에서 모두 상당히 큰 것을 볼 수 있다.

σ_{w0} 가 0.294일 때도 시험제제와 대조제제 모두의 CV 가 30%일 때만 $GMR = 1.25$ 에서 SABE의 소비자 위험이 5%를 만족하고 CV 가 30%보다 커지면 소비자 위험은 더 커질 것이다. 이는 불연속점의 영향도 있겠으나 SABE 허용한계가 σ_{WR} 에 비례해서 무한정 커질 수 있는 구조적인 문제에 기인하는 바가 더 큰 것으로 판단된다. 이와 같은 문제의 해결책의 일환으로 GMR constraint가 함께 사용된다.

Endrenyi와 Tothfalusi (2009)는 σ_{w0} 가 0.250, 0.294일 때 SABE에 대한 GMR constraint의 영향을 비교하였다. 비교 결과 σ_{w0} 가 0.250일 때는 CV_{WR} 이 큰 경우($CV_{WR} = 60\%$)에는 생동성 평가가 주로 GMR constraint의 결과에 의존하는데 반해 CV_{WR} 이 작은 경우($CV_{WR} = 35\%$)에는 주로 SABE의 결과에 따른다는 사실이 밝혀졌다. σ_{w0} 가 0.294일 때는 CV_{WR} 의 값에 상관없이 생동성 평가가 GMR constraint보다는 SABE의 결과에 따르는 것으로 나타났다. 결국 σ_{w0} 의 값으로는 0.250보다 0.294가 적당한 것으로 생각된다.

4.5. GMR constraint

SABE에서 사용하는 σ_{WR} 을 이용한 scaling은 두 평균의 차 $\mu_T - \mu_R$ 의 크기를 축소시키는 역할을 한다. 따라서 HVD 문제에 SABE를 적용하는 경우 Midha (2006)나 Tothfalusi 등 (2009)의 언급처럼 GMR 값이 80~125% 범위를 벗어나더라도 생동성 기준을 충족시키는 경우가 생길 수 있다. PBE나 IBE에도 같은 문제가 발생하는데 이를 해결하기 위하여 Benet (1999, 2006)은 GMR constraint의 사용을 제안하였다. 앞서 살펴본 대로 US FDA는 PBE나 IBE의 허용기준과 더불어 GMR의 점추정치가 80~125% 이내에 속할 것을 요구하고 있으며, Health Canada는 σ_{WR} 의 크기에 상관없이 C_{max} 에 대한 생동성 평가는 허용한계가 80~125%인 GMR constraint만 만족하면 되는 것으로 규정하고 있다. 이 규정은 SABE에도 그대로 적용될 수 있다. EMEA 또한 HVD에 관한 규정에 80~125%의 허용한계를 사용하도록 명시하고 있다.

GMR constraint의 역할에 대해 Benet (2006)은 통계적인 배경이나 이유보다는 다만 사람들에게 허가된 복제약은 오리지널 약과 같은 효용을 갖는다는 믿음을 주기위한 ‘정치적’ 성격이 강한 규정이라 설명하였다.

GMR 추정치의 허용범위로는 보통 $\pm 20\%$ 가 적용되는데, 이보다 좁은 범위가 제안되기도 한다. 한 예로, Benet (2006)은 $\sigma_{WR} \leq 20\%$ 인 경우 AUC에 대해서는 $\pm 20\%$, C_{max} 는 $\pm 15\%$ 를, 그리고, $\sigma_{WR} > 20\%$ 인 경우에는 AUC는 $\pm 15\%$, C_{max} 는 $\pm 20\%$ 를 제안하였다.

Endrenyi와 Tothfalusi (2009)는 모의실험을 통해 GMR 상한이 125%인 경우와 115%인 경우를 비교하였다. GMR이 115%인 경우에는 125%인 때와는 다르게 생동성 평가가 SABE의 결과 보다는 GMR constraint의 평가 결과에 주로 의존하게 되는 바람직하지 못한 현상이 나타났다. 따라서 이들은 GMR constraint로 80~125%의 범위가 적당하며 이 보다 작은 구간은 너무 엄격한 규정이라는 점을 지적하였다.

SABE 허용한계가 σ_{WR} 에 비례해서 무한정 커지는 것을 방지하기 위해서 허용한계의 최대값을 제한하는 방법을 생각할 수 있다. 실제 EMEA의 규정에는 GMR constraint와 함께 CV_{WR} 이 50% 이상이라도 허용한계에 제한을 두어 최대 69.84%~143.19%까지만 확장을 허용한다. 또, Karalis 등 (2004), Karalis 등 (2005)와 Kytariolos 등 (2006)은 σ_{WR} 에 비례해서 커지는 SABE의 허용한계에 최대허용한계와 GMR constraint를 포함시켜 leveling off 성질을 갖는 새로운 방법을 제안하였다. 이 방법은 소비자 위험이나 생산자 위험을 어느 정도 줄일 수 있다는 측면에서 장점을 갖으나 ABE에 비해 여전히 어느 정도 소비자 위험의 증가는 피할 수 없어 보이고, 허용한계를 결정하기 위하여 선택해야 하는 함수의 모양이 복잡하고 인위적인 느낌을 받게 한다.

GMR constraint나 최대허용한계의 설정 등 SABE와 함께 사용될 부차적인 생동성 기준은 SABE의 단점을 보완하기 위하여 고안되었다. GMR에 대한 허용범위로는 80~125%가 적절해 보이나 최대허용한계의 설정 여부나 그 값은 얼마로 하는 것이 바람직한지에 대해서는 좀 더 많은 연구가 필요해 보인다.

4.6. 이상치 문제

생동성 시험은 소수의 피험자를 대상으로 이루어지기 때문에 이상치가 존재하면 시험의 결과에 큰 영향을 미칠 수 있다. 잘 알려진 대로 이상치는 관측치의 평균과 표준편차에 영향을 주어 두 평균의 차이에 대한 추정치와 표준편차의 추정치를 커지게 한다. 이와 같은 이상치의 역할 때문에 ABE에서는 이상치가 존재하면 생동성 검정을 통과하기가 어려워 지고 생산자는 이상치가 생기지 않도록 주위를 기울이고 이상치의 원인 파악에 노력하게 된다. 그러나 Tothfalusi 등 (2009)의 언급처럼 SABE에서는 좀 다른 영향을 미치게 된다. 우선 이상치 때문에 $\widehat{\sigma_{WR}}$ 이 커지면 이 생동성 시험이 HVD로 잘못 분류될 가능성

이 높아진다. 만약 HVD가 아닌 시험이 HVD로 분류되면 ABE대신 SABE에 의해 생동성이 평가되고 이는 생산자에게 유리하게 작용하게 된다. 따라서 생산자는 이상치가 존재하기를 원하게 되고 이는 생동성 시험에 나쁜 영향을 미칠 수 있기 때문에 관리 당국에서는 이상치 유무를 엄밀히 감독하여야 한다. EMEA의 규정에는 피험자내 표준편차의 추정치 $\widehat{\sigma}_{WR}$ 이 믿을만하다는 것과 추정치가 이상치의 영향을 받지 않았다는 점을 생산자가 입증하도록 명시되어 있다. SABE에서 이상치가 생산자에 유리하게 작용하는 것만은 물론 아니다. ABE에서처럼 이상치는 두 평균의 차를 크게 하여 생동성 검정에서 GMR constraint의 기준을 통과하기 어렵게 만든다. 결국 생산자나 관리 당국 모두 이상치에 주의하여야 한다.

이상치 문제를 피하기 위하여 Tothfalusi 등 (2009)은 이상치에 대해 robust한 통계적 방법을 사용할 것을 제안하고는 있지만 생동성 시험 규정에 담을만한 연구를 찾아보기는 어렵다. 다만 생동성 시험에서 이상치 검출을 위한 통계적 방법들이 Chow와 Tse (1990), Liu와 Weng (1991), Wang과 Chow (2003), Ramsay와 Elkum (2005), 정규진 등 (2009) 등에 의하여 연구되었으나 이 분야도 아직 규정에 적용할 수 있는 결과를 얻지는 못한 것 같다.

4.7. σ_{WT} 의 영향

SABE는 기본적으로 평균의 차에만 관심이 있는 ABE를 기반으로 하고 있어 IBE에 포함되어있는 분산의 차이 $\sigma_{WT}^2 - \sigma_{WR}^2$ 에는 관심이 없다. 다만 시험제제와 대조제제의 변동 중 관심 모수의 분모에 scaling 인자로 σ_{WR} 만이 포함되어 있을 뿐이다. 생동성 검정이 시험제제보다는 대조제제를 기준으로 설계되는 한 어쩔 수 없는 일로 보인다. 비록 시험제제의 피험자내 표준편차 σ_{WT} 가 직접 SABE의 식에 포함되어 있지는 못하더라도 생동성 평가에 영향을 주리라는 것은 자명하다. 그러나 σ_{WT} 의 영향력에 대한 연구는 거의 없고 앞서 언급한 모의실험도 $CV_{WT} = CV_{WR}$ 인 경우만 다루고 있다. SABE의 정의에 의해 σ_{WR} 이 커지면 검정력이 높아진다. 반면에 Haidar 등 (2008b)의 그림 6을 보면 CV_{WT} 이 커질수록 검정력은 급격히 떨어진다. σ_{WR} 로 scaling을 하기 때문에 어쩔 수 없는 현상이긴 하지만 무언가 불편한 느낌이 드는 것도 사실이다. 생동성 평가 전반에서 σ_{WT} 의 역할에 대한 보다 깊이 있는 연구가 필요해 보인다.

5. 결론

HVD문제를 해결하기 위해서 제안된 SABE는 σ_{WR} 에 비례해서 생동성 허용한계를 확장한다. 이는 전통적 생동성 평가 방법인 ABE에 비해 생산자 위험은 감소시키지만 동시에 소비자 위험을 증가시킨다. 이를 바탕으로 SABE에 대한 평가를 여러 측면에서 살펴보았다. 이를 요약하면 다음과 같다:

- 1) SABE의 통계학적 모형으로 2×2나 교차의 교차설계법이 모두 가능하지만 σ_{WR} 의 분리 추정에 유리한 2×3나 3×3 잉여대조 교차설계법이 선호된다.
- 2) SABE에서 신뢰구간 추정방법으로는 비중심 t -분포를 이용하는 방법, Howe의 방법, Boddy 등 (1995)의 방법 중에서 Howe의 선형화 근사신뢰구간 계산법이 보다 유용한 것으로 평가된다.
- 3) SABE의 허용한계에 영향을 주는 요인 중 하나인 switching variation은 $\sigma_S = 0.294$ (or $CV = 30\%$)가 널리 받아들여지고 있으며 이 값을 경계로 $\widehat{\sigma}_{WR} < \sigma_S$ 면 ABE를, $\widehat{\sigma}_{WR} \geq \sigma_S$ 면 SABE를 사용하는 mixed strategy가 적용된다.
- 4) 규정상수 σ_{W0} 의 값으로 0.250과 0.294 두 개가 비교되는데 0.250일 때 생기는 생동성 허용구간의 불연속성이 생동성 평가에 여러 가지 문제를 야기하기 때문에 0.294가 적절한 것으로 평가된다.

- 5) SABE를 보완하기 위하여 GMR constraint의 사용이 권장된다. 이는 통계적 이유 보다 정치적인 성격을 띤 기준이긴 하지만 SABE의 허용한계가 너무 커지는 것을 견제하는 역할도 한다. GMR 허용범위로 보통 80~125%가 이용된다.
- 6) 이상치는 허용한계의 설정에 영향을 줄뿐만 아니라 생동성 시험을 HVD로 분류할 가능성을 높이기 때문에 생동성 평가의 전 과정에서 이상치의 존재 여부에 주의하여야하며 관리 규정에 어떤 형태로든 이에 대한 내용이 포함되어야 할 것으로 여겨진다.

마지막으로 SABE에 대한 향후 연구주제 중 몇 가지를 정리해 본다.

- 1) 모수의 다양한 조건에서 여러 모형의 비교 연구가 필요해 보인다. 비교 연구에 포함될 SABE의 모형으로는 2×2 , 2×3 , 3×3 , 2×4 , 2×3 잉여대조, 3×3 잉여대조 교차설계 모형 등을 생각할 수 있다. 잉여대조 모형과 보통 모형의 비교가 이에 포함되어야 한다.
- 2) GMR constraint는 SABE의 생동성 허용한계와 밀접한 관계가 있는 것으로 보인다. 이 주제는 최대 생동성 허용한계의 설정과도 관련 될 수 있다.
- 3) SABE에서 이상치의 영향력이 어느 정도인지, 또 이상치 판정과 처리는 어떤 방법으로 해야 하는지에 대한 종합적인 연구가 필요해 보인다.
- 4) SABE의 검정력은 σ_{WR} 뿐만 아니라 σ_{WT} 에도 영향을 받기 때문에 둘을 함께 고려하는 방법을 생각할 수도 있어 보인다. 만약 SABE의 scaling 인자로 σ_{WR} 과 σ_{WT} 의 대칭성을 고려한다면 $\sqrt{\sigma_{WT}^2 + \sigma_{WR}^2}$ 이 σ_{WR} 에 대한 자연스런 대안이 될 수 있다.
- 5) 표본크기에 따라 SABE의 성능이 어떻게 변하는지 연구할 필요성이 있어 보인다.

참고문헌

- 박상규, 남봉현, 정윤노, 이재영, 정규진 (2010). 생물학적 동등성 시험의 추가시험에 대한 통계적 고찰, <한국통계학회논문집>, **17**, 107-115.
- 박상규, 임남규, 이재영, 김병천 (2005). 2×2 교차설계법에서 모집단 생물학적 동등성 검정방법 비교, <응용통계연구>, **18**, 1-13.
- 식품의약품안전청 (2010). 생물학적동등성시험기준 (고시 제2010-43호).
- 정규진, 박상규, 김관엽 (2011). 공변량을 고려한 2×2 교차설계법에 평균 생물학적 동등성 평가, <응용통계연구>, **24**, 161-167.
- 정규진, 박상규, 우화형 (2009). 2×2 생물학적 동등성 시험에서 이상치 검출을 위한 통계적 방법, <한국통계학회 논문집>, **16**, 745-751.
- 정규진, 임남규, 박상규 (2010). 3×2 교차설계법에서 개인 생물학적 동등성 검정, <응용통계연구>, **21**, 139-150.
- Arieta, A. G. (2006). Highly variable drugs. Understanding Bioequivalence of Generic Drugs, *A Worldwide Review*, Monte Carlo.
- Benet, L. (1993). Bioavailability and bioequivalence: Definitions and difficulties in acceptance criteria. In Midha KK, Blume HH, editors, *Bio-International 92: bioavailability, bioequivalence and pharmacokinetics*, Medpharm, Stuttgart, 27-35.
- Benet, L. (1999). Individual bioequivalence: An overview, *AAPS International Workshop on Individual Bioequivalence: Realities and Implementation*, Montreal, Quebec.
- Benet, L. (2006). Why highly variable drugs are safer. Meeting of FDA Committee for Pharmaceutical Science, <http://www.fda.gov/ohrms/dockets/ac/06/slides/2006-4241s22.ppt> (assessed 3/17/2011).
- Bio-international (1992). Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies, Bad Homburg, Germany, May 20-22.

- Blume, H. H. (2010). Highly variable drugs: reasons for high variability and solutions to overcome BE problems, Modern Strategies for the Development of Generic Drugs 20th AGAH Annual Meeting, Hamburg, http://www.agah.info/uploads/media/Blume_H_BE-HVD_AGAH_2010_Einleitung1.pdf (assessed 3/17/2011).
- Boddy, A. W., Snikeris, F. C., Kringler, R. O., Wei, G. C. G., Oppermann, J. A. and Midha, K. K. (1995). An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs, *Pharmaceutical Research*, **12**, 1865–1868.
- Chow, S. C. and Liu, J. P. (2008). *Design and Analysis of Bioavailability and Bioequivalence Studies*, 3rd edition, CRC press.
- Chow, S. C., Shao, J. and Wang, H. (2002). Individual bioequivalence testing under 2×3 designs, *Statistics in Medicine*, **21**, 629–648.
- Chow, S. C. and Tse, S. K. (1990). Outliers detection in bioavailability/bioequivalence studies, *Statistics in Medicine*, **9**, 549–558.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Davit, B. M. (2004). Highly variable drugs: FDA case studies. Advisory Committee for Pharmaceutical Sciences. Office of Generic Drugs. US FDA Center for Drug Evaluation and Research, http://www.fda.gov/ohrms/dockets/ac/04/slides/4034s206_davit.ppt (assessed 3/17/2011)
- Davit, B. M., Conner, D. P., Fabian-Fritsch, B., Haidar, S. H., Jiang, X., Patel, D. T., Seo, P. R. H., Suh, K., Thompson, C. L. and Yu, L. X. (2008). Highly variable drugs: Observations from bioequivalence data submitted to the FDA for new generic drug applications, *The AAPS Journal*, **10**, 148–156.
- DiLiberti, C. E. (2007). BE for Highly Variable Drugs an Industry Perspective, Barr Laboratories, Inc. <http://www.aapspharmaceutica.com/meetings/files/90/21DiLiberti.pdf> (assessed 3/17/2011).
- EMA, European Medicines Agency, Committee for Medicinal Products for Human Use. (2008). *Draft Guideline on the Investigation of Bioequivalence*, London.
- EMA, European Medicines Agency, Committee for Medicinal Products for Human Use. (2010). *Draft Guideline on the Investigation of Bioequivalence*, London.
- Endrenyi, L., Fritsch, S. and Van, W. (1991). C_{max}/AUC is a clearer measure than C_{max} for absorption rates in investigations of bioequivalence, *International Journal of Clinical Pharmacology*, **29**, 394–399.
- Endrenyi, L. and Tothfalusi, L. (2009). Regulatory conditions for the determination of bioequivalence of highly variable drugs, *Journal of the Pharmaceutical Sciences*, **12**, 138–149.
- Haidar, S. H., Davit, B., Chen, M. L., Conner, D., Lee, L., Li, Q. H., Lionberger, R., Makhlouf, F., Patel, D., Schuirmann, D. J. and Yu, L. X. (2008a). Bioequivalence approaches for highly variable drugs and drug products, *Pharmaceutical Research*, **25**, 237–241.
- Haidar, S. H., Makhlouf, F., Schuirmann, D. J., Hyslop, T., Davit, B., Conner, D. and Yu, L. X. (2008b). Evaluation of a scaling approach for the bioequivalence of highly variable drugs, *AAPS Journal*, **10**, 450–454.
- Health Canada, Ministry of Health. (1992). *Guidance for industry: Conduct and analysis of bioavailability and bioequivalence studies. Part A: oral dosage formulations used for systemic effects*, Ottawa, Ontario.
- Health Canada. Therapeutic Products Directorate (2003). Discussion paper on 'Bioequivalence requirements - highly variable drugs and highly variable drug products: issues and options'. Expert Advisory Committee on Bioavailability and Bioequivalence (EAC-BB) Meeting.
- Howe, W. G. (1974). Approximate confidence limits on the mean of $X + Y$ where X and Y are two tabled independent random variables, *Journal of the American Statistical Association*, **69**, 789–794.
- Hyslop, T., Hsuan, F. and Holder, D. J. (2000). A small sample confidence interval approach to assess individual bioequivalence, *Statistics in Medicine*, **19**, 2885–2897.
- Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions-2*, Houghton Mifflin Company, Boston.
- Karalis, V., Macheras, P. and Symillides, M. (2005). Geometric mean ratio-dependent scaled bioequivalence limits with leveling-off properties, *European Journal of the Pharmaceutical Sciences*, **26**, 54–61.
- Karalis, V., Symillides, M. and Macheras, P. (2004). Novel scaled average bioequivalence limits based on GMR and variability considerations, *Pharmaceutical Research*, **21**, 1933–1942.

- Kytariolos, J., Karalis, V., Macheras, P. and Symillides, M. (2006). Novel scaled bioequivalence limits with leveling-off properties, *Pharmaceutical Research*, **23**, 2657–2664.
- Liu, J. P. and Weng, C. S. (1991). Detection of outlying data in bioavailability/bioequivalence studies, *Statistics in Medicine*, **10**, 1375–1389.
- Midha, K. K. (2006). Highly Variable Drugs & Drug Products-A Rationale for Solution of a Persistent Problem, <http://www.fda.gov/ohrms/DOCKETS/ac/06/slides/2006-4241s2.3.ppt> (assessed 11/3/2010).
- Patterson, S. D., Zariffa, N. M. D., Montague, T. H. and Howland, K. (2001). Non-traditional study designs to demonstrate average bioequivalence for highly variable drug products, *European Journal of Pharmaceutical Sciences*, **57**, 663–670.
- Ramsay, T. and Elkum, N. (2005). A comparison of four different methods for outlier detection in bioequivalence studies, *Journal of Biopharmaceutical Statistics*, **15**, 43–52.
- Schall, R. and Luus, H. G. (1993). On population and individual bioequivalence, *Statistics in Medicine*, **12**, 1109–1124.
- Scheiner, L. B. (1992). Bioequivalence revisited, *Statistics in Medicine*, **11**, 1777–1788.
- Schuirman, D. J. (1987). A comparison of two one-sided test procedures and the power approach for assessing the equivalence of average bioavailability, *Journal of Pharmacokinetics and Biopharmaceutics*, **15**, 657–680.
- Tanguay, M., Potvin, D., Haddad, J., Lavigne, J., Marier, J. F., DiMarco, M. and Ducharme, M. P. (2002). When will a drug formulation pass or fail bioequivalence criteria? Experience from 1200 studies, *AAPS Pharmaceutical Science*, **4**, Abstract R6193.
- Tothfalusi, L. and Endrenyi, L. (2003). Limits for the scaled average bioequivalence of highly variable drugs and drug products, *Pharmaceutical Research*, **20**, 382–389.
- Tothfalusi, L., Endrenyi, L. and Garcia Arieta, A. (2009). Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence, *Clinical Pharmacokinetics*, **48-11**, 725–743.
- US FDA Center for Drug Evaluation and Research. (2001). *Statistical Approaches to Establishing Bioequivalence: Guidance for Industry*, Rockville, Maryland.
- US Government. (2009). 21 CFR Part320 - Code of Federal Regulations Title 21–Food and Drugs, Part 320–Bioavailability and Bioequivalence Requirements. FDA Home > Medical Devices > Databases에 수록, 2009년 4월 1일 최종수정.
- Wang, W. and Chow, S. C. (2003). Examining outlying subjects and outlying records in bioequivalence trials, *Journal of Biopharmaceutical Statistics*, **13**, 43–56.
- Wellek, S. (2003). *Testing Statistical Hypotheses of Equivalence*, Chapman & Hall, London.

On Evaluation of Bioequivalence for Highly Variable Drugs

Gyu-Jin Jeong¹ · Sang-Gue Park²

¹Department of Business Statistics, Hannam University

²Department of Applied Statistics, Chung-Ang University

(Received September 2011; accepted November 2011)

Abstract

This paper reviews the definition of highly variable drug(HVD), the present regulatory recommendations and the approaches proposed in the literature to deal with the bioequivalence issues of HVD. The concept and the statistical approach of scaled average bioequivalence(SABE) is introduced and discussed with the current regulatory methods. The recommendations for SABE approach are proposed and the further study topics related to HVDs are also presented.

Keywords: Bioequivalence, highly variable drugs, scaled average bioequivalence.

²Corresponding author: Professor, Department of Applied Statistics, Chung-Ang University, 221 Heuksuk-Dong, Dongjak-Gu, Seoul 156-756, Korea. E-mail: spark@cau.ac.kr