

3 × 2 교차설계법에서 개인 생물학적 동등성 검정

정규진¹ · 임남규² · 박상규³

¹한남대학교 비즈니스통계학과, ²질병관리본부 임상연구지원센터, ³중앙대학교 통계학과
(2010년 1월 접수, 2010년 2월 채택)

요 약

최근들어 복제약(generic drug)의 개발이 활발히 이루어지고 있다. 현재 국내 생물학적 동등성 시험은 기존의 오리지널제제에 대하여 개발된 복제약이 동일한 질(quality), 안전성(safety) 그리고 효과(efficacy)를 갖는다는 것을 평균 생물학적 동등성(ABE)에 기초하여 증명하고 있다. 그러나 최근에 생물학적 동등성의 개념이 기존의 오리지널제제에 대하여 바꾸어 복용(switchability)할 수 있다는 개인 생물학적 동등성(IBE)의 개념으로 확장되었다. 미국 식약청은 개인 생물학적 동등성(IBE)을 평가하기 위하여 2 × 4 교차설계법을 권장하고 있다. 그러나 2 × 4 교차설계법은 실험기간이 길기 때문에 비용이나 피험자관리에 여러 가지 문제가 있어 실험기간을 줄인 2 × 3 교차설계법과 Chow 등 (2002)의 2 × 3 잉여대조(extra-reference)설계법이 제안되었다. 본 논문에서는 실험기간이 짧으면서 개인 생동성을 평가할 수 있는 3 × 2 교차설계법을 제안하고 모의실험을 통하여 2 × 3 교차설계법과 2 × 3 잉여대조설계법과 비교 및 그 결과를 고찰하였다.

주요용어: 교차설계법, 개인 생물학적 동등성, 잉여대조설계.

1. 서론

복제약의 약효가 오리지널 약에 대해 생물학적 동등성(bioequivalence, 이하 생동성)을 갖는 것을 입증하기 위하여 미국 식약청(US FDA)이 1992년에 제시한 기준에는 2 × 2 교차설계법을 이용하여 실험을 설계하고 이 실험에서 얻어진 두 약의 약효의 평균을 비교하여 통계학적으로 두 평균이 같음을 증명할 것을 요구하고 있다. 평균 약효가 동등할 것을 요구하는 기준을 평균 생동성 기준이라 하며 생동성 실험에서 보통 복제약을 실험제제(test formulation: T), 오리지널 약을 대조제제(reference formulation: R)로 표현한다.

평균 생동성 기준은 Anderson과 Hauck (1990)에 의해 두 제제의 평균뿐만 아니라 피험자 개개인에 대한 변동성을 동시에 고려한 개인 생동성의 개념으로 확장되었고 그 후 Sheiner (1992), Schall과 Luus (1993)나 Esinhart과 Chinchilli (1994) 등에 의해 개인 생동성에 대한 초기 연구가 이루어졌다. 이들의 연구를 토대로 미국 식약청(US FDA)에서도 2001년 새로이 개정된 규정을 통해 평균 생동성 뿐만이 아니라 개인 생동성의 사용을 동시에 증명할 것을 권장하고 있다. 특히, 평균 생동성 검정을 위해 사용되는 2 × 2 교차설계법으로는 피험자 내 분산(within-subject variation)의 추정값을 계산할 수 없기 때문에 FDA의 규정에는 2 × 2 교차설계법보다 실험 기간(period)이 두 배가 긴 2 × 4 교차설계법의 사용을 권장하고 있다.

2 × 3 교차설계법은 2 × 4 교차설계법의 실험 기간이 길기 때문에 발생하는 비용이나 피험자 관리에서 나타나는 여러 가지 문제점을 완화시키기 위해 제안되었는데 이 설계법은 2 × 2 교차설계법보다 여전히

본 연구는 2009년도 한남대학교 교비학술연구조성비 지원에 의하여 수행되었음.

³교신저자: (156-756) 서울시 동작구 흑석동 223, 중앙대학교 통계학과, 교수. E-mail: spark@cau.ac.kr

실험 기간이 길 뿐만 아니라 피험자 내 분산의 추정량이 두 실험 순서(sequence) 중 하나의 순서에서 얻어진 데이터만으로 계산되기 때문에 2×4 교차설계법에 비해 추정량의 자유도가 작아지는 단점을 갖게 된다. 자유도가 작아지는 단점을 보완하는 방법으로 Chow 등 (2002)은 실험 순서가 (TRT, RTR)인 2×3 교차설계법 대신 (TRR, RTR)의 실험 순서를 갖는 2×3 잉여대조(extra reference) 설계법을 제안하였다. 이 설계법은 두 제제에 대한 대칭성이 만족되지 못하기 때문에 잔류효과(carryover effect)가 존재하지 않는 모형이 필요한 반면, 각 실험순서에서 대조제제에 대한 관측값이 2개씩 존재하기 때문에 피험자 내 분산의 추정량이 2×3 교차설계법에 비해 2배나 큰 자유도를 갖게 된다. Chow 등 (2002)은 이 설계법이 2×4 교차설계법에 대해서도 비견될 수 있음을 보였다.

본 연구는 2×3 설계법의 세 번의 실험횟수도 너무 많은 경우 잉여대조 설계법을 이용하면 두 번의 기간을 갖는 3×2 교차설계법으로도 개인 생동성 검정방법을 만들 수 있다는 점에 착안하여 시작되었다. 2장에서는 실험순서가 (RT, TR, RR)인 3×2 교차설계법에 대한 혼합 모형(mixed model)을 제안하고, 이 모형 하에서 개인 생동성에 대한 검정방법을 3장에서 살펴본다. 4장에서는 3×2 교차설계법과 그리고 기존에 제안된 2×3 교차설계법과 2×3 잉여대조 설계법들이 모의실험을 통하여 비교된다. 기타 토의 및 결론이 5장에 기술된다.

2. 혼합모형(Mixed Model)

본 연구에서 관심이 있는 3×2 잉여대조 설계법에 대한 혼합 모형으로 Chow 등 (2002)이 제안한 모형을 변형한 모형을 고려한다. 여기서 실험 순서는 (RT, TR, RR)로 고정하고, 잔류효과는 충분한 휴약 기간이 주어져 무시할 수 있다고 가정한다.

y_{ijk} 를 생동성 실험의 k 번째 순서(sequence)에 참여한 i 번째 피험자(subject)로부터 j 번째 기간(period)에 얻어진 반응값($i = 1, \dots, n_k; j = 1, 2; k = 1, 2, 3$)이라 할 때 모형은 다음과 같다:

$$y_{ijk} = \mu + F_l + W_{ijk} + S_{ikl} + \varepsilon_{ijk}, \quad (2.1)$$

여기서 μ 는 전체평균, F_l 은 제제 l 의 효과 ($l = T$ or $R, F_T + F_R = 0$) 그리고 W_{ijk} 는 기간, 순서 그리고 교호작용 효과 또는 모형에 따라서 잔류효과를 포함하는 고정효과(fixed effects)로 $\sum_k \bar{W}_{lk} = 0$ (\bar{W}_{lk} 는 고정된 (l, k) 에 대한 W_{ijk} 의 평균)이다. S_{ikl} 은 k 번째 순서에서 제제 l 을 투여받은 i 번째 피험자의 랜덤 효과(random effect)로 (S_{ikT}, S_{ikR}) 가 서로 독립이고 동일하게 평균 0, 분산 공분산 행렬이

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

인 이변량 정규분포를 따르는 확률벡터이다. 실험제제 T 는 없고 대조제제 R 만 두 개 존재할 때는 (S_{ikT}, S_{ikR}) 의 분포는 평균 0, 분산 공분산 행렬의 모든 원소가 σ_{BR}^2 인 이변량 정규분포가 된다. 오차항 ε_{ijk} 는 서로 독립이며, 동일하게 평균 0, 분산 σ_{Wl}^2 ($l = T$ or R)인 정규분포를 따르고 S_{ikl} 와 ε_{ijk} 는 서로 독립이라 가정한다.

3. 개인 생물학적 동등성(Individual Bioequivalence)

식 (2.1)의 혼합모형을 적용하면 FDA에서 제시하는 개인 생동성에 대한 기준이 얻어진다:

$$\theta = \frac{\delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2}{\max(\sigma_{WR}^2, \sigma_0^2)} < \theta_I, \quad (3.1)$$

여기서 $\delta = F_T - F_R$ 는 제제간 평균 생동성의 차이를 나타내고, 제제와 피험자간 교호작용분산은 $\sigma_D^2 = \sigma_{BT}^2 + \sigma_{BR}^2 - 2\rho\sigma_{BT}\sigma_{BR}$ 로 정의된다. 분모의 두 모수값이 $\sigma_{WR}^2 \geq \sigma_0^2$ 인 경우 (3.1)의 기준을 reference scaled 기준, $\sigma_{WR}^2 < \sigma_0^2$ 인 경우 constant scaled 기준이라 하며 FDA 규정에는 $\sigma_0^2 = 0.04$ 를 제시하고 있다. 개인 생동성 지표 θ 가 상한 θ_I 보다 작으면 두 제제간의 개인 생동성을 인정하는데 FDA 권장값은 $\theta_I = 2.4948$ 이다.

식 (3.1)의 기준으로부터 우리는 개인 생동성에 대한 가설

$$H_0 : \lambda \geq 0, \quad H_1 : \lambda < 0$$

을 설정할 수 있는데, 여기서 모수 λ 는

$$\lambda = \delta^2 + \sigma_D^2 + \sigma_{WT}^2 - \sigma_{WR}^2 - \theta_I \max(\sigma_{WR}^2, \sigma_0^2) \quad (3.2)$$

이고 λ 에 대한 95% 신뢰구간의 상한이 0보다 작으면 귀무가설 H_0 를 기각한다.

Hyslop 등 (2000)은 λ 를 구성하고 있는 개별 모수들에 대한 추정량들을 이용하여 λ 에 대한 근사 신뢰구간을 구하는 방법을 제안하였는데 이는 Howe (1974)의 연구에 기초를 두고 있기 때문에 이 방법으로 구해진 신뢰구간을 Howe의 신뢰구간이라 부르기로 한다. λ 에 대한 Howe의 신뢰상한을 구하려면 개별 모수들에 대한 추정량들이 서로 독립이어야 한다. Chow 등 (2002)은 2 × 3 교차설계법에서 λ 에 대한 Howe의 신뢰상한을 구하였다.

이 방법의 장점중 하나는 λ 를 구성하고 있는 개별 모수들 중에 몇 개의 모수를 결합하여 추정량을 구하기 쉬운 새로운 모수를 만들 수 있는 경우에도 여전히 같은 방법으로 신뢰구간을 구할 수 있다는 점이다. 예를 들어 실험 순서가 RT이거나 TR인 경우 y_T 와 y_R 을 각각 실험제제와 대조제제의 특성값이라 하면, 혼합모형 (2.1)에서

$$\text{var}(y_T - y_R) = \sigma_D^2 + \sigma_{WT}^2 + \sigma_{WR}^2 \quad (3.3)$$

이기 때문에 식 (3.2)의 λ 는

$$\lambda = \delta^2 + \sigma_{1,1}^2 - 2\sigma_{WR}^2 - \theta_I \max(\sigma_{WR}^2, \sigma_0^2) \quad (3.4)$$

로 표현된다. 여기서 새로운 분산 $\sigma_{1,1}^2$ 은

$$\sigma_{a,b}^2 = \sigma_D^2 + a\sigma_{WT}^2 + b\sigma_{WR}^2 \quad (3.5)$$

의 $a = 1, b = 1$ 일 때 값으로 정의된다.

만약 δ^2 과 $\sigma_{1,1}^2$ 그리고 σ_{WR}^2 에 대한 서로 독립인 추정량을 구할 수 있다면 식 (3.4)로부터 λ 에 대한 Howe의 신뢰상한을 구할 수 있게 된다. 식 (3.3)에 의해 RT나 TR처럼 실험 기간이 2인 경우에도 $\sigma_{1,1}^2$ 의 불편추정량을 구할 수 있고, 하나 남은 문제가 σ_{WR}^2 에 대한 추정인데 이 문제는 대조제제만의 실험인 RR을 추가하면 해결된다. 식 (3.5)으로 정의된 분산 $\sigma_{a,b}^2$ 은 Hyslop 등 (2000)이나 Chow 등 (2002) 등에 의해 2 × 4 교차설계법이나 2 × 3 교차설계법의 연구에 사용되었다.

3.1. 3 × 2 교차설계법 (RT, TR, RR)

실험 순서가 (RT, TR, RR)인 3 × 2 교차설계법에서 순서 k 의 i 번째 피험자의 두 관측값의 차이를 d_{ik} ($k = 1, 2, 3; i = 1, \dots, n_k$)라 하면 δ^2 , $\sigma_{1,1}^2$ 그리고 σ_{WR}^2 에 대한 불편추정량과 분포는 다음과 같다:

$$\hat{\delta} = \frac{\bar{d}_1 - \bar{d}_2}{2} \sim N\left(\delta, \frac{\sigma_{1,1}^2}{4} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$\begin{aligned}\hat{\sigma}_{1,1}^2 &= \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_k)^2 \sim \sigma_{1,1}^2 \frac{\chi_{n_1+n_2-2}^2}{n_1 + n_2 - 2} \\ \hat{\sigma}_{WR}^2 &= \frac{1}{2(n_3 - 1)} \sum_{i=1}^{n_3} (d_{i3} - \bar{d}_3)^2 \sim \sigma_{WR}^2 \frac{\chi_{n_3-1}^2}{n_3 - 1}\end{aligned}\quad (3.6)$$

식 (3.6)의 세 통계량은 서로 독립이므로 λ 에 대한 Howe의 95% 신뢰상한 $\hat{\lambda}_U$ 을 구할 수 있다. 먼저 $\sigma_{WR}^2 \geq \sigma_0^2$ 일 때

$$\hat{\lambda}_U = \hat{\delta}^2 + \hat{\sigma}_{1,1}^2 - (2 + \theta_I)\hat{\sigma}_{WR}^2 + \sqrt{U}$$

이 된다. 여기서 $U = A^2 + B^2 + C^2$ 이고

$$\begin{aligned}A &= \left(\left| \hat{\delta} \right| + t_{0.95; n_1+n_2+n_3-3} \frac{\hat{\sigma}_{1,1}}{2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)^2 - \hat{\delta}^2 \\ B &= \hat{\sigma}_{1,1}^2 \left(\frac{n_1 + n_2 - 2}{\chi_{0.05; n_1+n_2-2}^2} - 1 \right) \\ C &= (2 + \theta_I)\hat{\sigma}_{WR}^2 \left(\frac{n_3 - 1}{\chi_{0.95; n_3-1}^2} - 1 \right)\end{aligned}$$

이다. 만약 $\sigma_{WR}^2 < \sigma_0^2$ 이면

$$\hat{\lambda}_U = \hat{\delta}^2 + \hat{\sigma}_{1,1}^2 - 2\hat{\sigma}_{WR}^2 - \theta_I\hat{\sigma}_0^2 + \sqrt{U'}$$

이 되고, 여기서 $U' = A^2 + B^2 + C'^2$ 이고

$$C' = 2\hat{\sigma}_{WR}^2 \left(\frac{n_3 - 1}{\chi_{0.95; n_3-1}^2} - 1 \right)$$

이다. 2×3 교차설계법과 2×3 잉여대조설계법에서 λ 에 대한 Howe의 신뢰상한은 $\sigma_{WR}^2 \geq \sigma_0^2$ 일 때와 $\sigma_{WR}^2 < \sigma_0^2$ 일 때로 나누어 Chow 등 (2002)의 식 (10), (12)와 식 (14), (16)으로 부터 각각 얻을 수 있다. Chow 등 (2002)이 제안한 검정법과의 비교는 다음 장에서 모의실험을 통해 이루어진다.

4. 모의실험결과(Simulation Results)

이 장에서는 모의실험을 통하여 다음의 내용을 살펴보고자 한다.

1. 3×2 , 2×3 교차설계법과 2×3 잉여대조 설계법의 제1종 오류의 확률을 계산하고 비교한다.
2. 검정력(statistical power)을 이용하여 3×2 , 2×3 교차설계법과 2×3 잉여대조 설계법의 검정 수행 능력을 비교한다.

개인 생동성에서는 식 (3.2)에 영향을 주는 4개의 모수, δ , σ_D^2 , σ_{WT}^2 그리고 σ_{WR}^2 가 있다. 본 논문에서는 이 모수의 값들을 Hyslop 등 (2000)과 Chow 등 (2002)의 값들과 동일한 값들로 고정하고 모의실험을 실시하였다. 즉, $\sigma_D^2 = 0, 0.2$, $\sigma_{WT} = 0.15, 0.2, 0.3, 0.5$, $\sigma_{WR} = 0.15, 0.2, 0.3$ 으로 하였다. 표본의 수는 $n = n_1 = n_2 = n_3 = 10(5)40$ 으로 하였고 σ_0 와 θ_U 는 각각 0.2와 2.4948로 하였다.

각 모수와 표본의 수에서 제 1종 오류확률과 검정력을 구하기 위하여 30,000번의 모의실험이 수행되었다. 그리고 그 결과를 표 4.1과 4.2에 보였다. 검정력에 대한 모의실험결과는 그림 4.1~4.4에 보였다. 그림 4.1과 4.2는 몇가지 고정된 λ 에 대하여 검정력 대 표본의 수의 관계를 나타내고 있고 반면에 그림 4.3과 4.3는 표본의 수가 $n = 20$ 으로 고정되어 있을 때 검정력과 λ 의 관계를 보였다.

표 4.1. Type 1 error probability of IBE tests when $\sigma_{WR} \leq 0.2$ (30,000 simulations).

δ	Parameter			n	3 × 2 cross-over	2 × 3 cross-over	2 × 3 extra-reference
	σ_D	σ_{WT}	σ_{WR}		RT/TR/RR	TRT/RTR	TRR/RTR
0.3158	0	0.15	0.15	10	0.0453	0.0334	0.0481
				15	0.0469	0.0397	0.0487
				20	0.0500	0.0437	0.0515
				25	0.0522	0.0454	0.0519
				30	0.0520	0.0468	0.0519
				35	0.0534	0.0481	0.0515
				40	0.0530	0.0517	0.0539
0.2868	0	0.2	0.15	10	0.0450	0.0282	0.0442
				15	0.0474	0.0331	0.0465
				20	0.0469	0.0388	0.0510
				25	0.0514	0.0426	0.0492
				30	0.0524	0.0436	0.0501
				35	0.0498	0.0456	0.0501
				40	0.0528	0.0476	0.0509
0.3424	0	0.15	0.2	10	0.0565	0.0494	0.0600
				15	0.0581	0.0516	0.0588
				20	0.0574	0.0537	0.0601
				25	0.0570	0.0541	0.0589
				30	0.0575	0.0531	0.0584
				35	0.0557	0.0534	0.0571
				40	0.0577	0.0558	0.0606
0.3158	0	0.15	0.15	10	0.0453	0.0464	0.0588
				15	0.0469	0.0496	0.0584
				20	0.0500	0.0532	0.0598
				25	0.0522	0.0500	0.0570
				30	0.0520	0.0508	0.0582
				35	0.0534	0.0533	0.0590
				40	0.0530	0.0530	0.0586
0.2231	0	0.3	0.2	10	0.0516	0.0343	0.0504
				15	0.0521	0.0410	0.0527
				20	0.0535	0.0434	0.0525
				25	0.0526	0.0431	0.0524
				30	0.0525	0.0474	0.0534
				35	0.0547	0.0495	0.0536
				40	0.0538	0.0490	0.0546
0.2444	0.2	0.15	0.15	10	0.0430	0.0238	0.0420
				15	0.0450	0.0313	0.0432
				20	0.0454	0.0339	0.0430
				25	0.0490	0.0364	0.0451
				30	0.0495	0.0408	0.0461
				35	0.0481	0.0401	0.0492
				40	0.0484	0.0423	0.0481
0.2779	0.2	0.15	0.2	10	0.0558	0.0407	0.0520
				15	0.0543	0.0456	0.0544
				20	0.0530	0.0482	0.0599
				25	0.0568	0.0492	0.0561
				30	0.0558	0.0510	0.0571
				35	0.0553	0.0509	0.0577
				40	0.0536	0.0509	0.0545
0.2444	0.2	0.2	0.2	10	0.0430	0.0372	0.0521
				15	0.0450	0.0425	0.0530
				20	0.0454	0.0471	0.0549
				25	0.0490	0.0498	0.0548
				30	0.0495	0.0474	0.0544
				35	0.0481	0.0484	0.0558
				40	0.0484	0.0501	0.0570

표 4.2. Type 1 error probability of IBE tests when $\sigma_{WR} > 0.2$ (30,000 simulations).

δ	Parameter			n	3 × 2 cross-over	2 × 3 cross-over	2 × 3 extra-reference
	σ_D	σ_{WT}	σ_{WR}		RT/TR/RR	TRT/RTR	TRR/RTR
0.5403	0	0.15	0.3	10	0.0495	0.0479	0.0485
				15	0.0517	0.0507	0.0515
				20	0.0502	0.0486	0.0522
				25	0.0498	0.0498	0.0496
				30	0.0500	0.0498	0.0506
				35	0.0533	0.0505	0.0521
				40	0.0532	0.0510	0.0520
0.4738	0	0.3	0.3	10	0.0495	0.0404	0.0460
				15	0.0492	0.0457	0.0488
				20	0.0505	0.0461	0.0493
				25	0.0495	0.0480	0.0504
				30	0.0490	0.0483	0.0488
				35	0.0483	0.0477	0.0495
				40	0.0498	0.0479	0.0493
0.2539	0	0.5	0.3	10	0.0410	0.0282	0.0403
				15	0.0415	0.0336	0.0407
				20	0.0446	0.0363	0.0407
				25	0.0430	0.0403	0.0442
				30	0.0459	0.0384	0.0433
				35	0.0458	0.0414	0.0439
				40	0.0469	0.0415	0.0453
0.9129	0	0.2	0.5	10	0.0484	0.0472	0.0506
				15	0.0493	0.0508	0.0536
				20	0.0533	0.0482	0.0513
				25	0.0508	0.0508	0.0505
				30	0.0500	0.0517	0.0517
				35	0.0524	0.0502	0.0520
				40	0.0535	0.0513	0.0520
0.7896	0	0.5	0.5	10	0.0486	0.0405	0.0482
				15	0.0488	0.0446	0.0478
				20	0.0502	0.0484	0.0487
				25	0.0519	0.0467	0.0494
				30	0.0500	0.0476	0.0471
				35	0.0509	0.0490	0.0506
				40	0.0495	0.0481	0.0501
0.4842	0.2	0.2	0.3	10	0.0498	0.0408	0.0486
				15	0.0504	0.0460	0.0507
				20	0.0494	0.0454	0.0499
				25	0.0516	0.0464	0.0502
				30	0.0491	0.0480	0.0499
				35	0.0492	0.0491	0.0517
				40	0.0523	0.0495	0.0506
0.4295	0.2	0.3	0.3	10	0.0490	0.0396	0.0465
				15	0.0484	0.0423	0.0481
				20	0.0487	0.0443	0.0463
				25	0.0521	0.0453	0.0488
				30	0.0518	0.0452	0.0475
				35	0.0473	0.0484	0.0483
				40	0.0518	0.0470	0.0500
0.8622	0.2	0.3	0.5	10	0.0494	0.0468	0.0481
				15	0.0496	0.0472	0.0496
				20	0.0500	0.0494	0.0498
				25	0.0532	0.0485	0.0499
				30	0.0521	0.0528	0.0508
				35	0.0502	0.0502	0.0507
				40	0.0502	0.0477	0.0501

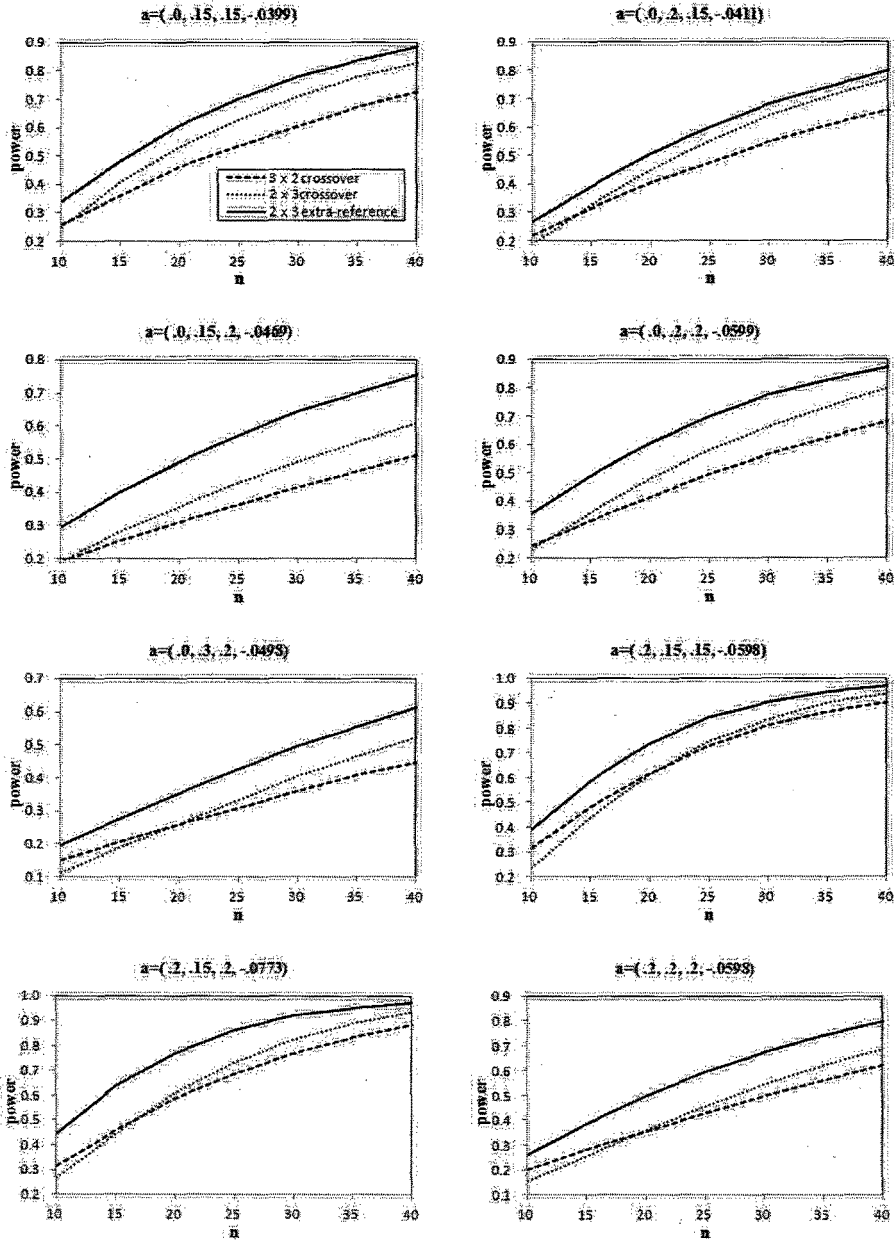


그림 4.1. Power of IBE tests versus n ; $a = (\sigma_D, \sigma_{WT}, \sigma_{WR}, \lambda)$.

모의실험결과는 다음과 같다. 제1종 오류의 확률은 3 × 2 교차설계법과 2 × 3 잉여대조 설계법 두 설계법 모두 유의수준 $\alpha = 0.05$ 를 잘 맞추어주고 있다 (표 4.1과 4.2). 반면에 2 × 3 교차설계법에서는 3 × 2와 2 × 3 잉여대조 설계법에 비하여 제1종 오류의 확률이 낮게 나오는 경향이 있다. 즉, 다분히 보수적(conservative)인 경향을 나타내고 있다. 특히 표본의 수가 적은 경우 그런 경향은 더욱 두드러진다.

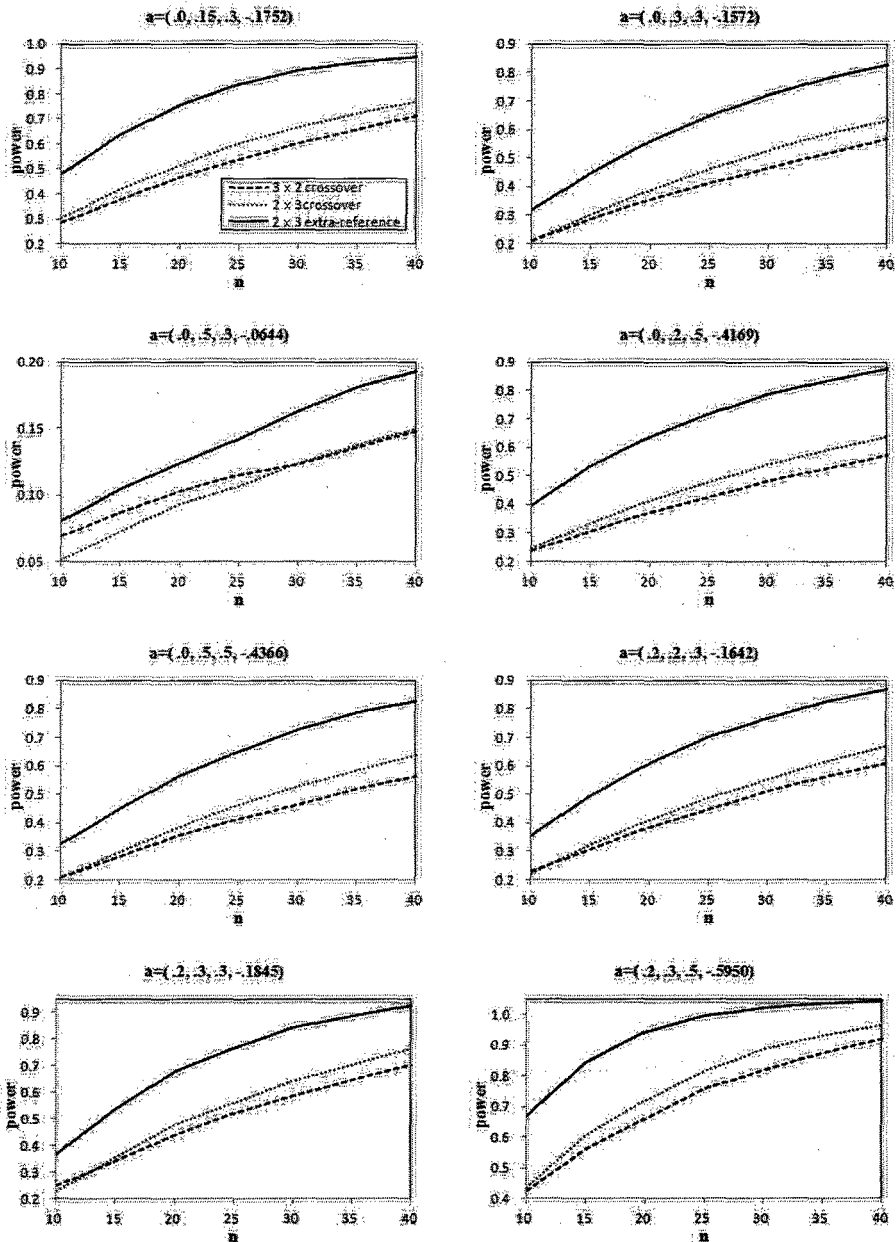


그림 4.2. Power of IBE tests versus n ; $a = (\sigma_D, \sigma_{WT}, \sigma_{WR}, \lambda)$.

n 이 커지면서 제1종 오류의 확률이 정해진 유의수준에 근접 하고는 있지만 유의수준을 맞추기 위해 너무 많은 표본의 수를 필요로 하는 경향이 있다.

다양한 모수 하에서 검정력을 살펴보았다 (그림 4.1과 4.2). Chow와 Wang (2002)에서 언급 한 바와 같이 모수값이 $\sigma_D = 0, \sigma_{WT} = 0.5, \sigma_{WR} = 0.3$ 인 경우 n 이 매우 커짐에도 불구하고 검정력이 매우 낮은

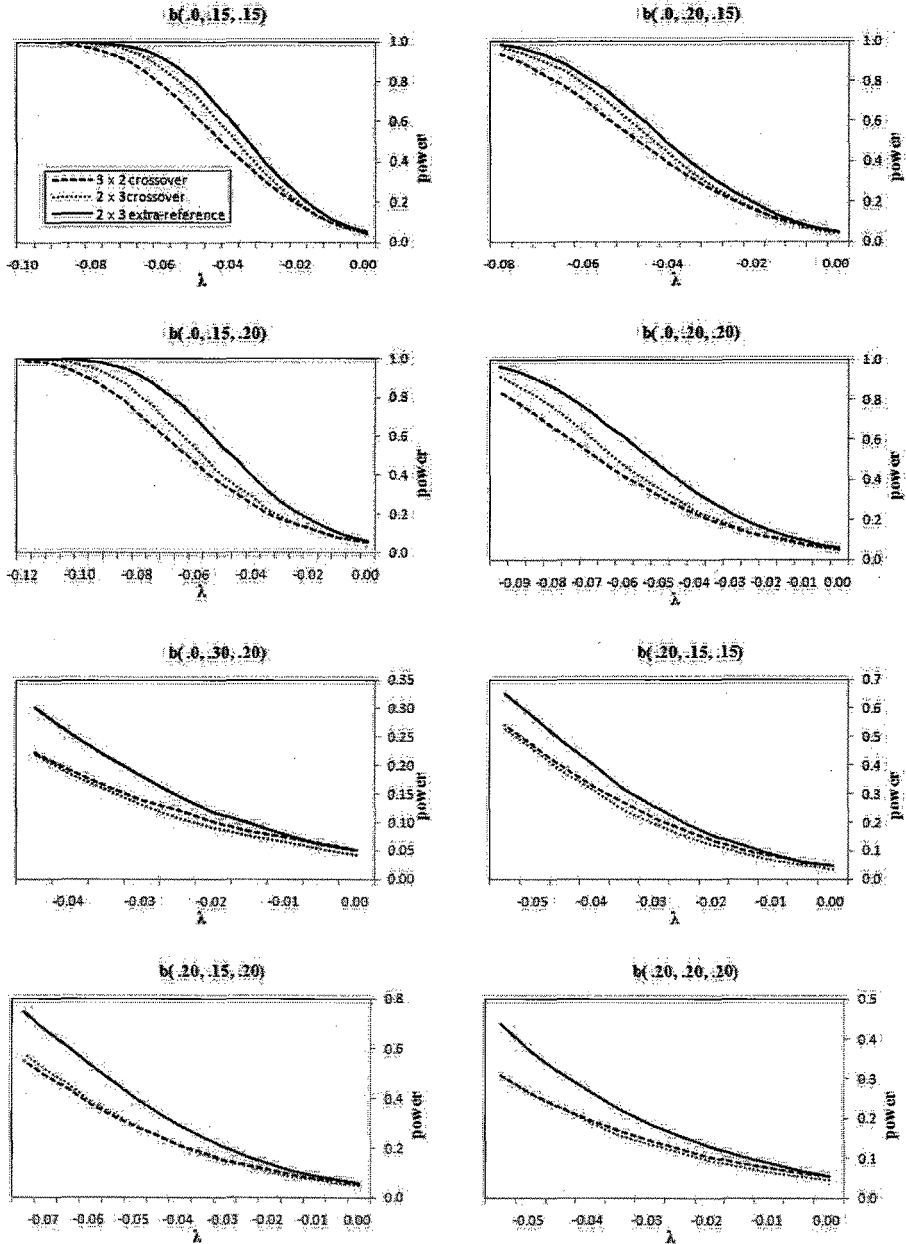


그림 4.3. Power of IBE tests versus λ ; $b = (\sigma_D, \sigma_{WT}, \sigma_{WR})$.

것을 확인할 수 있다. 이는 실험제제가 대조제제 보다 더 높은 피험자 내 분산(variability)을 가질 경우에 실제 개인 생동성이 인정되는 경우에도 개인 생동성이라고 주장하기 어려움을 의미한다. 그림 4.1과 4.2에서 2 × 3 잉여대조 설계법이 검정력이 가장 큰 것을 알 수 있다. 그리고 3 × 2와 2 × 3 교차설계법에서는 많은 경우에 2 × 3 교차설계법이 더 높은 검정력을 보이고 있다. 하지만 $\sigma_{WT} > \sigma_{WR}$ 이거나

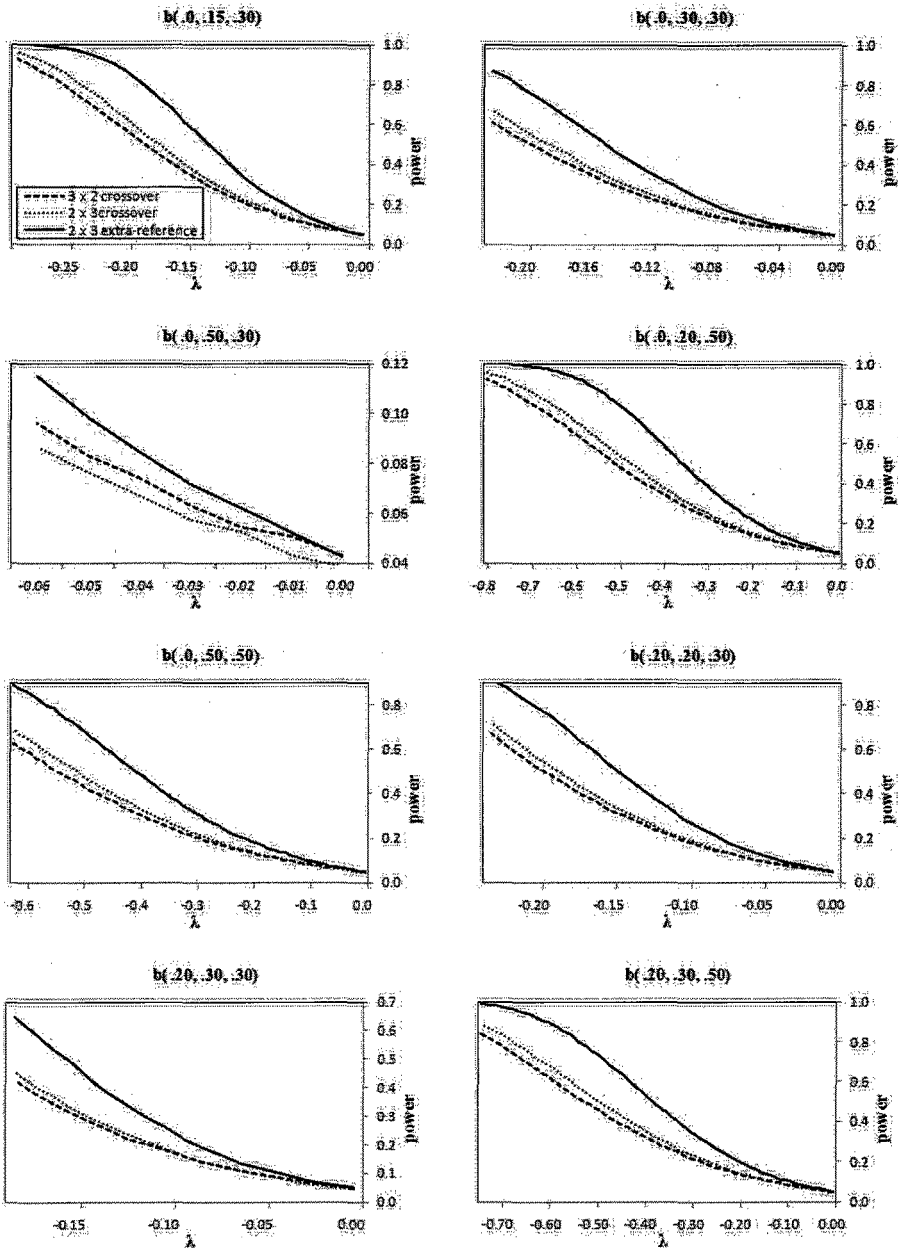


그림 4.4. Power of IBE tests versus λ ; $b = (\sigma_D, \sigma_{WT}, \sigma_{WR})$.

$\sigma_D > 0$ 인 경우에 표본의 수가 작은 경우 3×2 교차설계법이 2×3 교차설계법 보다 높은 검정력을 갖는다.

그림 4.3과 4.4에서는 세 교차설계법 모두 피험자의 수가 20인 경우 λ 값이 감소 할때 검정력의 변화를 살펴보았다. 2×3 잉여대조 설계법이 모든 λ 의 구간에서 높은 검정력을 갖고 있다. 그러나 3×2 와

2 × 3 교차설계법의 검정력의 비교에 있어서는 $\sigma_{WT} > \sigma_{WR}$ 이거나 $\sigma_D > 0$ 인 경우에 3 × 2가 2 × 3 교차설계법 보다 높은 검정력을 가지고 있는 것으로 나타나는데, 표 4.1의 제1종 오류를 살펴보면 $\sigma_{WT} > \sigma_{WR}$ 이거나 $\sigma_D > 0$ 인 경우 2 × 3 교차설계법이 다른 두 방법보다 훨씬 더 보수적으로 나타나므로 이 영향도 상당히 받았을 것으로 보인다.

5. 결론(Conclusions)

개인 생동성은 현재 미국이나 선진국에서 평균 생동성에 추가로 새롭게 채택이 되고 있는 생동성 기준이다. 본 논문에서는 현재 생동성 시험에서 개인 생동성을 밝히는데 이용하는 실험계획법인 2 × 3 잉여대조 설계법과 2 × 3 교차설계법에 대하여 고찰해보고 이 실험설계법이 가지고 있는 단점 중 하나인 세번의 실험 기간으로 인해 발생하는 여러가지 문제점들을 개선하기 위하여 시험의 기간이 하나 짧은 3 × 2 교차설계법의 사용가능성에 대하여 모의실험을 통하여 알아 보았다. 물론 2 × 3 교차설계법이나 2 × 3 잉여대조설계법은 3 × 2 교차설계법에 비하여 n 만큼 적은 피험자 수를 요구한다. 반면에 3 × 2는 피험자의 수는 n 만큼 늘어나 나지만 전체 관측치는 $6n$ 으로 동일한 수의 관측치를 이용하고 반면에 더 적은 시험기간을 요구하므로 부득이하게 시험의 기간을 길게 가져가기 어려운 경우나 실험의 기간을 길게 가져감으로서 생기는 결측치의 가능성이 높은 경우에 세번의 기간이 있는 2 × 3 잉여대조설계법 및 2 × 3 교차설계법의 대안으로 사용될 수 있을 것이다.

참고문헌

- Anderson, S. and Hauck, W. W. (1990). Consideration of individual bioequivalence, *Journal of Pharmacokinetics and Pharmacodynamics*, **18**, 259-273.
- Chow, S. C., Shao, J. and Wang, H. (2002). Individual bioequivalence testing under 2 × 3 designs, *Statistics in Medicine*, **21**, 629-648.
- Esinhart, J. D. and Chinchilli, V. M. (1994). Extension to the use of tolerance intervals for the assessment of individual bioequivalence, *Journal of Biopharmaceutical Statistics*, **4**, 39-52.
- 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.
- Sheiner, L. B. (1992). Bioequivalence revisited, *Statistics in Medicine*, **11**, 1777-1788.
- Schall, R. and Luus, H. G. (1993). On population and individual bioequivalence, *Statistics in Medicine*, **12**, 1109-1124.
- US Food and Drug Administration (1992). *Guidance on Statistical procedures for Bioequivalence using a standard two-treatment crossover design*, Rockville, MD.
- US Food and Drug Administration (2001). *Statistical Approaches to Establishing Bioequivalence*, Rockville, MD.

Individual Bioequivalence Tests under 3×2 Design

Gyu-Jin Jung¹ · Nam-Kyoo Lim² · Sang-Gue Park³

¹Department of Business Statistics, Hannam University

²National Clinical Research Coordination Center, Korea Centers for Disease Control and Prevention

³Department of Statistics, Chung-Ang University

(Received January 2010; accepted February 2010)

Abstract

In recent years, more generic drug products became available. The current regulation for assessing the bioequivalence of two drug formulations is based on the concept of average bioequivalence. This approach has been indicated to be insufficient for assessing switchability between two drug formulations and US FDA has adopted individual bioequivalence as one of the bioequivalence criterion since 2001. The US FDA recommends that individual bioequivalence be assessed based on 2×4 crossover design, while a 2×3 crossover design may be used as an alternative design to reduce the length and cost of the study. In this paper, a statistical procedure for assessment of individual bioequivalence under 3×2 crossover designs is proposed and some statistical points are discussed with 2×3 crossover design and 2×3 extra-reference design through simulation studies.

Keywords: Cross-over design, extra-reference design, individual bioequivalence.

This paper has been supported by the 2009 Hannam University Research Fund.

³Corresponding author: Professor, Department of Statistics, Chung-Ang University, 223 Heuksuk-Dong, Dongjak-Gu, Seoul 156-756, Korea. E-mail: spark@cau.ac.kr