

활성대조군을 이용한 두 군 설계와 위약군을 포함한 세 군 설계의 비열등성 시험

이지선¹, 김동재²

¹가톨릭대학교 의학통계학과; ²가톨릭대학교 의학통계학과

(2008년 9월 접수, 2008년 10월 채택)

요약

비열등성 시험시에 치료군(treatment group)과 활성대조군(active control group)을 이용한 모형으로 Hauschke 등 (1999)이 제안한 모수적 검정법이 있다. 이 방법은 위약군(placebo group)과 직접적인 비교가 불가능하므로 Pigeot 등 (2003)이 세 군으로 확장한 검정법을 제안하였다. 그러나 이와 같은 두 검정법은 구체적인 분포가정이 필요하다. 이런 단점을 보완하기 위하여 본 논문에서는 비모수적 방법으로서 두 군 설계에 Wilcoxon 순위합 검정 (Wilcoxon, 1945)을 이용한 방법을, 세 군 설계에 Scheirer 등 (1976)이 제안한 선형대비검정을 확장한 새로운 방법을 제안한다. 또한 모의실험을 통하여 모수적 방법과 비모수적 방법간의 검정력을 비교하였다.

주요용어: 비열등성, 비모수, 가설검정.

1. 서론

비열등성 시험이란 새로운 치료약의 효과가 통계적, 임상적으로 기존 치료약의 효과에 비해 열등하지 않음을 보이는 것에 목적을 두고 있으며 현재 국내외 많은 제약회사에서 주로 시행되고 있다. 이처럼 비열등성 시험에 관심을 갖는 이유는 주평가변수에서는 새로운 치료약이 기존치료약보다 열등하지 않지만(비슷하지만) 그 외 다른 측면에서 장점이 있는 경우이다. 예를 들면, 약의 독성이 덜 하거나, 투여하기 쉽거나, 비용이 싸거나 하는 등의 이유이다. 이와 같은 비열등성 시험을 위한 가설은 치료군과 활성대조군의 모집단 평균의 차로써 표현된다.

$$H_0 : \mu_1 - \mu_2 \leq \delta \quad vs. \quad H_1 : \mu_1 - \mu_2 > \delta, \quad \delta < 0,$$

여기서 μ_1 , μ_2 는 각각 치료군과 활성대조군의 모집단의 평균이며 δ 는 실험 전에 임상적인 정보를 이용하여 결정되어지는 비열등성의 한계이다. δ 의 선택은 가장 중요한 사항중에 하나이다.

두 군 설계상에서 치료군의 비열등성 시험을 위한 기존의 방법으로는 Hauschke 등 (1999)이 동등성 검정시에 제안한 방법이 있다. δ 를 활성대조군의 평균 반응값에 대한 일정 비율로 정하고 이에 따른 가설에 대하여 모수적으로 검정하는 방법이다. 그러나 시험의 목적이 단순히 새로운 치료약이 기존 치료약에 비해 비열등함을 보이는 것 뿐만 아니라 위약에 대한 효능을 입증하고자 하는 경우 위약처방이 윤리적으로 문제가 된다면 기존치료약과의 비열등성 평가로서 목적을 달성할 수 있다. 그러나 이러한 두 군

¹(137-701) 서울시 서초구 반포동 505, 가톨릭대학교 의학통계학과, 석사과정. E-mail: yjs0229@hanmail.net

²교신저자: (137-701) 서울시 서초구 반포동 505, 가톨릭대학교 의학통계학과, 교수.

E-mail: djkim@catholic.ac.kr

시험은 설계상의 몇가지 어려움이 있다. 이 설계는 내적 타당도의 측정방법이 결여되어 있어 외적 타당도 검정이 필요하며 설계나 시험의 수행에서 발생하는 많은 결점들로 인하여 결과를 비열등한 쪽으로 비뚤리게 하는 경향이 있다. 따라서 대조치료약의 선택에 주의를 필요로 한다. 새로운 임상시험은 대조치료약이 임상적으로 명백한 유효성을 보인 우위성 평가시험과 동일한 일차변수, 대조치료약의 용량, 연구대상 포함기준 등과 같은 설계상의 특징을 가지고 있어야 한다. 그러나 위약의 반응값이 크고 대조치료약의 효과가 작거나 지나치게 변동성이 심한 질병인 우울증, 치매, 울혈성심부전증의 운동부하, 중후성 위식도역류 등의 경우 대조치료약의 효과가 일관성을 유지하기 어렵다고 할 수 있다 (D'Agostino 등, 2003; Hung 등, 2003; Hwang과 Morikawa, 1999).

이런 단점을 보완하기 위하여 위약군을 포함시킨 세 군 설계상의 비열등성 시험은 Pigeot 등 (2003)이 제안한 방법이 있다. Hauschke 등 (1999)이 제안한 방법을 세 군으로 확장시킨 것으로 δ 를 활성대조군의 평균 반응값과 위약의 평균 반응값의 차에 대한 일정 비율로 정하고 이에 따른 가설에 대하여 모수적으로 검정하는 방법이다. 위약처방에 윤리적인 문제가 없는 경우라면 세 군 설계는 위약에 대한 새로운 치료약의 우위성을 확립하고 시험설계의 타당성을 확인하며 동시에 대조치료약에 대한 치료약의 유효성 및 안전성의 유사성을 평가 등의 여러가지 목표를 동시에 성취할 수 있는 좋은 방법이다.

위에서 언급한 바와 같이 경우에 따라서 두 군과 세 군 설계의 비열등성 시험을 실시하게 되는데 지금까지 제시된 방법들은 모집단에 대하여 구체적인 분포함수를 가정했을 때 사용하는 모수적 방법이다. 따라서 분포함수의 정보가 없거나 정규성을 만족하지 못하여 특정 분포를 가정할 수 없는 경우, 비모수적 방법을 사용함으로써 통계적 검정력을 높일 수 있다.

따라서 본 논문에서는 두 군 모형에 대한 비모수적인 검정 방법으로서 Wilcoxon 순위합 검정 (Wilcoxon, 1945)을 이용하는 방법을 제안하였으며 세 군 모형의 경우 Scheirer 등 (1976)이 제안한 선형대비 검정을 각 군의 표본수가 다른 경우로 확장하여 비열등성 시험에 이용하는 새로운 방법을 제안하였다. 그리고 모의실험을 통하여 각 설계상에서 기존의 모수적 방법과 본 논문에서 제안한 비모수적 방법간의 검정력을 비교하였다.

2. 두 군 모형의 가설검정

두군 비열등성 시험의 가설은 다음과 같다.

$$H_0 : \mu_1 - \mu_2 \leq \delta \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 > \delta, \quad (2.1)$$

여기서 μ_1, μ_2 는 각각 치료군과 활성대조군의 모집단 평균이며 $\delta < 0$ 은 비열등성 한계이다. 치료군이 활성대조군의 치료효과의 특정 비율을 유지하는지에 시험 목적을 두고 비열등성 한계를 $\delta = \lambda\mu_2, -1 < \lambda < 0$ 로 표현한다. λ 는 대조군 치료효과에 대하여 허용할 수 있는 정도의 감소율을 말한다. 즉, 대조군 효능의 어느 정도 감소율까지 치료군과 효능 차이가 없다고 할 수 있는가를 표현하고 있다. 일반적으로 20%감소($\lambda = -1/5$)는 임상적으로 중요하게 생각하지 않는다. 이 때 $\theta = 1 + \lambda$ 로 정의하면 식 (2.1)은

$$\begin{aligned} H_0 : \frac{\mu_1}{\mu_2} \leq \theta \quad \text{vs.} \quad H_1 : \frac{\mu_1}{\mu_2} > \theta, \quad \theta \in [1 + \lambda, 1) \\ H_0 : \mu_1 - \theta\mu_2 \leq 0 \quad \text{vs.} \quad H_1 : \mu_1 - \theta\mu_2 > 0, \quad \theta \in [1 + \lambda, 1) \end{aligned} \quad (2.2)$$

으로 표현된다.

본 논문에서는 가설 (2.2)에 대한 검정으로 기존의 모수적 방법을 제시하고 Wilcoxon 순위합 검정을 이용하는 새로운 방법을 제안하였다.

2.1. Hauschke의 모수적 방법

확률표본 X_{ij} ($i = 1, 2, j = 1, 2, \dots, n_i$)는 평균이 μ_i 이고 분산이 σ^2 인 정규분포에서 추출된 확률변수이다. $i = 1, 2$ 는 순서대로 치료군과 활성대조군을 말한다. \bar{X}_i, S_i^2 ($i = 1, 2$)을 각각 표본평균과 표본 분산이라고 할 때 가설 (2.2)을 검정하는 검정통계량

$$T = \frac{\bar{X}_1 - \theta \bar{X}_2}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{\theta^2}{n_2}}} \tag{2.3}$$

은 자유도가 $(n_1 + n_2 - 2)$ 인 t 분포를 따르며 $\hat{\sigma}^2 = \{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2\} / (n_1 + n_2 - 2)$ 이다. 이 검정의 기각역은 $T > t_{\alpha}(n_1 + n_2 - 2)$ 이며 t_{α} 는 자유도가 $(n_1 + n_2 - 2)$ 인 t 분포의 상위 100α 백분위수이다.

2.2. Wilcoxon 순위합 검정을 이용한 비모수적 방법

확률표본 X_{ij} ($i = 1, 2, j = 1, 2, \dots, n_i$)는 각각의 분포함수 $F_i(x) = F(x - \Delta_i)$ 에서 추출된 확률변수이다. 이 때 활성대조군과 치료군의 평균차이는 $\Delta_2 - \Delta_1$ 이다.

가설 (2.2)의 θ 는 알려진 상수이고 일반적으로 0.8의 근치값들로 정해지기 때문에 활성대조군에 θ 를 곱해도 분산에는 큰 변동 없다. 따라서 확률표본 X_{1j} ($j = 1, 2, \dots, n_1$)과 θX_{2j} ($j = 1, 2, \dots, n_2$)의 두 분포함수에 대한 동일분산 가정에 무리가 없기 때문에 본 논문에서는 Wilcoxon 순위합 검정을 이용하여 가설 (2.2)에 대한 검정을 시행한다. 검정통계량 W 는 통합순위에서 치료군의 순위합($W = \sum_{i=1}^{n_1} R_i$)이고, 표준화된 검정통계량

$$W^* = \frac{W - E_0(W)}{\sqrt{\text{var}_0(W)}} = \frac{W - \left\{ \frac{n_1(n_1 + n_2 + 1)}{2} \right\}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \tag{2.4}$$

은 $n_1, n_2 \rightarrow \infty$ 일 때, 점근적으로(asymptotically) 표준정규분포를 따른다. 따라서 이 검정의 기각역은 $W^* > z_{\alpha}$ 이며 z_{α} 는 표준정규분포의 상위 100α 백분위수이다.

3. 세 군 모형의 가설검정

세 군 모형은 두 군 모형과는 다르게 시험목적은 치료군이 위약군보다 효율적이며 위약군 대비 활성대조군 치료효과의 특정 비율을 유지하는지에 두고 있다. μ_1, μ_2, μ_3 을 각각 치료군, 활성대조군, 위약군의 모집단 평균이라 할 때 가설 (2.1)에서 비열등성 한계를 $\delta = \lambda(\mu_2 - \mu_3)$, $-1 < \lambda < 0$ 으로 표현한다. 먼저 $\mu_2 > \mu_3$ 임을 가정하고 $\theta = 1 + \lambda$ 로 정의하면 가설 (2.1)은

$$H_0 : \frac{\mu_1 - \mu_3}{\mu_2 - \mu_3} \leq \theta \quad \text{vs.} \quad H_1 : \frac{\mu_1 - \mu_3}{\mu_2 - \mu_3} > \theta, \quad \theta \in [1 + \lambda, 1)$$

$$H_0 : \mu_1 - \theta\mu_2 - (1 - \theta)\mu_3 \leq 0 \quad \text{vs.} \quad H_1 : \mu_1 - \theta\mu_2 - (1 - \theta)\mu_3 > 0, \quad \theta \in [1 + \lambda, 1) \tag{3.1}$$

으로 표현된다. $\theta = 0$ 인 경우 가설 (3.1)은 $H_0 : \mu_1 \leq \mu_3$ vs. $H_1 : \mu_1 > \mu_3$ 이므로 위약군에 대한 치료군의 효율성 검정을 내포하고 있음을 알 수 있다. 그리고 $\mu_2 > \mu_3$ 을 만족한다는 가정하에 세워진 가설이므로 가설 (3.1)의 검정 전에 우선적으로

$$\widetilde{H}_0 : \mu_2 \leq \mu_3 \quad \text{vs.} \quad \widetilde{H}_1 : \mu_2 > \mu_3 \tag{3.2}$$

의 검정이 필요하다.

본 논문에서는 활성대조의 위약에 대한 우위성 가정에 가설 (3.1)에 대한 검정으로 기존의 모수적 방법을 제시하고 새로운 비모수적 방법을 제안하였다.

3.1. Pigeot의 모수적 방법

확률표본 X_{ij} ($i = 1, 2, 3, j = 1, 2, \dots, n_i$)는 평균이 μ_i 이고 분산이 σ^2 인 정규분포에서 추출된 확률변수이다. 여기서 $i = 1, 2, 3$ 은 순서대로 치료군, 활성대조군, 위약군을 말한다. \bar{X}_i, S_i^2 ($i = 1, 2, 3$)을 각각 표본평균과 표본분산이라고 할 때 가설 (3.1)은 다음과 같이 선형대비를 이용하여 검정할 수 있다. 선형대비는

$$\phi(\mu) = \mu_1 - \theta\mu_2 - (1-\theta)\mu_3, \quad \mu = (\mu_1, \mu_2, \mu_3)$$

이며, 이 선형대비의 불편추정량은

$$\phi(\hat{\mu}) = \bar{X}_1 - \theta\bar{X}_2 - (1-\theta)\bar{X}_3, \quad \hat{\mu} = (\bar{X}_1, \bar{X}_2, \bar{X}_3)$$

이고, 추정된 분산은

$$\widehat{\text{var}}(\phi(\hat{\mu})) = \text{MSE} \left(\frac{1}{n_1} + \frac{\theta^2}{n_2} + \frac{(1-\theta)^2}{n_3} \right) = S_P^2 \left(\frac{1}{n_1} + \frac{\theta^2}{n_2} + \frac{(1-\theta)^2}{n_3} \right),$$

$$S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2 + (n_3-1)S_3^2}{n_1 + n_2 + n_3 - 3}$$

이므로, 검정통계량

$$T = \frac{\bar{X}_1 - \theta\bar{X}_2 - (1-\theta)\bar{X}_3}{S_P \sqrt{\frac{1}{n_1} + \frac{\theta^2}{n_2} + \frac{(1-\theta)^2}{n_3}}} \quad (3.3)$$

은 자유도가 $(n_1 + n_2 + n_3 - 3)$ 인 t 분포를 따른다. 이 검정의 기각역은 $T > t_\alpha(n_1 + n_2 + n_3 - 3)$ 이며 t_α 는 자유도가 $(n_1 + n_2 + n_3 - 3)$ 인 t 분포의 상위 100α 백분위수이다.

3.2. 제안된 비모수적 방법

확률표본 X_{ij} ($i = 1, 2, 3, j = 1, 2, \dots, n_i$)를 각각의 분포함수 $F_i(x) = F(x - \Delta_i)$ 에서 추출된 확률변수이다. Scheirer 등 (1976)이 Kruskal-Wallis 검정에서 선형대비로 확장하여 제안한 검정법을 각 군의 표본수가 다른 경우, 비열등성 검정에 적용할 수 있는 방법을 제안한다.

Kruskal-Wallis 검정통계량은 다음의 가정 하에 유도된다. R_1, R_2, R_3 은 각각 통합순위에서 치료군, 활성대조군, 위약군의 순위합이고 N 은 전체 표본수라 하자. 모든 $i = 1, 2, 3$ 에 대하여 $\lim_{N \rightarrow \infty} n_i/N = v_i$ ($0 < v_i < 1$)일 때, Wald-Wolfowitz (1944) 정리에 의하여

$$\frac{\{R_i - E(R_i)\}^2}{\text{var}(R_i)}, \quad i = 1, 2, 3$$

은 점근적으로 자유도가 1인 카이제곱분포를 따른다. 그러므로 R_i 가 서로 독립인 경우

$$H' = \sum_{i=1}^3 \frac{\{R_i - E(R_i)\}^2}{\text{var}(R_i)}$$

은 자유도가 3인 카이제곱분포를 따른다. 그러나 $\sum_{i=1}^3 R_i = N(N+1)/2$ 이 성립하기 때문에 R_i 는 실제로 독립일 수 없고 H' 통계량의 점근적 분포를 확신하기 어렵다. 따라서 Kruskal (1952)은 H' 통계량의 각 군에 대한 분산 $\text{var}(R_i)$ 에 보정항(correction factor) $N/(N - n_i)$ 을 곱한 통계량

$$H = \sum_{i=1}^3 \frac{N - n_i}{N} \frac{\{R_i - E(R_i)\}^2}{\text{var}(R_i)} = \frac{12}{N(N+1)} \left(\sum_{i=1}^3 \frac{R_i^2}{n_i} \right) - 3(N+1)$$

이 점근적으로 자유도가 2인 카이제곱분포를 따른다는 것을 보였다.

위의 과정은 각 군의 순위합의 평균 $\bar{R}_i = R_i/n_i$ ($i = 1, 2, 3$)에 대해서도 성립하므로 표준화된 선형대비는

$$C = \bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3$$

이고

$$G = \frac{C - E(C)}{\sqrt{\text{var}(C)}} = \frac{\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3 - E(\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3)}{\sqrt{\text{var}(\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3)}}$$

는 \bar{R}_i ($i = 1, 2, 3$)가 서로 독립일 때 점근적으로 표준정규분포를 따른다.

여기서, $H'_0 : \mu_1 = \mu_2 = \mu_3$ 가 참일 때, $E(\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3) = 0$ 이고 $\text{var}(C)$ 는

$$\begin{aligned} \text{var}(\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3) &= \text{var}(\bar{R}_1) + \theta^2 \text{var}(\bar{R}_2) + (1 - \theta)^2 \text{var}(\bar{R}_3) \\ &= \frac{(N - n_1)(N + 1)}{12n_1} + \frac{(N - n_2)(N + 1)}{12n_2} \theta^2 + \frac{(N - n_3)(N + 1)}{12n_3} (1 - \theta)^2 \end{aligned}$$

이다. 그러므로

$$G = \frac{\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3}{\sqrt{\frac{(N - n_1)(N + 1)}{12n_1} + \frac{(N - n_2)(N + 1)}{12n_2} \theta^2 + \frac{(N - n_3)(N + 1)}{12n_3} (1 - \theta)^2}}$$

이고

$$G^2 = \frac{\{\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3\}^2}{\frac{(N - n_1)(N + 1)}{12n_1} + \frac{(N - n_2)(N + 1)}{12n_2} \theta^2 + \frac{(N - n_3)(N + 1)}{12n_3} (1 - \theta)^2}$$

은 자유도가 1인 카이제곱분포를 따른다. 그러나 실제로 \bar{R}_i ($i = 1, 2, 3$)은 독립일 수 없으므로 각 분산항에 $N/(N - n_i)$ 을 곱하면

$$H^{**} = \frac{\{\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3\}^2}{\frac{N(N + 1)}{12} \left\{ \frac{1}{n_1} + \frac{\theta^2}{n_2} + \frac{(1 - \theta)^2}{n_3} \right\}}$$

이다. 따라서 가설 (3.1)을 검정하는 검정통계량은

$$H^* = \frac{\bar{R}_1 - \theta \bar{R}_2 - (1 - \theta) \bar{R}_3}{\sqrt{\frac{N(N + 1)}{12} \left\{ \frac{1}{n_1} + \frac{\theta^2}{n_2} + \frac{(1 - \theta)^2}{n_3} \right\}}} \quad (3.4)$$

이고 표준정규분포를 따른다. 이 검정의 기각역은 $H^* > z_\alpha$ 이며 z_α 는 표준정규분포의 상위 100α 백분위수이다.

표 4.1. 두 군 모형상에서 두 방법 사이의 검정력 비교($\alpha = 0.05, \theta = 0.8, \mu_2 = 4, N = 60$)

$n_1:n_2$	μ_1	정규 분포		이중지수 분포		코쉬 분포	
		T	Wilcoxon	T	Wilcoxon	T	Wilcoxon
1:1 (30, 30)	3.2	0.0550	0.0525	0.0515	0.0485	0.0336	0.0465
	3.3	0.1200	0.1185	0.0938	0.0981	0.0392	0.0765
	3.4	0.2240	0.2092	0.1552	0.1789	0.0468	0.1177
	3.5	0.3658	0.3463	0.2360	0.2861	0.0575	0.1693
	3.6	0.5273	0.4988	0.3437	0.4174	0.0675	0.2335
	3.7	0.6882	0.6575	0.4633	0.5603	0.0781	0.3105
	3.8	0.8206	0.7918	0.5769	0.6838	0.0903	0.3870
	3.9	0.9144	0.8960	0.6899	0.7958	0.1052	0.4759
	4.0	0.9640	0.9541	0.7815	0.8763	0.1198	0.5545
1:2 (20, 40)	3.2	0.0493	0.0521	0.0539	0.0545	0.0478	0.0508
	3.3	0.1138	0.1178	0.0880	0.1018	0.0539	0.0778
	3.4	0.2035	0.2036	0.1419	0.1763	0.0596	0.1180
	3.5	0.3225	0.3210	0.2202	0.2749	0.0689	0.1680
	3.6	0.4690	0.4657	0.3091	0.3891	0.0779	0.2275
	3.7	0.6230	0.6084	0.4136	0.5134	0.0890	0.2934
	3.8	0.7559	0.7382	0.5196	0.6316	0.0997	0.3686
	3.9	0.8532	0.8408	0.6215	0.7385	0.1126	0.4425
	4.0	0.9258	0.9152	0.7140	0.8277	0.1258	0.5162
2:1 (40, 20)	3.2	0.0551	0.0470	0.0538	0.0461	0.0390	0.0491
	3.3	0.1137	0.0988	0.0869	0.0873	0.0450	0.0766
	3.4	0.2144	0.1799	0.1509	0.1576	0.0540	0.1139
	3.5	0.3478	0.3025	0.2368	0.2596	0.0623	0.1607
	3.6	0.5090	0.4475	0.3397	0.3850	0.0731	0.2186
	3.7	0.6734	0.6104	0.4508	0.5194	0.0846	0.2874
	3.8	0.8097	0.7583	0.5722	0.6525	0.0978	0.3643
	3.9	0.8999	0.8660	0.6786	0.7642	0.1133	0.4407
	4.0	0.9559	0.9338	0.7710	0.8460	0.1266	0.5181

T: Hauschke의 모수적 방법, Wilcoxon: 비모수적 방법

4. 모의 실험의 계획 및 결과

두 군과 세 군의 설계상의 모수적 방법과 비모수적 방법에 의한 비열등성 검정이 본 논문에서 제시되었다. 각 설계상의 제시된 모수적 방법과 비모수적 방법간의 검정력을 비교하고 세 군 설계상에서 활성대조의 위약에 대한 우위성 검정을 포함한 2단계 검정시의 최종검정력 변화를 모수적 방법과 비모수적 방법에 대해서 비교하기 위하여 Monte Carlo 모의실험을 시행하였다.

이 모의실험에서 모집단의 분포로는 정규분포, 이중지수(Double exponential)분포, 코쉬(Cauchy)분포를 채택하고, SAS의 RANNOR함수를 이용하여 정규분포와 코쉬분포의 난수를 생성하였다. 이중지수 분포의 난수생성은 균일분포의 RANUNI함수의 역변화 기법을 이용하여 생성하였다. 그리고 유의수준 α 는 0.05로 하였으며 각 검정법들의 검정력을 비교하기 위하여 10,000번 반복 실험하였다.

두 군 설계에서는 전체 표본수 $N = 60$ 으로 하고 각 군의 표본수 비율을 1:1($n_1 = 30, n_2 = 30$), 1:2($n_1 = 20, n_2 = 40$), 2:1($n_1 = 40, n_2 = 20$)으로 3가지 경우를 고려하였다. $\mu_2 = 4, \theta = 0.8$ 로 정하였으며 귀무가설하의 치료군의 모평균은 $\mu_1 = 3.2$ 이고 μ_1 을 0.1 간격으로 증가시키며 검정력의 변화

표 4.2. 세 군 모형상에서 두 방법 사이의 검정력 비교($\alpha = 0.05, \theta = 0.8, \mu_2 = 4.5, \mu_3 = 3, N = 90$)

$n_1:n_2:n_3$	μ_1	정규 분포		이중지수 분포		코쉬 분포	
		T	KWLC	T	KWLC	T	KWLC
1:1:1 (30, 30, 30)	4.2	0.0550	0.0441	0.0528	0.0459	0.0408	0.0552
	4.3	0.1161	0.0949	0.0867	0.0837	0.0465	0.0831
	4.4	0.2165	0.1724	0.1542	0.1549	0.0528	0.1228
	4.5	0.3511	0.2865	0.2403	0.2516	0.0593	0.1714
	4.6	0.5187	0.4441	0.3302	0.3588	0.0647	0.2312
	4.7	0.6770	0.5994	0.4498	0.4964	0.0735	0.2954
	4.8	0.8110	0.7435	0.5672	0.6238	0.0823	0.3632
	4.9	0.9045	0.8533	0.6848	0.7372	0.0929	0.4448
	5.0	0.9599	0.9278	0.7757	0.8356	0.1047	0.5239
	2:1:1 (45, 23, 22)	4.2	0.0530	0.0443	0.0507	0.0478	0.0420
4.3		0.1133	0.0977	0.0928	0.0887	0.0477	0.0865
4.4		0.2165	0.1797	0.1501	0.1537	0.0524	0.1302
4.5		0.3752	0.3153	0.2461	0.2612	0.0616	0.1786
4.6		0.5516	0.4830	0.3589	0.3951	0.0664	0.2400
4.7		0.7126	0.6340	0.4746	0.5294	0.0778	0.3118
4.8		0.8443	0.7812	0.5914	0.6532	0.0898	0.3913
4.9		0.9243	0.8779	0.7148	0.7736	0.1001	0.4730
5.0		0.9706	0.9449	0.8085	0.8632	0.1035	0.5513
2:2:1 (36, 36, 18)		4.2	0.0529	0.0453	0.0545	0.0484	0.0339
	4.3	0.1204	0.0989	0.0942	0.0906	0.0412	0.0888
	4.4	0.2429	0.2008	0.1601	0.1650	0.0492	0.1331
	4.5	0.4017	0.3480	0.2654	0.2942	0.0567	0.1922
	4.6	0.5809	0.5145	0.3796	0.4233	0.0654	0.2607
	4.7	0.7459	0.6790	0.5024	0.5700	0.0739	0.3432
	4.8	0.8696	0.8196	0.6252	0.7036	0.0856	0.4278
	4.9	0.9437	0.9094	0.7371	0.8137	0.0955	0.5192
	5.0	0.9813	0.9668	0.8298	0.8954	0.1064	0.6049

T: Pigeot의 모수적 방법, KWLC: 비모수적 방법

를 비교하였다. 그 결과는 표 4.1에 제시하였다.

활성대조군의 우위성 가정하의 세 군 설계에서는 전체 표본수 $N = 90$ 으로 하고 각 군의 표본수 비율을 1:1:1 ($n_1 = 30, n_2 = 30, n_3 = 30$), 2:1:1 ($n_1 = 45, n_2 = 23, n_3 = 22$), 2:2:1 ($n_1 = 36, n_2 = 36, n_3 = 18$)으로 3가지 경우를 고려하였다. $\mu_2 = 4.5, \mu_3 = 3, \theta = 0.8$ 으로 정하였으며 귀무가설하의 치료군의 모평균은 $\mu_1 = 4.2$ 이고 μ_1 을 0.1 간격으로 증가시키며 검정력의 변화를 비교하였다. 그 결과는 표 4.2에 제시하였다.

활성대조군의 우위성 검정을 포함한 2단계 검정은 동일 표본수 $n_1 = n_2 = n_3 = 30$ 인 경우로 $\mu_1 = 4.5, \mu_3 = 3, \theta = 0.8$ 으로 정하고 μ_2 를 4.5에서 0.1씩 감소시키면서 검정력 변화를 비교하였다. 그 결과는 표 4.3에 제시하였다.

표 4.1에서 제시한 두 군 설계의 결과를 보면 정규분포에서는 $\mu_1 = 3.2$ 일 때 표본할당이 1:2인 경우를 제외하고 모수적 방법이 0.0550 정도로 0.05보다 약간 높은 성향을 보였으며 검정력 비교 결과 비모수적 방법에 비해 높은 검정력을 보였다. 이중지수분포에서는 할당 비율에 상관없이 $\mu_1 = 4.0$ 일 때 비모수적 방법이 모수적 방법 보다 약 0.1정도 높은 검정력을 보였다. 코쉬분포에서는 모수적 방법이 $\mu_1 = 3.2$ 일

표 4.3. 세 군 모형상에서 두 단계 검정 과정의 검정력 비교($\alpha = 0.05, \theta = 0.8, \mu_1 = 4.5, \mu_3 = 3, n_1 = n_2 = n_3 = 30$)

분포	모수				비모수		
	μ_2	μ_2 vs. μ_3 우월성	μ_1 vs. μ_2 비열등성	multiple	μ_2 vs. μ_3 우월성	μ_1 vs. μ_2 비열등성	multiple
정규	4.5	1.0000	0.3547	0.3547	1.0000	0.2903	0.2903
	4.4	1.0000	0.4792	0.4792	1.0000	0.4191	0.4191
	4.3	0.9999	0.6159	0.6158	0.9997	0.5549	0.5547
	4.2	0.9992	0.7343	0.7337	0.9986	0.6928	0.6918
	4.1	0.9963	0.8325	0.8294	0.9938	0.8004	0.7954
	4.0	0.9870	0.9012	0.8895	0.9829	0.8846	0.8695
	3.9	0.9686	0.9507	0.9209	0.9581	0.9369	0.8976
	3.8	0.9256	0.9781	0.9053	0.9082	0.9713	0.8821
	3.7	0.8539	0.9910	0.8462	0.8276	0.9881	0.8178
	3.6	0.7518	0.9966	0.7492	0.7205	0.9961	0.7177
	3.5	0.6180	0.9986	0.6171	0.5876	0.9981	0.5865
	3.4	0.4663	0.9998	0.4662	0.4369	0.9992	0.4366
이중지수	4.5	0.9904	0.2343	0.2321	0.9978	0.2485	0.2480
	4.4	0.9811	0.3163	0.3103	0.9953	0.3494	0.3478
	4.3	0.9661	0.4035	0.3898	0.9902	0.4618	0.4573
	4.2	0.9421	0.4937	0.4651	0.9765	0.5752	0.5617
	4.1	0.9171	0.5880	0.5393	0.9615	0.6872	0.6607
	4.0	0.8552	0.6813	0.5863	0.9262	0.7803	0.7227
	3.9	0.7952	0.7569	0.6919	0.8792	0.8553	0.7520
	3.8	0.7178	0.8276	0.5941	0.8081	0.9077	0.7335
	3.7	0.6274	0.8766	0.5450	0.7273	0.9461	0.6881
	3.6	0.5123	0.9155	0.4690	0.6133	0.9676	0.5934
	3.5	0.4108	0.9458	0.3885	0.4899	0.9824	0.4812
	3.4	0.3036	0.9671	0.2936	0.3615	0.9914	0.3584
코쉬	4.5	0.2209	0.0614	0.0136	0.8735	0.1760	0.1537
	4.4	0.2049	0.0683	0.0140	0.8377	0.2219	0.1859
	4.3	0.1889	0.0750	0.0142	0.7966	0.2746	0.2188
	4.2	0.1754	0.0826	0.0145	0.7426	0.3415	0.2536
	4.1	0.1619	0.0909	0.0147	0.6870	0.4123	0.2833
	4.0	0.1462	0.0989	0.0145	0.6257	0.4875	0.3050
	3.9	0.1306	0.1074	0.0140	0.5573	0.5613	0.3128
	3.8	0.1178	0.1169	0.0134	0.4851	0.6265	0.3039
	3.7	0.1043	0.1268	0.0132	0.4106	0.6873	0.2822
	3.6	0.0923	0.1368	0.0126	0.3385	0.7450	0.2522
	3.5	0.0792	0.1477	0.0117	0.2688	0.7949	0.2137
	3.4	0.0675	0.1577	0.0106	0.2095	0.8363	0.1752

때 표본 할당이 1:1, 2:1인 경우 각각 0.0336, 0.0390으로 유의수준 0.05를 만족하지 못한다는 것을 알 수 있다. 또한 검정력의 증가율이 매우 낮아서 $\mu_1 = 4.0$ 일 때 약 0.12, 0.52로 비모수적 방법과 현저히 차이가 난다. 표본 할당 비율에 따른 결과는 대부분 검정력이 1:1 > 2:1 > 1:2 순으로 나타났으며 코쉬분포에서 모수적 방법은 상승률이 불규칙적이고 비모수적 방법은 2:1, 1:2가 거의 같은 크기로 증가하였다.

표 4.2에서 제시한 세 군 설계의 결과를 보면 정규분포에서는 표본 할당이 1:1:1이고 $\mu_1 = 4.2$ 일 때 모수적 방법이 0.0550으로 약간 높았고 비모수적 방법이 0.0441로 약간 낮은 성향을 보였다. 검정력은 모수적 방법이 비모수적 방법보다 높게 나타났다. 이중지수분포에서는 $\mu_1 = 5.0$ 일 때 비모수적 방법이 모수적 방법 보다 약 0.06정도 높은 검정력을 보였다. 코쉬분포에서는 $\mu_1 = 4.2$ 일 때 할당 비율에 상관없이 모수적인 방법은 0.05보다 상당히 낮은 성향을 보였다. 특히 2:2:1일 때 0.0339로 유의수준 0.05를 만족하지 못한다는 것을 알 수 있다. 비모수적인 방법은 약 0.055로 조금씩 높게 나타났으며 검정력은 모수적 방법 보다 현저히 높게 나타났다. 표본 할당 비율에 따른 결과는 2:2:1 > 2:1:1 > 1:1:1 순으로 나타났으며 단지 코쉬 분포에서 모수적 방법은 상승률이 불규칙적이다.

표 4.3에서 제시한 2단계 검정에 의한 검정력 변화를 보면 코쉬분포에서 모수적 방법을 제외하고 모든 경우에 μ_2 가 3.8인 경우부터 감소하는 성향을 보이고 있으며 정규분포에 비해서 이중지수분포와 코쉬분포인 경우에 대조군과 위약군에 대한 우월성의 검정력이 낮으므로 치료군과 활성대조군에 대한 비열등성의 검정력과 최종 검정력의 차이가 더 크게 나타났다.

5. 결론 및 고찰

본 논문에서는 두 군과 세 군 설계상의 비모수적 방법을 제안하고 기존의 모수적 방법들과 Monte Carlo 모의시험을 통하여 비교해 보았다.

두 군 모형상의 비열등성 검정의 검정력은 각 군의 표본 할당비율이 동일할 때 가장 높았고 각 군의 표본 할당비율이 다른 경우에는 치료군이 활성대조군 보다 클 때 검정력이 높은 경향을 나타냈다. 또한 정규분포에서 모수적 방법과 비모수적 방법의 검정력은 유사하였으며 이중지수분포, 코쉬분포에서 비모수적 방법의 검정력이 높게 나타났다.

세 군 모형상의 비열등성 검정의 검정력은 치료군과 활성대조군의 표본 할당비율이 클수록 높은 경향을 보였으며 정규분포를 제외한 이중지수분포, 코쉬분포 경우 비모수적 방법의 검정력이 모수적 방법보다 높게 나타났다. 또한 2단계 검정시에는 활성대조군 보다 치료군의 평균이 높더라도 활성대조군의 위약군에 대한 우월성 검정의 검정력이 낮으면 결과적으로 최종검정력이 현저하게 떨어진 것을 볼 수 있었다.

따라서 본 논문에서 제시한 두 군 설계상의 비모수적 방법은 동일 표본수 일 때 검정력이 높으며 모수적 방법과 비교하였을 때 꼬리가 두꺼운 분포에서 검정력이 좋을 뿐만 아니라 일반적인 분포에서도 서로 유사한 결과를 보였다. 세 군 설계상의 비모수적 방법은 전체 표본에서 위약군의 표본수가 작을수록 검정력은 높아지며 그 중 치료군과 활성대조군이 동일 표본일 때 가장 좋은 검정력을 나타냈다. 이는 세 군이 동일 표본수인 경우와 같은 검정력을 목표로 할 때 필요한 전체 표본수가 작아지는 것을 나타낸다. 또한 모수적 방법과 비교하였을 때 꼬리가 두꺼운 분포에서 좋은 검정력을 나타냈다.

그러나 세 군 설계에서 제안한 검정방법은 각 군의 표본수가 큰 경우에만 적용할 수 있다는 단점이 있으며 분포에 따라서 유의수준 0.05보다 약간 낮거나 높은 결과를 보이므로 검정통계량의 근사과정에 보완이 필요할 것으로 보인다. 또한 2단계 검정과정을 시행하려는 경우에 활성대조약의 효과가 위약에 비해서 미약하고 새로운 치료약의 효과가 그에 비해 상당히 큰 경우에는 비열등성 검정 보다 우월성 검정이 더 적합할 것이다.

참고문헌

- D'Agostino, R. B., Massaro, J. M. and Sullivan, L. M. (2003). Non-inferiority trials: Design concepts and issues-the encounters of academic consultants in statistics, *Statistics in Medicine*, **22**, 169-186.

- Hauschke, D., Kieser, M., Diletti, E. and Burke, M. (1999). Sample size determination for proving equivalence based on the ratio of two means for normally distributed data, *Statistics in Medicine*, **18**, 93–105.
- Hung, H. M. J., Wang, S. J., Tsong, Y., Lawrence, J. and O'Neil, R. T. (2003). Some fundamental issues with non-inferiority testing in active controlled trials, *Statistics in Medicine*, **22**, 213–225.
- Hwang, I. K. and Morikawa, T. (1999). Design issues in noninferiority/equivalence trials, *Drug Information Journal*, **33**, 1205–1218.
- Kruskal, W. H. (1952). A nonparametric test for the several sample problem, *The Annals of Mathematical Statistics*, **23**, 525–540.
- Pigeot, I., Schäfer, J., Röhmel, J. and Hauschke, D. (2003). Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo, *Statistics in Medicine*, **22**, 883–899.
- Scheirer, C. J., Ray, W. S. and Hare, N. (1976). The analysis of ranked data derived from completely randomized factorial designs, *Biometrics*, **32**, 429–434.
- Wald, A. and Wolfowitz, J. (1944). Statistical tests based on permutations of the observations, *The Annals of Mathematical Statistics*, **15**, 358–372.
- Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics Bulletin*, **6**, 80–83.

Non-Inferiority Test in a Two-Arm Trial and a Three-Arm Trial Including a Placebo

Ji Sun Lee¹ · Dongjae Kim²

¹Dept. of Biostatistics, Catholic University; ²Dept. of Biostatistics, Catholic University

(Received September 2008; accepted October 2008)

Abstract

Two-arm non-inferiority trials is often applied to parametric procedure suggested by Hauschke *et al.* (1999). Since this design does not allow a direct comparison of a new treatment group with placebo group, parametric procedure in a three-arm non-inferiority trial with a placebo group was suggested by Pigeot *et al.* (2003). But, procedures in these designs are necessary for distribution assumptions. Therefore we propose, in this paper, nonparametric procedures employing Wilcoxon rank sum test in a two-arm design and linear contrast test suggested by Scheirer *et al.* (1976) in a three-arm design. The proposed nonparametric procedures and parametric procedures are compared by Monte Carlo simulation study.

Keywords: Non-inferiority, nonparametric, hypothesis testing.

¹Graduate Student, Dept. of Biostatistics, Catholic University, Seoul 137-701, Korea.

E-mail: yjs0229@hanmail.net

²Corresponding author: Professor, Dept. of Biostatistics, Catholic University, Seoul 137-701, Korea.

E-mail: djkim@catholic.ac.kr