

층화된 임상시험에서 생존분포의 비교를 위한 표본수의 결정*

김선우¹⁾ 박미라²⁾ 이재원³⁾

요약

생존시간을 비교하기 위한 임상시험에서는 생존시간과 관련된 위험인자들을 고려한 층화된 연구설계가 종종 요구된다. 이 경우 필요한 표본수의 결정은 층이 없는 경우에 비하여 다양한 연구상황과 복잡한 표본수의 산출 절차를 수반한다. 본 논문에서는 층이 있는 경우 생존시간을 비교할 때 필요한 표본수를 결정하는 방법들을 실례와 함께 설명하고, 연구자가 주어진 상황하에서 적절한 방법을 선택하는데 도움이 될 수 있도록, 다양한 상황을 설정하여 이에 대한 표본수를 비교하였다.

1. 서론

생존시간(survival time)을 비교하기 위한 임상시험에서는 사망, 병의 재발과 같은 어떤 특정한 사건이 발생할 때까지 걸리는 시간을 비교하게 된다. 특히 대부분의 임상시험에서는 환자들이 동시에 연구에 참여하기보다는 순차적으로 수집이 되며 환자수집기간이 종료된 후에도 환자를 계속 추적한다면가 또한 연구종료이전에 사전에 정의된 사건(event)이 발생하는 것이 아니라 다른 이유로 인해서 중도절단(censoring)이 발생하는 등의 자료 형태를 갖게 된다. 이러한 연구에서 필요한 표본수의 계산 방법은 Pasternack and Gilbert (1971), George and Desu (1974), Rubinstein, Gail and Santner (1981), Lachin (1981), Freedman (1982)에 의해 연구된 바 있으며, 박미라 등 (1998)은 이러한 방법들을 모의실험을 통해서 비교연구 하였다.

연령, 환자의 신체적 상태, 질병의 상태 등과 같은 위험인자에 대해 처리(treatment)를 균형에 맞게 할 필요가 있다거나, 이러한 위험인자에 대한 각 범주에 동일한 환자 수를 할당할 수 없는 경우 각 위험인자를 층(stratum)으로 고려하여 층화된 연구설계를 하게 되는데 이 경우에는 보다 복잡한 방법으로 표본수를 계산하게 된다. 예를 들어 한 임상의가 간세포암의 파열로 혈복강이 발생했을 때 기존의 보존적치료보다 경동맥색전술에 의한 치료가 생존율을 높이는지 알아보기 위한 연구를 한다고 하자. 이때 생존율이 치료방법이외에 간기능상태, 간암의 크기, 간암의 유형 및 간정맥의 침범여부 등의 다른 변수들과 관계되어 있을 수 있다면 이들 변수들을 고려하여 층화된 연구설계로 표본수를 계산하여야 할 것이다. Bernstein and Lagakos (1978), Schoenfeld (1983), Palta and Amini (1985), Lachin and Foulkes (1986) 등은 층이 있는 경우 생존시간을 비교할 때 필요한 표본수를 결정하는 방법을 제시한 바있다. 본 논문에서는 이러한 여러 방법들에 대해서 비교 설명하기로 한다. 또한 다양한 상황에서의 표본수를 비교하므로써 각각의 상황에 가장 적합한 표본수의 결정 방법을 제시하게 될 것이다.

* 이 연구는 삼성생명과학연구소 연구비(C-97-044)의 보조로 이루어졌음.

1) (135-230) 서울시 강남구 일원동 50, 삼성생명과학연구소 임상의학연구센터, 수석연구원
2) (301-112) 대전시 중구 용두 2동 143-5, 을지외과대학 의예과, 전임강사
3) (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과, 부교수

2. 표본수 결정을 위한 여러 방법들

표본수 결정방법들을 고찰함에 있어서 앞서 소개한 간세포암에서의 경동맥색전술 효과의 연구를 예제로 각 방법들을 설명하고자 한다. 본 논문에서는 고려된 총화변수들 중 간정맥 침범여부만을 총화변수로 이용하기로 한다. 대체적으로 간정맥을 침범한 경우가 1/3, 그렇지 않은 경우가 2/3이며 확률화를 통하여 같은 비율로 보존적치료군과 경동맥색전술군에 환자를 할당한다고 하자. 간정맥 침범여부와 관계없이 보존적치료군의 한달 생존확률을 20%로 가정하며 경동맥색전술을 했을 때 40%의 생존확률 효과가 있는지를 알아본다고 하자. 이때 5%의 유의수준, 80%의 검정력을 갖기 위해 필요한 총 환자수를 각 방법들의 설명과 함께 계산해 보기로 한다.

2.1. 모형, 가설 및 기호정의

먼저 우리가 고려하는 모형과 기호들을 정의하기로 하자. 연구에 들어온 각 환자들은 K 개의 층들중 하나로 분류되고 각 층안에서 확률화(randomization)를 통해 θ , $1 - \theta$ 의 비율로 시험군 E (experimental group) 또는 대조군 C (control group)로 각각 할당되어 치료를 받는다. 시험군과 대조군에서의 위험율(hazard rate)을 각각 λ_E , λ_C 로 정의한다.

단측검정을 고려했을 때 두 그룹에서의 위험율이 동일하다는 귀무가설과 시험군의 위험율이 대조군에서의 위험율보다 작다는 대립가설은 각각

$$H_0 : \lambda_E = \lambda_C \quad \text{vs.} \quad H_1 : \lambda_E < \lambda_C$$

와 같다. 두 그룹의 위험율의 비 $\Delta = \lambda_C / \lambda_E$ 로 정의하면 가설은

$$H_0 : \Delta = 1 \quad \text{vs.} \quad H_1 : \Delta > 1$$

와 같이 된다. 환자수집기간이 T , 추가추적기간을 τ 라 정의하고, Z_α 는 표준정규분포에서 확률 α 에 해당하는 표준정규분포의 임계치(upper 100 α -th percentile)를 의미한다. 이제 층이 있는 경우 시험군에서의 위험율이 대조군에서의 위험율보다 작은지 알아보기 위한 시험에서 요구되는 최소한의 표본수를 계산하는 방법에 대해 차례로 알아보기로 하자.

2.2. BERNSTEIN AND LAGAKOS (1978) 방법

환자들은 수집기간 $(0, T)$ 에 걸쳐 균일분포에 따라 연구에 들어오고 각 층에서의 생존시간은 지수분포를 따른다고 가정한다. 층 s 에서 대조군과 시험군의 생존시간의 위험율을 $\lambda_C^{(s)}$, $\lambda_E^{(s)}$ 라하고 두 위험율의 비가 모든 층에서 일정하다고 가정하자 (즉 $\Delta = \lambda_C^{(s)} / \lambda_E^{(s)}$, $s = 1, 2, \dots, K$).

총 연구기간중 층 s 에서 대조군 및 치료군의 환자가 사망할 확률을 각각 $\pi_C^{(s)}$ 와 $\pi_E^{(s)}$ 라할 때

$$\pi_i^{(s)} = 1 - \frac{e^{-\lambda_i^{(s)}\tau}(1 - e^{-\lambda_i^{(s)}T})}{\lambda_i^{(s)}T}, \quad i = C, E$$

와 같다. Bernstein and Lagakos는 유의수준 α 와 $1 - \beta$ 의 검정력을 갖기 위해 단위시간당 필요한 총 환자수 N^* 은

$$N^* = \frac{(\gamma^{-1/2}(1)Z_\alpha + \gamma^{-1/2}(\Delta)Z_{1-\beta})^2}{T\theta(1-\theta)(\ln \Delta)^2}$$

와 같이 계산됨을 보였다. 따라서 연구에 필요한 총 환자수 N 은 N^*T 와 같이 된다. 여기서 p_s 가 층 s 에 속하는 환자들의 비율일 때 $\gamma(\Delta)$ 은

$$\gamma(\Delta) = \sum \frac{p_s \pi_C^{(s)} \pi_E^{(s)}}{(1-\theta)\pi_C^{(s)} + \theta\pi_E^{(s)}}$$

와 같고, $\gamma(1)$ 은 $\pi_C^{(s)} = \pi_E^{(s)}$ ($s = 1, 2, \dots, K$)일 경우이다.

앞서 소개한 예제에서 생존시간이 지수분포를 따른다고 가정하면 층 s ($s=1,2$)의 대조군과 시험군의 위험율은 각각 $\lambda_C^{(s)} = -\ln(0.2) = 1.6094$ 와 $\lambda_E^{(s)} = -\ln(0.4) = 0.9163$ 으로 계산되어 위험율의 비 $\Delta = 1.6094/0.9163 = 1.7564$ 가 된다. 대조군과 시험군에 같은 수의 환자를 할당하고 ($\theta = 0.5$), 환자참여분포가 균일분포라고 가정하며, 6개월동안 환자를 수집하고 ($T = 6$), 2달동안의 추가추적기간을 갖는다고 하면 ($\tau = 2$), 모든 층에 대해 $\pi_C^{(s)}$ 및 $\pi_E^{(s)}$ 는 각각 0.9959와 0.9710으로 계산되며 p_1 과 p_2 는 각각 1/3과 2/3이므로 $\gamma(1)$, $\gamma(\Delta)$ 및 N^* 는 다음과 같이 계산된다.

$$\begin{aligned} \gamma(1) &= \frac{1/3 \times 0.9959 \times 0.9959}{0.5 \times 0.9959 + 0.5 \times 0.9959} + \frac{2/3 \times 0.9959 \times 0.9959}{0.5 \times 0.9959 + 0.5 \times 0.9959} = 0.9959 \\ \gamma(\Delta) &= \frac{1/3 \times 0.9959 \times 0.9710}{0.5 \times 0.9959 + 0.5 \times 0.9710} + \frac{2/3 \times 0.9959 \times 0.9710}{0.5 \times 0.9959 + 0.5 \times 0.9710} = 0.9833 \\ N^* &= \frac{(1.645/\sqrt{0.9959} + 0.841/\sqrt{0.9833})^2}{6 \times 0.5 \times (1 - 0.5) \times (\ln 1.7564)^2} = 13.0961 \end{aligned}$$

따라서 필요한 총 환자수는 N 은 $N^* \times T = 13.0961 \times 6 = 78.58$ 로 약 79명이 된다.

2.3. SCHOENFELD (1983) 방법

처리효과가 부분우도(partial likelihood) (Cox, 1972)에 근거한 방법으로 검정된다고 할 때 유의수준 α 와 $1 - \beta$ 의 검정력을 갖기 위해 필요한 총 사망자수는

$$(z_\alpha + z_{1-\beta})^2 / (\theta(1-\theta)(\ln \Delta)^2)$$

로 계산될 수 있다. 이 때 연구기간 중 사망할 환자들의 비율 d 를 보정해주면 필요한 총환자수 N 은

$$N = \frac{(Z_\alpha + Z_{1-\beta})^2}{\theta(1-\theta)(\ln \Delta)^2 d}$$

와 같이 계산된다. 이 공식은 지수분포를 따르는 두 그룹의 생존시간을 F-검정법으로 비교할 때 사용되는 공식과 같으며, 또한 비례위험율을 갖는 두 그룹의 생존시간을 로그순위검정법으로 비교할 때 사용되는 공식과 같다.

층화된 임상시험을 고려할 때 Schoenfeld는 비례위험모형하에서 환자들이 임상시험에 일정한 비율로 들어온다고 가정하고 생존할 환자들의 비율은 시점 τ 에서 $T + \tau$ 까지의 생존곡선의 평균으로 추정하였다. 즉 대조군의 생존곡선의 추정치를 $S_C(t)$ 라고 할 때 대조군에서 사망할 환자들의 비율 d_C 의 근사치를

$$d_C = 1 - 1/6\{S_C(\tau) + 4S_C(\tau + 0.5T) + S_C(\tau + T)\}$$

와 같이 추정하였다. 또한 환자들을 K 개의 층으로 나누고 각 층에서의 빈도를 p_1, p_2, \dots, p_K , 대조군의 생존곡선을 $S_C^{(1)}(t), \dots, S_C^{(K)}(t)$ 라 할 때 처리군에서 사망할 환자들의 비율 d_E 의 추정치를

$$d_E = 1 - 1/6 \sum p_s \{S_C^{(s)}(\tau)^{1/\Delta} + 4S_C^{(s)}(\tau + 0.5T)^{1/\Delta} + S_C^{(s)}(\tau + T)^{1/\Delta}\}$$

와 같이 제시하였다. 따라서 연구기간 중 사망할 환자들의 비율 d 는 $\theta d_E + (1 - \theta)d_C$ 와 같고 위의 공식에 의해 유의수준 α 와 $1 - \beta$ 의 검정력을 갖기 위해 필요한 총 환자수 N 을 구하게 된다.

앞서 소개한 연구에서 생존시간이 지수분포를 따른다고 가정할 때 2.2절에서 계산된 $\lambda_C^{(s)}$, $\lambda_E^{(s)}$ 와 Δ 로부터 환자수집기간이 6개월, 추가추적기간이 2개월이라할 때 사망할 환자들의 비율 d_C 와 d_E 는 각각 다음과 같이 계산된다.

$$\begin{aligned} d_C &= 1 - 1/6\{S_C(2) + 4S_C(2 + 0.5 \times 6) + S_C(2 + 6)\} = 0.9931, \\ d_E &= 1 - 1/6\{(1/3)[S_C^{(1)}(2)^{1/1.7564} + 4S_C^{(1)}(2 + 0.5 \times 6)^{1/1.7564} + S_C^{(1)}(2 + 6)^{1/1.7564}] \\ &\quad + (2/3)[S_C^{(1)}(2)^{1/1.7564} + 4S_C^{(1)}(2 + 0.5 \times 6)^{1/1.7564} + S_C^{(1)}(2 + 6)^{1/1.7564}]\} = 0.8666. \end{aligned}$$

따라서 연구기간중 사망할 환자들의 비율 d 와 필요한 총 표본수 N 은 각각 0.9298과 83.78로 계산되어 약 84명의 환자가 필요하게 된다.

2.4. PALTA AND AMINI (1985) 방법

Palta and Amini는 환자참여 및 생존분포에 대해 어떤 특정한 형태를 고려하지 않고 보다 일반적인 경우에도 적용 가능한 방법을 제시하였다. K 개의 층이 있는 경우 층화로 그순위검정법(stratified logrank test)를 사용하여 가설검정을 수행한다고 할 때, Palta and Amini는 Schoenfeld (1981)가 층이 하나인 경우 유도한 공식을 확장해서

$$(z_\alpha + z_{1-\beta})^2 / \mu^2$$

과 같은 일반적인 형태로 유도하였다. g_s 가 층 s 에서 시험군에 할당되는 환자들의 비율이고 p_s 는 층 s 의 환자비율이라 할 때, μ 는

$$\mu = \frac{\sum p_s \int_0^\infty \ln(\lambda_C^{(s)}(t)/\lambda_E^{(s)}(t)) v_s(t) g_s (1 - g_s) dt}{\sqrt{\sum p_s \int_0^\infty v_s(t) g_s (1 - g_s) dt}}$$

와 같다. 여기서 $v_s(t)$ 는 층의 하나인 경우 Schoenfeld (1981)가 정의한 함수 $v(t)$ 를 층 s 에 대해 정의한 것으로, $f_C^{(s)}$ 와 $f_E^{(s)}$ 를 각각 대조군과 시험군의 확률밀도함수라하고 $H_C^{(s)}$ 와 $H_E^{(s)}$ 를 각각 대조군과 시험군의 중도절단분포함수라 할 때

$$v_s(t) = (1 - g_s)f_C^{(s)}(1 - H_C^{(s)}) + g_sf_E^{(s)}(1 - H_E^{(s)})$$

와 같다. 두 그룹의 위험율의 비가 일정하고 모든 층에 대하여 같을 때 μ 는

$$\mu = (\ln \Delta) \sqrt{\sum p_s g_s (1 - g_s) V_s}$$

와 같이 간단한 형태가 된다. Palta and Amini는 그룹 i ($i = C, E$)의 층 s 에서 생존시간이 위험율 $\lambda_i^{(s)}$ 인 지수분포를 따르고 중도절단이 없으며 환자들이 균일분포에 따라 연구에 참여한다고 할 때, $v_s^{(t)}$ 를 연구기간에 대해 적분한 값 V_s 를

$$V_s = g_s [1 - 1/\lambda_E^{(s)} (e^{-\lambda_E^{(s)}(R-1)} - e^{-\lambda_E^{(s)}R})] + (1 - g_s) [1 - 1/\lambda_C^{(s)} (e^{-\lambda_C^{(s)}(R-1)} - e^{-\lambda_C^{(s)}R})]$$

와 같이 유도하였다. 여기서 R 은 환자수집기간에 대한 연구기간의 시간으로, 예를 들어 4년 동안 환자를 수집하고 1년 동안 추가추적기간이 있었다면 R 은 1.25년이 된다.

앞서 소개한 연구에서는 6개월의 환자수집기간과 2개월의 추가추적기간을 가정하므로 6개월을 단위시간으로 하면 총 연구기간 R 은 1.3333 단위시간이 된다. 생존시간이 지수분포를 따른다고 가정할 때 대조군과 시험군에서 단위시간 동안 생존할 확률은 각각 $e^{-6 \times 1.6094}$ 와 $e^{-6 \times 0.0163}$ 이 되어 위험율 $\lambda_C^{(s)}$ 와 $\lambda_E^{(s)}$ ($s = 1, 2$)은 각각 9.6564와 5.4978로 계산된다. 환자참여 분포가 균일분포라 하면 V_s 는

$$V_1 = 0.5 [1 - 1/5.4978 (e^{-5.4978 \times 0.3333} - e^{-5.4978 \times 1.3333})] + (1 - 0.5) [1 - 1/9.6564 (e^{-9.6564 \times 0.3333} - e^{-9.6564 \times 1.3333})] = 0.9835$$

와 같이 계산된다. 따라서 $g_1 = g_2 = 0.5$ 이므로 $V_2 = V_1$ 이고,

$\mu = (\ln 1.7564) \sqrt{(1/3)(0.5)(1 - 0.5)(0.9835) + (2/3)(0.5)(1 - 0.5)(0.9835)} = 0.2793$ 이므로 $N = 79.22$ 로 계산되어 약 80명의 환자가 필요하게 된다.

2.5. LACHIN AND FOULKES (1986) 방법

생존분포가 지수분포를 따른다는 가정하에서 두 그룹의 위험율이 같다는 가설은 최대 우도추정치 $\hat{\lambda}_E, \hat{\lambda}_C$ 및 이들의 근사적 분산 $\sigma^2(\hat{\lambda}_E), \sigma^2(\hat{\lambda}_C)$ 에 기초한 통계량을 사용하여 검정할 수 있다 (Gross and Clark, 1975). Lachin (1981)은 이러한 가설검정에서 필요한 총 표본수 N 과 검정력 $1 - \beta$ 간의 관계식이

$$\sqrt{N} |\lambda_E - \lambda_C| = Z_\alpha \sqrt{\psi(\bar{\lambda})(1/\theta + 1/(1 - \theta))} + Z_{1-\beta} \sqrt{\phi(\lambda_E)(1/\theta) + \phi(\lambda_C)(1/(1 - \theta))}$$

와 같다는 것을 보였다. 여기서 $\bar{\lambda} = \theta\lambda_E + (1 - \theta)\lambda_C$ 이고, $\Phi(\lambda)$ 는 $\sigma^2(\hat{\lambda}) = \Phi(\lambda)/N$ 을 만족하도록 구해진다. Lachin (1981)은 환자참여분포가 균일분포를 따르고 연구종료외에 다른 중도절단은 없다고 가정할 때 $\Phi(\lambda)$ 는

$$\Phi(\lambda) = \frac{\lambda^2}{1 - \frac{e^{-\lambda\tau} - e^{-\lambda(T+\tau)}}{\lambda T}}$$

와 같이 되어 총 표본수 N 을 쉽게 구할 수 있음을 보였다. 일반적으로 환자참여분포를 $g(z)$ ($0 \leq z \leq T$)라 할 때 이에 대한 $\Phi(\lambda)$ 은

$$\Phi(\lambda) = \frac{\lambda^2}{\int_{\tau}^{T+\tau} g(T + \tau - f)(1 - e^{-\lambda f})df}$$

와 같이 표현된다. Lachin and Foulkes는 연구기간중 중도절단이 있는 경우에도 함수 Φ 의 산출방법을 제시하였다. 중도절단이 임의로 일어나고 중도절단분포 $H(t)$ 가 생존분포와 독립일 때 $\Phi(\lambda)$ 는

$$\Phi = \frac{\lambda^2}{\int_0^T \int_0^{T+\tau-z} \lambda e^{-\lambda t} g(z)[1 - H(t)]dt dz}$$

와 같이 구할 수 있다. Lachin and Foulkes는 두 개의 층이 있는 경우 각 층에 대한 환자 수 집계기간과 추가추적기간을 각각 T_s, τ_s ($s = 1, 2$)라 하고 각 층에서 두 그룹의 위험율을 $\lambda_E^{(s)}, \lambda_C^{(s)}$ 라 할 때 검정에 필요한 총 표본수와 검정력과의 관계식을 유도하였는데 이를 두 개 이상의 층으로 일반화하면

$$\sqrt{N}|\bar{\lambda}_E - \bar{\lambda}_C| = Z_{\alpha}\sqrt{1/\Omega} + Z_{1-\beta}\sqrt{1/\Omega^2(\sum p_s\Psi_{1s}/\Psi_{0s}^2)}$$

와 같다. 여기서 p_s 는 s 번째 층의 비율이고 θ_s 는 층 s 에서 시험군에 할당되는 비율이라 할 때 $\bar{\lambda}_E, \bar{\lambda}_C, \Omega, \Psi_{0s}, \Psi_{1s}$ 및 필요한 공식은 다음과 같이 계산된다.

$$\begin{aligned} \bar{\lambda}_s &= \theta_s\lambda_E^{(s)} + (1 - \theta_s)\lambda_C^{(s)} \\ \Psi_{0s} &= \Phi_s(\bar{\lambda}_s)(1/\theta_s + 1/(1 - \theta_s)) \\ w_s &= p_s/\Psi_{0s}(\sum p_s/\Psi_{0s})^{-1} \\ \bar{\lambda}_E &= \sum w_s\lambda_E^{(s)} \\ \bar{\lambda}_C &= \sum w_s\lambda_C^{(s)} \\ \Psi_{1s} &= \Phi_s(\lambda_E^{(s)})/\theta_s + \Phi_s(\lambda_C^{(s)})/(1 - \theta_s) \\ \Omega &= \sum p_s/\Psi_{0s} \end{aligned}$$

앞서 소개한 연구에서 위의 공식들을 설명해 보기로 한다. 생존시간분포와 환자참여분포가 각각 지수분포와 균일분포라 하고 중도절단이 없다고 가정할 때 N 을 계산하기 위해

필요한 위의 공식들은 다음과 같이 계산된다.

$$\begin{aligned} \bar{\lambda}_1 &= \bar{\lambda}_2 = 0.5 \times 0.9163 + (1 - 0.5) \times 1.6094 = 1.2629 \\ \Phi_1(\bar{\lambda}_1) &= \Phi_2(\bar{\lambda}_2) = (1.2629)^2 \left[1 - \frac{e^{-1.2629 \times 2} - e^{-1.2629 \times 8}}{1.2629 \times 6} \right] = 1.6118 \\ \Psi_{01} &= \Psi_{02} = 1.6118 \times (1/0.5 + 1/(1 - 0.5)) = 6.4473 \\ \Omega &= (1/3)/(6.4473) + (2/3)/(6.4473) = 0.1551 \\ \Phi_1(\lambda_E^{(1)}) &= \Phi_2(\lambda_E^{(2)}) = (0.9163)^2 \left[1 - \frac{e^{-0.9163 \times 2} - e^{-0.9163 \times 8}}{0.9163 \times 6} \right] = 0.8646 \\ \Phi_1(\lambda_C^{(1)}) &= \Phi_2(\lambda_C^{(2)}) = (1.6094)^2 \left[1 - \frac{e^{-1.6094 \times 2} - e^{-1.6094 \times 8}}{1.6094 \times 6} \right] = 2.6011 \\ \Psi_{11} &= \Psi_{12} = 0.8646/0.5 + 2.6011/(1 - 0.5) = 6.9314 \\ w_1 &= (1/3)/6.4473 \times ((1/3)/6.4473 + (2/3)/6.4473)^{-1} = 0.3333 \\ w_2 &= (2/3)/6.4473 \times ((1/3)/6.4473 + (2/3)/6.4473)^{-1} = 0.6667 \\ \bar{\lambda}_E &= 0.3333 \times 0.9163 + 0.6667 \times 0.9163 = 0.9163 \\ \bar{\lambda}_C &= 0.3333 \times 1.6094 + 0.6667 \times 1.6094 = 1.6094 \end{aligned}$$

따라서 $N = 85.02$ 로 계산되어 약 86명의 환자가 필요하게 된다.

3. 여러 가지 방법들의 비교

3.1. 각 방법에서의 가정

위에서 소개한 네 가지 방법들을 비교하기 위해서, 먼저 각 방법의 가정을 정리하여 보자. 가장 일반적으로 사용 가능한 방법은 Palta and Amini 방법이다. 이 방법은 환자참여 분포나 생존시간에 관한 분포에 대한 어떤 특별한 가정이 필요하지 않고 각 층별 위험율의 비도 모두 같을 필요가 없기 때문이다. 이에 비하여 Bernstein and Lagakos 방법과 Lachin and Foulkes 방법은 생존시간의 분포로 지수분포를 가정하고 있으며 Bernstein and Lagakos 방법과 Schoenfeld 방법은 모든 층에 있어서 위험율의 비가 모두 같다고 할 수 있어야 사용할 수 있는 방법이다. 또한 환자참여분포에 대한 가정으로 Bernstein and Lagakos 방법은 균일분포를 가정하고 있다. 그러나 네 가지 방법 모두 각 그룹에 환자를 1:1로 할당할 필요는 없으며 전체 환자수에 대한 각 층에 속하는 환자의 비율도 모두 같을 필요는 없다. Bernstein and Lagakos 방법과 Lachin and Foulkes 방법은 생존시간이 지수분포를 따른다는 가정하에서 위험율의 비의 최대우도추정량을 이용한 근사적 모수적 검정에 근거한 것이고, Schoenfeld 방법은 부분우도추정법을 이용한 검정에 근거한 방법이다. 반면 Palta and Amini 방법은 층화로그순위검정법(stratified logrank test)에 근거하여 유도된 방법이다. 그러나 Palta and Amini 방법과 Lachin and Foulkes 방법은 연구종료이전에 중도절단이 발생하는 경우도 사용할 수 있는 방법이다. 표 3.1은 이러한 가정들을 정리한 것이다.

표 3.1: 각 방법별 가정

	Bernstein and Lagakos	Schoenfeld	Palta and Amini	Lachin and Foulkes
환자참여형태	균일분포	가정불필요	가정불필요	가정불필요
생존시간분포	지수분포	가정불필요	가정불필요	지수분포
할당비 (그룹)	일반적*	일반적*	일반적*	일반적*
할당비 (층)	일반적**	일반적**	일반적**	일반적**
층별 위험률 비	일정	일정	일반적***	일반적***
중도절단	없음	없음	있음	있음

* 그룹별 표본수를 임의의 비율로 할당 가능함

** 층별 표본수를 임의의 비율로 할당 가능함

*** 층별 위험률 비가 동일할 필요가 없음

3.2. 각 방법별 표본수의 비교

다양한 연구설계 상황하에서 각 방법에 따라 필요한 최소의 표본수를 비교하기 위하여 몇 가지 연구설계 상황을 설정하였다. 우선 생존시간의 분포를 지수분포로 가정하였으며, 각 그룹별 층별 위험률은 다음과 같이 산출하였다. 대조군의 첫 번째 층에 속하는 환자가 1년이상 생존할 확률을 PROB라 표시하고, 대조군에서 인접한 층간 위험률의 비를 B라 표시하자 (즉 $B = \lambda_C^{(s+1)} / \lambda_C^{(s)}$). PROB의 값을 먼저 설정하여 대조군의 첫 번째 층에 속하는 환자의 위험률을 산출한 후, B의 값을 설정하여 대조군의 다른 층들의 위험률을 산출하였다. 또한 중도절단을 없다고 가정하였다.

3개의 층이 있고 PROB을 0.5, B를 0.5로 설정하였을 때 네 가지 방법에 따른 표본수를 계산하였다. 이 때 두 그룹에 대한 표본할당비는 동일하게 하였고, 3개 층에 대한 표본비를

표 3.2: 필요한 총 표본수 ($K = 3, T = 2, \alpha = 0.05$, 단측검정)

Design	$\tau = 0$				$\tau = 0$			
	power=0.8		power=0.9		power=0.8		power=0.9	
	$\Delta = 1.5$	$\Delta = 2.0$	$\Delta = 1.5$	$\Delta = 2.0$	$\Delta = 1.5$	$\Delta = 2.0$	$\Delta = 1.5$	$\Delta = 2.0$
Bernstein and Lagakos	433.9	156.3	610.6	223.2	206.5	73.2	288.9	103.5
Palta and Amini	471.1	177.1	653.0	245.2	219.0	80.8	303.2	111.8
Schoenfeld	530.5	210.2	734.4	291.0	239.9	94.8	332.0	131.2
Lachin and Foulkes	521.8	204.4	723.0	283.6	241.9	92.8	335.5	129.2

도 모두 같다고 가정하였다. 두 그룹의 위험율의 비는 1.5와 2.0으로 하고 검정력이 0.8, 0.9일 때 필요한 최소의 표본수를 계산한 결과가 표 3.2에 있다. 표 3.2에서 볼 수 있듯이 이러한 경우 Bernstein and Lagakos 방법을 사용할 때가 항상 가장 적은 표본수가 필요하다. 추가추적기간이 없을 때 필요한 표본수는 Bernstein and Lagakos 방법, Palta and Amini 방법, Lachin and Foulkes 방법 그리고 Schoenfeld 방법의 순이다. 그러나 추가추적기간이 2년 인 경우 Lachin and Foulkes 방법과 Schoenfeld 방법의 사용 결과는 거의 차이가 없다.

3.3. 층의 수에 따른 표본수의 변화

층의 수가 많아질 때 각 방법에 따른 표본수를 조사하기 위하여 환자수집기간과 추가추적기간을 모두 2년으로 하고 층이 1개일 때부터 10개일 때까지 네 가지 방법에 따른 표본수를 계산하였다. 이 경우도 두 그룹에 대한 표본할당비를 같게 하고 층간 표본수의 차이에 따른 교란효과(confounding effect)에 영향을 받지 않게 하기 위하여 모든 층에 대한 표본비율도 모두 같게 하였다. 표 3.3에서 볼 수 있듯이 층이 없는 경우 ($K=1$) 보다 층이 있는 경우 더 많은 표본이 필요하지만 층이 있는 경우 층의 수가 늘어날수록 표본수의 차이는 거의 없으나 Lachin and Foulkes 방법에서는 점차 감소하는 경향과 함께 다른 방법에 비하여 표본수의 차이가 어느 정도 나타난다.

표 3.3: 층의 수에 따른 총 표본수 ($T=2, \tau=2, \text{PROB}=0.5, B=0.5, \Delta=1.5$)

K / Design	Bernstein and Lagakos	Palta and Amini	Schoenfeld	Lachin and Foulkes
1	178.7	187.3	205.8	191.8
2	207.2	219.5	238.2	251.5
3	206.5	219.0	239.9	241.9
4	206.3	218.9	240.4	238.7
5	206.2	218.8	240.7	237.1
6	206.1	218.7	240.8	236.2
7	206.1	218.7	240.9	235.5
8	206.1	218.7	241.0	235.1
9	206.0	218.7	241.1	234.7
10	206.0	218.6	241.1	234.5

3.4. 연구기간에 따른 표본수의 변화

표 3.4는 층이 세 개일때 환자수집기간 및 추가추적기간에 따른 표본수의 변화를 보여 준다. 이 경우도 두 그룹에 대한 표본할당비를 같게 하고 모든 층에 대한 표본비율도 모두 같게 하였다. 층화된 연구에서도 환자수집기간이 길수록 추가추적기간이 길수록 필요한 표

본수는 작아짐을 알 수 있다. 또한 연구기간이 일정할 때 상대적으로 추가추적기간이 긴 연구가 환자수집기간이 긴 연구에 비해 더 적은 수의 표본이 필요하다. 연구기간의 증가에 따른 표본수의 감소율은 네 가지 방법에서 모두 비슷하나, Schoenfeld 방법에서 약간 더 감소하였다. 예를 들어 환자수집기간이 1년일 때 추가추적기간이 없을 때는 Schoenfeld의 방법이 표본수가 가장 많이 필요하였으나 추가추적기간이 2년 이상이 되면 Lachin and Foulkes 방법에서 필요한 표본수가 더 많아진다.

표 3.4: 연구기간에 따른 총 표본수 ($K = 3$, $PROB = 0.5$, $B = 0.5$, $\Delta = 1.5$)

Design	T/τ	0	1	2	3	4	5
Bernstein and Lagakos	1	749.7	305.4	223.8	191.7	175.7	166.7
	2	433.9	258.3	206.5	183.4	171.1	163.9
	3	330.4	231.5	195.1	177.4	167.6	161.8
	4	279.8	214.4	187.1	173.0	165.0	160.1
	5	250.1	202.7	181.2	169.7	163.0	158.9
Palta and Amini	1	822.9	330.6	238.9	202.2	183.3	172.4
	2	471.7	277.4	219.0	192.3	177.7	168.9
	3	356.0	246.8	205.7	185.2	173.5	166.3
	4	299.1	227.1	196.2	179.9	170.2	164.1
	5	265.6	213.6	189.2	175.8	167.7	162.5
Schoenfeld	1	936.0	369.9	263.4	220.1	197.6	184.1
	2	530.5	307.7	239.9	208.2	190.6	179.7
	3	396.9	271.8	223.9	199.5	185.3	176.1
	4	331.2	248.7	212.5	193.0	181.1	173.3
	5	292.5	232.7	204.1	187.8	177.7	171.0
Lachin and Foulkes	1	912.1	365.4	263.8	223.3	202.8	191.0
	2	521.8	306.4	241.9	212.6	196.7	187.3
	3	393.5	272.6	227.3	204.9	192.2	184.5
	4	330.5	251.0	217.0	199.1	188.7	182.3
	5	293.5	236.2	209.4	194.8	186.1	180.6

3.5. 두 그룹에 대한 표본할당비에 따른 표본수의 변화

두 그룹에 다른 수의 표본을 할당할 경우 필요한 표본수를 알아보기 위하여 층을 세 개로 하고 모든 층에 대한 표본비율은 모두 같게 하였다. 환자수집기간과 추가추적기간을 2년으로 하였을 때 계산된 표본수는 표 3.5와 같다. 총화된 연구에서도 두 그룹에 같은 수의 표본을 할당할 경우가 가장 작은 표본수가 필요하고, 할당비가 1:1에서 벗어날수록 더 많은 수의 표본이 필요하다. 대조군에 비해 시험군에 상대적으로 매우 작은 수의 표본을 할당할

경우에는 ($\theta \leq 0.2$), Bernstein and Lagakos 방법보다 Palta and Amini 방법이나 Schoenfeld 방법에 의한 표본수가 더 작으며 그렇지 않은 경우에는 Bernstein and Lagakos 방법이 가장 작은 수의 표본을 산출한다. 시험군에 더 작은 수의 표본을 할당할 경우 Schoenfeld 방법이 다른 방법에 비해서 상대적으로 할당비가 1:1인 경우에 대해 표본증가율이 작으며 시험군에 더 많은 수의 표본을 할당할 경우는 Lachin and Foulkes 방법이 다른 방법에 비해서 상대적으로 할당비가 1:1인 경우에 대해 표본증가율이 작다. 그러나 시험군의 할당비율이 작

표 3.5: 두 그룹에 대한 표본할당비에 따른 표본수의 변화
($K = 3, T = 2, \tau = 2, \text{PROB} = 0.5, B = 0.5, \alpha = 1.5$)

θ / Design	Bernstein and Lagakos	Palta and Amini	Schoenfeld	Lachin and Foulkes
0.1	589.4	563.5	561.1	726.1
0.2	329.4	322.9	328.6	401.1
0.3	249.3	250.7	261.1	299.9
0.4	216.6	223.7	238.7	257.2
0.5	206.5	219.0	239.9	241.9
0.6	213.6	232.8	262.1	246.6
0.7	242.4	271.5	315.1	275.6
0.8	315.9	363.9	436.1	353.5
0.9	557.4	661.0	820.0	613.5

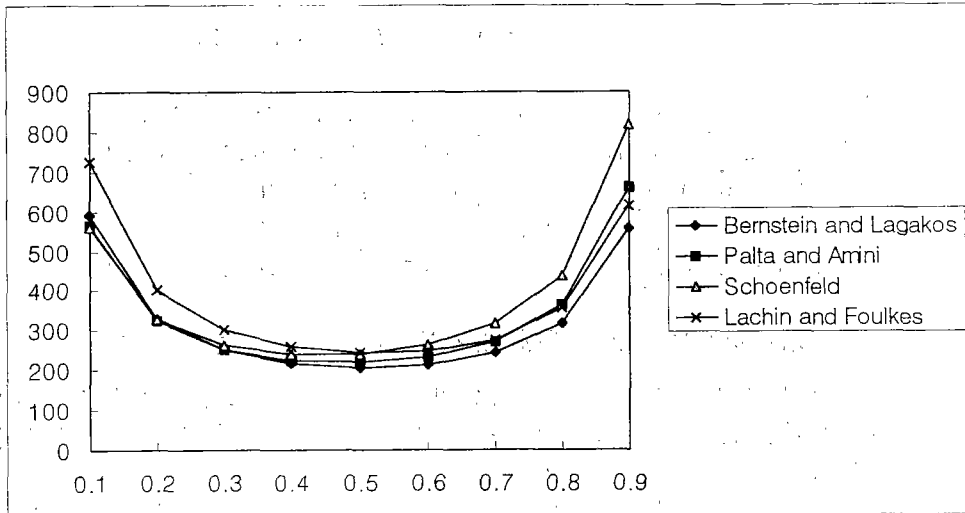


그림 3.1: 두 그룹에 대한 표본할당비에 따른 표본수의 변화

을 때는 Lachin and Foulkes 방법을 제외한 세 가지 방법에 의한 표본수는 비슷하게 되고 시험군의 할당비율이 클 때는 다른 세 가지 방법들에 비해 Schoenfeld의 방법에 의한 표본수가 상대적으로 크다. 그림 3.1은 이를 그래프로 그린 것이다.

3.6. 층에 대한 표본할당비율에 따른 표본수의 변화

이제 두 그룹에 같은 수의 표본을 할당하되 층별 표본수가 다른 경우에 필요한 표본수를 알아보기로 한다. 층이 세 개이고 환자수집기간 및 추가추적기간을 2년인 경우에 계산된 표본수는 표 3.6과 같다. 여기서 표본할당비는 첫 번째 층 : 두 번째 층 : 세 번째 층의 표본할당비이다. B 가 0.5인 경우는 표본이 첫 번째 층, 두 번째 층, 세 번째 층의 순서로 많이 할당됨에 따라 표본의 수는 더 많이 필요하게 되고, B 가 2.0인 경우는 그 반대이다. 따라서 위험율이 작은 층에 표본을 많이 할당할수록 더 많은 수의 표본이 필요하게 되고 반대로 위험율이 큰 층에 표본을 많이 할당하면 필요한 표본수는 작아진다.

표 3.6: 층별 표본할당비율에 따른 표본수의 변화

($K = 3, T = 2, \tau = 2, \text{PROB} = 0.5, B = 0.5, \Delta = 1.5$)

표본할당비	6:3:1	2:2:1	1:1:1	1:2:2	1:3:6
Bernstein and Lagakos	191.4	200.0	206.5	213.2	224.4
Palta and Amini	201.8	211.7	219.0	226.7	239.6
Schoenfeld	229.3	235.6	239.9	244.4	251.3
Lachin and Foulkes	219.0	231.7	241.9	246.7	256.8

($K = 3, T = 2, \tau = 2, \text{PROB} = 2.0, B = 0.5, \Delta = 1.5$)

표본할당비	6:3:1	2:2:1	1:1:1	1:2:2	1:3:6
Bernstein and Lagakos	171.2	167.2	165.0	162.5	159.5
Palta and Amini	177.9	173.0	170.2	167.5	163.4
Schoenfeld	187.8	183.6	181.1	178.5	174.9
Lachin and Foulkes	190.8	188.7	188.7	184.2	179.0

3.7. 생존확률 및 층간 위험률 비의 변화에 따른 표본수의 변화

대조군의 첫 번째 층에 속한 환자가 1년이상 생존할 확률 PROB의 변화와 층간 위험률의 비 B 의 변화에 따른 표본수의 변화를 알아보기로 한다. 층이 세 개 있으며 세 개 층에 할당되는 표본비율은 모두 같고, 두 그룹의 표본할당비를 같게 하였다. 환자수집기간 및 추가추적기간이 모두 2년일 때 필요한 표본수는 표 3.7과 같다. B 가 0.2일 때, PROB이 0.3인 경우에서 0.5로 증가하면 표본은 약 1.3배가 필요하나 0.7에서 0.9로 증가하면 약 2.7배의 표본이 필요하게 되어 PROB 값이 커질수록 표본의 수가 급격히 늘어난다는 것을 알 수

있다. 그 증가율은 네 가지 방법에서 모두 비슷하지만 B 값이 크면 그 증가율이 약간 줄어든다. B 값이 크면 필요한 표본수는 줄어드는데 그 감소율은 B 값이 커질수록 작아진다. Bernstein and Lagakos 방법, Palta and Amini 방법 및 Schoenfeld 방법의 B 값에 따른 감소율은 비슷한 반면 Lachin and Foulkes 방법에서는 다른 세 가지 방법보다 그 감소폭이 크다. 또한 PROB값이 크면 그 감소율은 커진다. 즉 PROB가 0.1일 때 B 가 0.2에서 0.5로 커지면 Lachin and Foulkes 방법에서는 약 32%의 표본이 덜 필요하고 0.5에서 0.8로 커지면 약 8%정도의 표본이 덜 필요하지만 다른 세 가지 방법에서는 B 가 0.2에서 0.5로 커질 때와 0.5에서 0.8로 커질 때 각각 10%와 2% 정도의 표본이 덜 필요하게 된다. 그리고 PROB가 0.9일 때 B 가 0.2에서 0.5로 커지면 Lachin and Foulkes 방법에서는 필요한 표본수가 약 46% 정도 줄어들고 0.5에서 0.8로 커지면 약 22%의 표본이 덜 필요하게 되는 반면, 다른 세 가지 방법에서는 B 가 0.2에서 0.5로 커질 때와 0.5에서 0.8로 커질 때 각각 25%와 17% 정도의 표본이 덜 필요하게 된다.

표 3.7: 생존확률 및 중간 위험율의 비의 변화에 따른 표본수의 변화 ($K = 3, T = 2, \tau = 2$)

Design	B / PROB	0.1	0.3	0.5	0.7	0.9
Bernstein and Lagakos	0.2	171.0	204.9	260.9	390.8	1046.4
	0.5	154.1	170.4	206.5	299.6	786.3
	0.8	151.5	160.5	186.1	259.0	655.0
Palta and Amini	0.2	176.0	214.9	278.3	423.8	1152.7
	0.5	156.2	176.8	219.0	324.1	865.7
	0.8	152.4	164.9	195.9	279.1	720.3
Schoenfeld	0.2	178.7	218.7	289.1	455.8	1296.6
	0.5	162.1	189.3	239.9	362.4	988.2
	0.8	156.6	176.4	216.0	314.6	825.7
Lachin and Foulkes	0.2	254.9	317.3	417.8	644.4	1775.0
	0.5	174.1	195.8	241.9	358.2	960.2
	0.8	159.4	171.0	202.1	287.6	744.2

4. 결론

환자를 대상으로 하는 임상시험에서는 적절한 표본수를 결정하는 것이 매우 중요하다. 지금까지 생존시간을 비교하기 위한 임상시험에서 충화된 연구설계를 할 경우 필요한 표본수를 계산하는 네 가지 방법들에 대하여 알아보았다.

가정별로 보면 네 가지 방법들의 가정이 모두 상이하다. Bernstein and Lagakos 방법은 환자참여분포, 생존시간분포 및 층별 위험율의 비에 대해 특정형태를 가정하고 있으며 다른 세 가지 방법들은 Bernstein and Lagakos 방법의 몇 개 또는 모든 가정들이 완화된 상태에서 표본수의 계산이 가능하다. 그러나 특히 가정이 완화될 경우에 사용 가능한 Palta and Amini 방법과 Lachin and Foulkes 방법에 의한 표본수 계산은 매우 복잡하게 된다. 그

러나 Palta and Amini 방법은 Schoenfeld 방법이나 Lachin and Foulkes 방법의 가정과 비교할 때 가장 완화된 가정을 하고 있으면서도 본 논문에서 설정된 상황의 대부분 경우에 더 작은 표본수를 산출한다.

대조군에서 인접한 층간 위험율의 비를 0.5로 할 때 층이 없는 경우보다는 층이 있는 경우 필요한 표본수는 커지지만 층의 수가 세 개 이상으로 늘어날 때 필요한 총 표본수는 큰 차이가 없다. 그러나 각 층의 위험율이 본 논문에서 설정된 설계와는 다른 방법으로 설정될 경우 다른 결과를 주므로 (Moussa, 1988), 층의 수에 따른 총 표본수의 변화는 각 층의 위험율의 분포에 따라 달라지게 된다. 연구기간에 따라 살펴보면, 층이 없는 경우와 마찬가지로 환자수집기간이 길수록, 추가추적기간이 길수록, 더 작은 수의 표본이 필요하며, 연구기간이 일정할 경우에는 추가추적기간이 길수록 더 작은 수의 표본이 필요하고, 연구기간에 따른 표본수의 감소율은 네 가지 방법에서 모두 비슷하다. 시험군과 대조군에 같은 비율로 표본을 할당할 수 없을 경우 시험군에 할당할 표본비율에 따른 각 방법의 표본수와 그 변화율은 차이가 있다. 또한 층에 대한 표본 할당비율이 같지 않은 경우에는, 위험율이 작은 층에 더 많은 표본을 할당하면 더 많은 표본이 필요하게 된다.

이상의 비교는 환자참여분포와 생존시간분포가 각각 균일분포와 지수분포이고 층별 위험율의 비가 모두 같을 때를 가정하고 계산된 총 표본수에 근거한 것이다. 이러한 가정을 만족하는 경우에는 대조군에 비하여 시험군에 상대적으로 매우 적은 표본을 할당하는 경우를 제외하고는 Bernstein and Lagakos 방법이 가장 작은 표본수를 산출할 수 있게 되지만 이 방법은 최대우도추정량을 이용한 근사적 모수적 검정을 요구한다. Palta and Amini 방법은 대부분의 상황에서 Bernstein and Lagakos 방법 다음으로 작은 표본수를 주며, 층화로그순위검정법을 이용하여 두 그룹간의 생존시간 차이를 검정할 수 있다. 그러나 환자참여분포와 생존분포가 각각 균일분포와 지수분포가 아닐 경우, 그리고 층별 위험율의 비가 다를 경우는 Bernstein and Lagakos 방법을 제외한 방법들간의 비교연구가 필요하다.

참고문헌

- [1] Bernstein, D. and Lagakos, S. W. (1978). Sample size and power determination for stratified trials, *Journal of Statistical Computing and Simulation*, 8, 65-73.
- [2] Cox, D. R. (1972). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society, Series B*, 34, 187-220.
- [3] Freedman, L. S. (1982). Tables of the number of patients required in clinical trials using the logrank test, *Statistics in Medicine*, 1, 121-129.
- [4] George, S. L. and Desu, M. M. (1974). Planning the size and duration of a clinical trial studying the time to some critical event, *Journal of Chronic Diseases*, 27, 15-24.
- [5] Lachin, J. M. (1981). Comparing sample size determination and power analysis for clinical trials, *Controlled Clinical Trials*, 2, 93-113.

- [6] Lachin, J. M. and Foulkes, M. A. (1986). Evaluation of sample size and power for analyses of survival with allowance for nonuniform entry, losses to follow-up, noncompliance, and stratification, *Biometrics*, 42, 507-519.
- [7] Moussa, M. A. A. (1988). Planning the size of survival time clinical trials with allowance for stratification, *Statistics in Medicine*, 7, 559-569.
- [8] Palta, M. and Amini, S. (1985). Consideration of covariates and stratification in sample size determination for survival time studies. *Journal of Chronic Diseases*, 38, 801-809.
- [9] Pasternack, B. S., Gilbert, H. S. (1971). Planning the duration of long-term survival time studies designed for accrual by cohorts, *Journal of Chronic Diseases*, 24, 681-700.
- [10] Rubinstein, L. V., Gail, M. H., and Santner, T. J. (1981). Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation, *Journal of Chronic Diseases*, 34, 469-479.
- [11] Schoenfeld, D. A. (1981). The asymptotic properties of nonparametric tests for comparing survival distributions, *Biometrika*, 68, 316-319.
- [12] Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model, *Biometrics*, 39, 499-503.
- [13] 박미라, 김선우, 이재원 (1998). 생존함수의 비교연구를 위한 표본수의 결정, <응용통계연구>, 11(2), 269-285.

[1998년 6월 접수, 1999년 1월 최종수정]

Sample Size Determination Comparing Survival Distributions in the Stratified Clinical Trials *

Seonwoo Kim ¹⁾ Mira Park ²⁾ Jae Won Lee ³⁾

ABSTRACT

Clinical trials for comparing survival distributions often require the stratified study designs considering the risk factors related to survival times. In this case, sample size determination is attended with the various study conditions and the complicated computing procedures. This paper explained several methods for sample size determination for comparing survival distributions with the real case as an example, and compared the sample size in the various study designs in order to provide the researcher guidelines given his own study condition.

* This work was supported by the Samsung grant, #SBRI C-97-044.

1) Center for Clinical Research, Samsung Biomedical Research Institute

2) Department of Pre-Medicine, Eulji Medical College

3) Department of Statistics, Korea University