

생존함수의 비교연구를 위한 표본수의 결정*

박미라¹⁾ 김선우²⁾ 이재원³⁾

요약

임상시험연구의 설계에서 적절한 표본수의 결정은 매우 중요한 문제 중의 하나이다. 본 논문에서는 생존분포를 비교하기 위한 여러 가지 방법들을 소개하고, 각 방법에서의 가정들을 고찰하였다. 또한 다양한 상황에서의 표본수와 검정력 등을 비교제시하고 모의실험을 통해 각 방법들의 이론상의 검정력과 실제 검정력을 알아보았다. 그 결과로서 의학연구자들이 처한 여러 상황에 적합한 표본수의 결정방법을 제시하였다.

1. 서론

적절한 표본수의 결정은 임상시험연구의 설계에 있어서 매우 중요한 문제 중의 하나이다. 언급된 정도의 차이를 찾는 데 있어서 일정한 수준의 검정력을 지키는 것이 요구되는데, 사전에 적절한 표본의 수를 고려하지 못한 임상시험 중에서는 실제로 임상적으로 의미 있는 차이가 있더라도 검정력이 작아서 이를 찾아내지 못하는 경우가 종종 발생한다 (Friedman et al., 1978). 최근 들어 많은 의학저널에서 주어진 차이를 찾기 위한 표본수에 대한 정보를 논문에 제시할 것을 요구하고 있다.

생존시간(survival time)의 비교를 위한 임상시험에서는 통상적인 실험과는 다른 형태의 문제에 접하게 된다. 생존시간이란 어떤 정의된 시점부터 실패(failure)가 발생할 때까지 걸리는 시간으로 정의되는데, 여기서 실패는 환자의 사망 뿐 아니라 병의 재발, 또는 특정한 반응 등으로 정의될 수 있다. 임상시험에서 환자들은 모두 같은 시간에 임상시험에 참여하는 것이 아니고 순차적으로 임상시험에 참여하게 되며, 일부 환자들에 대한 관측치는 분석 시점 이전에 실패(failure)가 관측되지 않음으로써 중도절단(censored)된 자료의 형태를 갖게 된다. 이 밖에 다기관 공동시험(multicenter clinical trial)에서 병원과 같은 층(strata)을 고려할 필요가 있다거나, 대조군(control group)과 시험군(experimental group)에 동일한 환자수를 할당할 수 없는 경우에는 문제가 더욱 복잡하게 된다.

지금까지 여러 연구자들이 생존시간의 비교시 필요한 표본수를 결정하는 방법들을 제시한 바 있다. Pasternack and Gilbert(1971)는 코호트연구(cohort study)에서 생존시간의 중앙값(median survival time)을 비교하는데 필요한 연구기간과 표본수를 계산한 바 있다. George and Desu(1974)는 정규근사를 이용하여 이를 일반화한 방법을 제시하였다. 또한 Freedman(1982)은 두 그룹의 전반적인 분포비교를 위하여 로그순위검정(logrank test)을 이용할 때 필요한 표본의 크기를 유도하였다. Lachin(1981)과 Schuster(1990)는 두 그룹에

* 이 연구는 삼성전자 부설 삼성생명과학연구소 연구비(C-96-020)의 보조로 수행되었음.

1) (301-112) 대전시 중구 용두 2동 143-5, 을지의과대학 의예과, 전임강사

2) (135-230) 서울시 강남구 일원동 50, 삼성 생명과학연구소 임상의학연구센터, 수석연구원

3) (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과, 부교수

대한 환자의 할당비율이 서로 다른 경우에 대한 표본수의 결정방법을 제시하였으며, Rubinstein et al.(1981)은 보다 현실적으로 연구종료이외의 중도절단과 추가추적기간(additional follow-up period)을 고려하여 환자수를 계산하는 방법을 제안하였다. 최근에 Lakatos and Lan(1992)는 Rubinstein et al.(1981)의 방법과 Freedman(1982)의 방법, 그리고 Lakatos (1988)의 방법의 장단점을 비교한 바 있다. 이외에도 여러 가지 상황에 따라 대상환자수를 계산하는 방법이 다양하게 개발되어 있다 (Donner, 1984; Moussa, 1988; Shuster, 1990; Hsieh, 1992).

본 논문에서는 이러한 여러 방법들에 대해서 설명하고 이들 방법에서의 가정들을 고찰하기로 한다. 또한 다양한 상황에서의 표본수와 검정력 등을 비교제시하고 모의실험을 통해 각 방법들의 이론상의 검정력과 실제로 검정을 실시한 후에 관측된 검정력을 알아보기로 한다. 그 결과로서 의학연구자들이 처한 다양한 상황에 가장 적합한 표본수의 결정방법을 제시하게 될 것이다.

2. 표본수 결정을 위한 여러 방법들

2.1. 모형, 가설 및 기호정의

필요한 표본수를 계산하는 여러 가지 방법들에 대해 고찰하기 위하여 우선 우리가 고려하는 모형과 기호들을 정의하기로 하자.

먼저 연구에 들어온 환자들은 확률화(randomization)를 통해 시험군 E(experimental group) 또는 대조군 C(control group)로 할당되어 치료를 받는다. 시험군과 대조군에서의 위험율(hazard rate)을 각각 λ_E, λ_C 로 정의하자. 생존시간이 지수분포(exponential distribution)를 따른다고 할 때에는 각 그룹에서의 평균 생존시간(mean survival time)이

$$\mu_C = \lambda_C^{-1}, \quad \mu_E = \lambda_E^{-1}$$

이 되고 생존시간의 중앙값(median survival time)은 각각

$$M_C = \log 2 \times \lambda_C^{-1}, \quad M_E = \log 2 \times \lambda_E^{-1}$$

이 된다.

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

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

두 그룹의 위험율의 비 $\delta = \lambda_C / \lambda_E$ 로 정의하면 가설은 다음과 같이 된다.

$$H_0 : \delta = 1 \quad vs. \quad H_1 : \delta > 1$$

환자수집기간을 T, 추가추적기간을 τ 라고 하고 대조군과 시험군에서의 중도절단율(loss-to-follow-up rate)을 각각 ϕ_C, ϕ_E 라고 정의하자. 또한 Z_α 는 표준정규분포에서 확률 α 에 해당하

는 표준정규분포의 임계치(upper $100\alpha - th$ percentile)를 의미한다. 예컨대 $Z_{0.05} = 1.645$ 이다.

이제 시험군에서의 위험율이 대조군에서의 위험율보다 유의적으로 작는지 알아보기 위한 시험에서 요구되는 최소한의 표본수를 계산하는 방법에 대해 차례로 알아보기로 하자.

2.2. PASTERNAK AND GILBERT(1971)의 방법

환자들이 연구에 들어오는 시점이 균일분포(uniform distribution)를 따르며, 각 환자들은 두 그룹에 확률화를 통해 동일한 비율로 할당된다고 하자. 또한 생존시간은 각각 위험율이 λ_C, λ_E 인 지수분포를 따른다고 하자. Pasternack and Gilbert(1971)는 모든 환자를 사망할 때까지 관측하고 생존시간의 중앙값의 차이가 정규분포를 따른다는 가정 하에서 근사적으로 필요한 사망자수를 계산하였다. 유의수준 α 와 $1 - \beta$ 의 검정력을 갖기 위해 각 그룹당 필요한 최소한의 사망자수 d 는

$$d = (M_E - M_C)^{-2} \{z_\alpha \sqrt{(M_C + M_E)^2/2} + z_\beta \sqrt{M_C^2 + M_E^2}\}^2$$

또는

$$d = (\delta - 1)^{-2} \{z_\alpha \sqrt{(\delta + 1)^2/2} + z_\beta \sqrt{\delta^2 + 1}\}^2$$

와 같다. 앞서 정의한대로 M_E 와 M_C 는 각각 시험군과 대조군에서의 생존시간의 중앙값(median survival time)을 말하고, δ 는 두 그룹의 위험율의 비($\delta = \lambda_C/\lambda_E$)이다.

2.3. GEORGE AND DESU (1974)의 방법

Pasternack and Gilbert(1971)의 근사적 방법에 의한 표본수는 정확한 분포에서 유도하는 것보다 항상 큰 값을 계산하게 된다. George and Desu(1974)는 지수분포를 따르는 변량의 합을 로그 변환한 것이 근사적으로 정규분포를 따른다는 것을 이용하여 이러한 점을 보완하고 보다 간편하게 쓸 수 있는 방법을 개발하였다.

여기서도 각 환자들은 두 그룹에 동일한 수로 할당되며, 생존시간은 각각 위험율이 λ_C, λ_E 인 지수분포를 따른다고 가정한다. 단, 여기서 환자들은 균일분포로 들어오지 않고, 포아송과정(Poisson process)으로 연구에 들어오게 된다고 가정한다. 이제 각 그룹당 필요한 총 사망자수 d 는 다음과 같이 구해진다.

$$d = \frac{2(Z_\alpha + Z_\beta)^2}{(\ln \delta)^2}$$

2.4. RUBINSTEIN, GAIL AND SANTNER (1981)의 방법

앞에서의 방법들은 각 그룹의 모든 환자들에게서 실패(failure)가 관측된다는 것을 가정한 방법이다. 그러나 실제로는 연구의 종료나 또는 연구 도중 여러 가지 사유로 인해 중도절단되는 사례가 많이 있다. Rubinstein et al.(1981)은 보다 현실적으로 연구종료와 추적불가(loss to follow up)등의 중도절단사유가 있고, 추가추적기간(additional follow-up

period)을 고려했을 때 필요한 환자수를 계산하는 방법을 제안하였다. 이 때 환자들은 포아송 과정으로 연구에 들어오고, 각 환자들은 두 그룹에 확률화를 통해 동일한 비율로 할당된다고 가정한다. 각 그룹에서의 생존시간이 각각 위험율이 λ_C, λ_E 인 지수분포를 독립적으로 따르며 중도절단시간 역시 독립적으로 지수분포를 따르고 중도절단을 각각 ϕ_C, ϕ_E 라고 하자.

환자수집기간이 T 년, 추가추적기간이 τ 년일 때 환자가 포아송과정에 따라 연구에 들어온다고 가정하면 한 해에 필요한 총 환자수는

$$N_T = \left(\frac{Z_\alpha + Z_\beta}{\log \delta} \right)^2 \sum_{i=1}^2 \frac{2\lambda_i^*}{T\lambda_i} \left\{ 1 - \frac{e^{-\tau\lambda_i^*(1-e^{-T\lambda_i^*})}}{T\lambda_i^*} \right\}^{-1}$$

과 같이 구해진다. 여기서 $\lambda_i^* = \lambda_i + \phi_i$ 이다. 따라서 각 그룹당 필요한 총 환자수는

$$N = N_T \times T/2$$

가 된다.

2.5. LACHIN (1981)의 방법

환자들이 두 그룹에 확률 q_C 와 q_E 로 할당되고 ($q_C + q_E = 1$), 각 그룹에서 생존시간이 지수분포를 따른다고 하자. Lachin (1981)은 두 그룹에서 필요한 총 환자수를 다음과 같이 제시하였다.

$$TN = \frac{(Z_\alpha + Z_\beta)^2}{(\lambda_C - \lambda_E)^2} \left(\frac{\Phi(\lambda_C)}{q_C} + \frac{\Phi(\lambda_E)}{q_E} \right)$$

동일한 비율로 할당($q_C = q_E = 1/2$)했을 때, 각 그룹당 필요한 총 환자수는 다음과 같다.

$$N = \frac{(Z_\alpha + Z_\beta)^2}{(\lambda_C - \lambda_E)^2} (\Phi(\lambda_C) + \Phi(\lambda_E))$$

여기서 $\Phi(\lambda_i)$ 는 다음의 세 가지 경우에 따라 다르게 된다.

(i) 환자들이 연구에 균일분포로 들어오고 ($U(0, T)$), 추가추적기간이 없는 경우 :

$$\Phi(\lambda_i) = \lambda_i^3 T / (\lambda_i T - 1 + e^{-\lambda_i T}), \quad i = C, E.$$

(ii) 모든 환자들이 동시에 연구에 들어오거나, 환자들을 연구시작 후 각각 T 시간씩 관측하는 경우 :

$$\Phi(\lambda_i) = \lambda_i^2 / (1 - e^{-\lambda_i T}), \quad i = C, E.$$

(iii) 환자들이 연구에 균일분포로 들어오고, 추가추적기간(τ)이 있는 경우 :

$$\Phi(\lambda_i) = \lambda_i^2 \left[1 - \frac{e^{-\lambda_i \tau} - e^{-\lambda_i (T+\tau)}}{\lambda_i T} \right]^{-1}, \quad i = C, E.$$

가 된다.

Lachin(1981)은 환자수가 고정되어 있을 때 1년간의 환자참여율(entry rate)이 작고 연구기간이 긴 경우보다는 환자들이 일찍 연구에 참여해 오랜 시간 동안 관찰하는 것이 낫다고 보인 바 있다.

2.6. FREEDMAN (1982)의 방법

Freedman(1982)은 생존시간 분포를 가정하지 않는 경우에 로그-순위검정(log rank test)을 하기 위한 표본수의 결정방법을 제시하였다. 이 때 환자들은 포아송과정으로 임상시험연구에 들어온 것을 가정하며, 확률화를 통해 두 그룹에 할당된다. 여기에서는 마지막 환자가 연구에 들어온 시점으로부터 T^* 시간 후에 분석하며 이후의 데이터는 중도절단된다. 환자들을 두 그룹에 1:1로 할당하는 경우에 각 그룹당 필요한 총 사망자수는

$$d = \frac{(Z_\alpha + Z_\beta)^2(1 + \delta)^2}{2(1 - \delta)^2}$$

이 된다. 각 그룹당 필요한 총 환자수는 다음과 같이 계산할 수 있다.

$$N = \frac{2d}{(2 - P_C - P_E)}$$

여기서 P_i 는 시간 T^* 에서의 생존율(survival proportion)이다. 만약 생존시간이 지수분포를 따른다면 $P_i = \exp(-\lambda_i T^*)$, $i = C, E$ 이 된다.

환자들을 두 그룹에 $r : 1$ 로 할당하는 경우에는 두 그룹에 필요한 총 사망자수는

$$TD = \frac{(Z_\alpha + Z_\beta)^2(1 + \delta r)^2}{r(1 - \delta)^2}$$

이 되며, 두 그룹에서 필요한 총 환자수는 다음과 같이 계산된다.

$$TN = \frac{2d(1 + r)}{r(1 - P_C) + (1 - P_E)}$$

이 방법은 시간 T^* 이후에 중도절단되는 것을 가정하므로 필요한 사망자수를 과대 추정하는 경향이 있다. 또한 필요한 환자수는 환자참여율이나 실패율과는 독립이고 T^* 이후의 두 그룹의 생존율에만 의존하게 된다.

3. 여러 가지 방법들의 비교

3.1. 각 방법에서의 가정

지금까지 표본수를 계산하는 다섯 가지 방법을 소개하였는데, 가장 중요한 것은 어느 상황에 어느 공식을 사용하는 것이 적당한 것인가 또는 쓰지 말아야 할 것인가를 결정하는 것이다. 이를 위해서 먼저 각 방법들이 가지고 있는 가정을 살펴보는 것이 중요하다. Pasternack-Gilbert 방법과 George-Desu 방법은 필요한 환자수가 아닌 사망자수를 산출하

도록 되어 있다. 생존시간은 두 방법 모두 지수분포를 이용하며, 환자를 두 그룹에 같은 수를 할당한 경우에만 계산할 수가 있다. 환자들이 연구에 참여하는 데에 있어서 Pasternack-Gilbert 방법이 균일분포(uniform distribution)를 가정하는데 반해, George-Desu 방법은 포아송 분포(Poisson distribution)를 가정한다.

Rubinstein-Gail-Santner 방법과 Lachin 방법은 중도절단을 고려하여 이 두 방법들을 확장한 것으로, 연구에 참여한 환자수를 계산한다. 생존시간은 역시 두 방법 모두 지수분포를 이용하며 추가추적기간을 갖는다. 차이점은 Lachin 방법에서는 환자가 연구에 균일분포 형태로 참여하고 연구의 종료에 유일한 중도절단 사유로 생각하는 반면, Rubinstein-Gail-Santner 방법은 환자가 연구에 포아송분포 형태로 참여하고, 연구를 종료해서 중도절단이 발생할 뿐만 아니라 별도로 중도절단이 지수분포 형태를 가지고 발생한다는 것이다. 또한 Rubinstein-Gail-Santner 방법에서는 환자를 1 : 1로 할당한 경우에만 계산할 수가 있으나, Lachin 방법에서는 환자를 $r : 1$ 로 할당한 경우에도 계산할 수가 있다. 위의 4가지 방법은 생존시간이 지수분포를 따른다는 가정 하에서 위험율비(hazard ratio)의 최대우도추정량을 이용한 모수적 검정에 근거한 것이다. 반면 Freedman 방법은 로그순위검정에 근거해서 사망자수를 산출한 것이다. 여기서 환자는 연구에 포아송분포 형태로 참여하고 추적기간을 거친 후, 연구의 종료로 인해 중도절단이 발생한다. Freedman 방법은 Lachin 방법과 같이 환자수를 $r : 1$ 로 할당한 경우에도 계산할 수가 있다. 지금까지 설명한 각 방법에 따른 가정들을 정리하면 표 3.1과 같다.

표 3.1: 각 방법별 가정

방법	표본수	환자참여분포	중도절단의 형태	생존분포	할당비
Pasternack-Gilbert	사망자수	균일분포	없음	지수분포	1:1
George-Desu	사망자수	포아송분포	없음	지수분포	1:1
Freedman	사망자수	포아송분포	연구종료	분포무관	$r:1$
Rubinstein-Gail-Santner	참여환자수	포아송분포	지수분포+연구종료	지수분포	1:1
Lachin	참여환자수	균일분포	연구종료	지수분포	$r:1$

3.2. 각 방법별 표본수의 비교

각 방법에 따른 필요한 최소의 표본수를 비교하여 보자. 이를 위하여 추적기간은 없고 ($\tau = 0$), 2년간의 환자 수집기간 ($T = 2$)만을 가정하였다. 이 때 Lachin 방법과 Freedman 방법에서 두 그룹에 대한 표본할당비는 동일하게 하였다. 두 환자군의 위험율의 비(δ)는 1.5와 2.0으로 하고, 검정력(power)이 0.8, 0.9일 때 필요한 최소 표본의 수를 계산한 결과가 표 3.2이다. 표 3.2에서 볼 수 있는 바와 같이 George-Desu 방법은 Pasternack-Gilbert 방법보다 더 정확한 근사적 방법으로 더 적은 크기의 표본수를 준다. Freedman 방법은 분석시점 이후의 정보가 무시되므로, 환자 참여분포가 같은 George-Desu 방법보다 표본수가 과대 추정됨을 알 수 있다. 그러나 표에서와 같이 지수분포를 따르는 경우에서도 표본수의 차이가 크

지 않기 때문에 분석시점 이후의 사망률이 낮고 생존분포가 지수분포를 따른다고 가정하기 어려운 경우에는 생존분포에 대한 가정이 필요하지 않은 Freedman방법을 사용할 수 있겠다. 또한 Rubinstein-Gail-Santner방법은 Lachin방법보다 더 적은 환자수를 주기는 하나 그 차이는 크지 않다.

표 3.2: 각 그룹당 필요한 표본수 ($T = 2, \tau = 0, \alpha = 0.05$, 단측검정)

Design	$\delta = 1.5$		$\delta = 2.0$	
	power=0.8	power=0.9	power=0.8	power=0.9
Pasternack-Gilbert *	78.32	108.91	28.85	40.38
George-Desu *	75.21	104.18	25.73	35.65
Freedman *	77.28	107.05	27.82	38.54
Rubinstein-Gail-Santner	150.25	208.12	57.65	79.85
Lachin	153.26	212.29	60.37	83.62

(*는 사망자수를 제시한것임)

3.3. 연구기간에 따른 검정력변화

표 3.3은 연구기간에 따른 검정력의 변화를 보기 위해 표본수를 고정시켰을 때의 각 방법의 검정력을 계산한 결과이다. 추가추적기간을 갖는 것이 검정력이 커진다는 것을 알 수 있다. 환자 추적기간은 없고 환자 수집기간만 4년인 경우 그룹당 50명, $\alpha = 0.05, \delta = 1.5$ 일 때 Rubinstein-Gail-Santner방법에서의 검정력의 크기는 0.520인 반면 환자 수집기간 및 추적기간이 각각 2년인 경우의 검정력의 크기는 0.608이며, Lachin방법에서는 각각 0.507과 0.591이다. 환자 수집기간(accrual period; T)과 환자 추적기간(follow-up period; τ)의 합을 연구기간이라 하자. 이로부터 두 방법 모두 연구기간이 일정할 때 상대적으로 추적기간이 긴 연구가 환자수집기간이 긴 연구에 비해 검정력이 크다는 것을 알 수 있다.

Rubinstein-Gail-Santner방법과 Lachin방법을 비교해 보면 Rubinstein-Gail-Santner방법의 검정력이 Lachin의 검정력보다 크며, 환자수집기간이 길수록, 환자추적기간이 길수록 검정력의 차이는 커짐을 알 수 있다. 환자수집기간 및 환자추적기간의 증가에 따른 검정력의 증가율은 Rubinstein-Gail-Santner방법이 Lachin방법보다 크다. 예를 들어 $\delta = 1.5$ 이고 그룹당 환자수가 50명일 때 $\tau = 0$ 인 경우 환자수집기간이 1년에서 3년으로 증가하면 Rubinstein-Gail-Santner방법에서는 약 58%의 검정력이 증가한 반면, Lachin방법에서는 55% 증가하였고, 환자수집기간이 1년, 환자추적기간이 2년일 때 추적기간이 없는 것에 비해 Rubinstein-Gail-Santner방법에서는 검정력이 약 95% 증가한 반면 Lachin방법에서는 약 90%가 증가하였다. 그러나 두 방법 모두에 있어서 환자수집기간 및 환자추적기간의 증가에 따른 검정력의 증가율은 감소한다. 예컨대 $\delta = 1.5$ 이고 그룹당 환자수가 50명, $\tau = 2$ 인

경우 Rubinstein-Gail-Santner 방법에서 환자수집기간이 1년씩 증가함에 따라 검정력의 증가는 2.7%에서 0.8%로 감소하며 Lachin방법에서 $T = 2$ 인 경우 환자추적기간이 1년씩 증가함에 따라 검정력의 증가는 약 32%에서 0.5%로 감소한다. 또한 δ 가 작을수록 검정력의 증가율이 크다. 또한 어느 시점에서든 환자수집기간을 증가시키는 것보다 환자추적기간을 증가시키면 더 큰 검정력을 얻을 수 있음을 쉽게 알 수가 있다.

Pasternack-Gilbert, George-Desu방법 및 Freedman방법에서는 사망환자수를 계산하므로 연구기간에 따른 변화는 의미가 없다.

표 3.3: 연구기간에 따른 검정력 변화 : ($\alpha = 0.05$, 단측검정, 그룹당 50명)

Design	T	$\delta = 1.5$			$\delta = 2.0$		
		$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 0$	$\tau = 1$	$\tau = 2$
Pasternack-Gilbert	2	0.636			0.945		
	3						
	4						
George-Desu	2	0.649			0.966		
	3						
	4						
Freedman	2	0.638			0.954		
	3						
	4						
Rubinstein-Gail-Santner	2	0.417	0.555	0.608	0.749	0.899	0.938
	3	0.481	0.579	0.618	0.832	0.919	0.945
	4	0.520	0.594	0.624	0.874	0.931	0.950
Lachin	2	0.411	0.543	0.591	0.732	0.876	0.912
	3	0.471	0.564	0.599	0.807	0.894	0.918
	4	0.507	0.577	0.604	0.845	0.904	0.922

3.4. 예상 검정력(PREDICTED POWER)과 관측 검정력(OBSERVED POWER)의 비교

위의 공식들을 사용할 때 예상되는 검정력(predicted power)과 실제로 로그순위검정법(log rank test)을 이용하였을 때 관측되는 검정력(observed power)을 비교하여 보자. 관측검정력을 계산하기 위해서 모의실험을 실시하였다. 각 그룹당 50명의 환자가 연구에 참여한다고 가정하고, 세 가지 경우의 환자수집기간(accrual period) ($T = 2, 3, 4$ 년), 세 가지 경우의 환자추적기간(follow-up period) ($\tau = 0, 1, 2$ 년), 두 가지 경우의 위험율비(hazard ratio) ($\delta = 1.5, 2.0$), 두 가지 형태의 환자참여(patient entry) (균일분포, 포아송분포), 네 가지 형태의 중도절단(censoring) (중도절단이 없는 경우, 연구의 종료가 원인, 균일분포,

지수분포)에 대해서 각 1000번의 모의실험을 실시하였다. 모든 경우에 (0, T)사이엔 환자가 연구에 참여하고, T + τ 시점에서 연구를 종료하고 분석을 실시하는 것으로 가정하였다. 중도절단이 균일(uniform)분포를 따르는 경우에는 U(0, 2)를 따르고, 지수분포의 경우는 두 그룹에서 똑같이 exp(1)을 따른다고 가정하였다. 앞에서 소개한 공식들이 연구에 참여하는 환자수 또는 사망자수를 산출하도록 제안되었기 때문에, 공식들을 사용할 때 예상되는 검정력과 실제로 로그순위검정을 실시하였을 때 관측되는 검정력을 비교하기 위해서는 먼저 일정한 수의 환자가 연구에 참여하였을 때 환자참여의 형태나 중도절단의 형태에 따라서 얼마나 많은 사망자가 발생하는가를 알아볼 필요가 있다. 이를 위해서 앞에서 소개한 관계식을 이용할 수도 있으나 여기에서는 별개의 모의실험을 실시하여 사망자수를 구하였다. 표 3.4는 그룹당 50명의 환자가 참여하였을 때 여러 가지 경우에 대해서 총 사망자수를 계산한 것인데 이는 100번의 모의임상실험에서 구해진 사망자수를 평균한 것이다.

표 3.4: 관측된 사망자수 (그룹당 50명의 환자가 참여)

방법	분포형태	T	δ = 1.0			δ = 1.5			δ = 2.0		
			τ = 0	τ = 1	τ = 2	τ = 0	τ = 1	τ = 2	τ = 0	τ = 1	τ = 2
Pasternack -Gilbert	entry: uniform censoring: no censoring	2	100			100			100		
		3									
		4									
George -Desu	entry: Poisson censoring: no censoring	2	100			100			100		
		3									
		4									
Freedman	entry: Poisson censoring: end-of-study	2	57	84	94	51	78	90	47	73	85
		3	68	88	96	62	83	92	58	78	88
		4	76	91	97	70	87	94	66	83	90
Rubinstein -Gail -Santner	entry: Poisson censoring: exponential	2	37	48	50	32	43	45	30	39	42
		3	41	49	50	36	44	45	33	41	41
		4	44	50	50	39	44	45	36	41	42
Lachin	entry: uniform censoring: end-of-study	2	57	84	94	51	78	90	47	73	86
		3	69	89	91	63	83	92	59	79	88
		4	76	92	97	71	87	94	67	83	91

3.1절에서 다섯 가지 방법에 따른 가정들을 비교하였는데, 여기에서는 표 3.4에서 정리된 바와 같이 환자참여-중도절단형태에 관한 조합 중에서 각 방법의 가정에 맞는 조합을 찾아 예상되는 검정력과 실제로 관측되는 검정력을 비교하였다. 표 3.5와 표 3.6은 각각 그룹당 50명의 환자가 참여하고 위험율의 비가 1.5와 2.0일 때 방법별로 공식을 사용해서 구한 예상 검정력과 로그순위검정을 실시해서 관측된 검정력을 보여준다. 예상되는 검정력을 구하기 위해서는 참여환자수나 표 3.4에서 구한 사망자수를 각각의 경우에 맞게 공식에 대입하였다. 다시 말해서, Pasternack-Gilbert 방법과 George-Desu 방법은 사망자수를 이용하지만 중도절단이 없이 모두 관측된 경우와 비교하였기 때문에 참여한 환자수를 대입하였고, Rubinstein-Gail-Santner 방법과 Lachin 방법에서는 참여환자수를, Freedman 방법에서는 표

3.4에서 구한 사망자수를 대입하였다. 앞에서 설명한 바와 같이 Pasternack-Gilbert 방법과 George-Desu 방법은 중도절단없이 모든 환자가 사망할 때까지 관측한 경우(즉, $\tau = \infty$)와 비교하였기 때문에 환자추적기간($\tau = 0, 1, 2$ 년)의 값은 무시해도 좋겠다. 또한 Rubinstein-Gail-Santner방법의 경우 중도절단이 두 그룹에서 똑같이 $\phi_C = \phi_E = 1$ 인 지수분포를 따른다고 가정하고 검정력을 계산하였다.

표 3.5: 예상 검정력과 관측된 검정력의 비교 (그룹당 50명, $\delta = 1.5$, $\alpha = 0.05$, 단측검정)

Design	T	$\tau = 0$		$\tau = 1$		$\tau = 2$	
		기대치	관측치	기대치	관측치	기대치	관측치
Pasternack-Gilbert	2		0.681		0.681		0.681
	3	0.636	0.665	0.636	0.665	0.636	0.665
	4		0.656		0.656		0.656
George-Desu	2		0.681		0.681		0.681
	3	0.638	0.665	0.638	0.665	0.638	0.665
	4		0.656		0.656		0.656
Freedman	2	0.412	0.442	0.548	0.566	0.598	0.620
	3	0.472	0.469	0.568	0.616	0.608	0.636
	4	0.511	0.536	0.585	0.610	0.614	0.637
Rubinstein-Gail-Santner	2	0.312	0.323	0.373	0.390	0.383	0.395
	3	0.336	0.342	0.377	0.385	0.383	0.379
	4	0.348	0.329	0.379	0.383	0.384	0.377
Lachin	2	0.411	0.433	0.543	0.572	0.591	0.623
	3	0.471	0.485	0.564	0.621	0.599	0.628
	4	0.507	0.522	0.577	0.626	0.604	0.640

이제 표 3.5와 표 3.6의 결과를 토대로 각 방법들을 사용할 때 예상되는 검정력과 실제로 로그순위검정을 실시하였을 때 관측되는 검정력을 비교해보기로 하자. Pasternack-Gilbert 공식과 George-Desu 방법은 중도절단이 없이 사망이 모두 관측된 경우와 비교하였는데, 두 방법의 차이점은 단지 환자참여의 형태에 있다. 다시 말해서, Pasternack-Gilbert 방법은 균일분포를 가정한 반면에 George-Desu 방법은 포아송(Poisson) 분포를 가정하였다. 두 공식 모두가 위험율 비(δ)의 최대우도추정량(maximum likelihood estimator)에 근거해서 모수적 검정방법을 실시하는 것을 전제로 제안되었음에도 불구하고, 상당히 관측치에 근접한 예상검정력을 계산해주고 있다. 특히 George-Desu 공식이 Pasternack-Gilbert 공식에 비해서 더욱 관측치에 근접한 검정력을 가지고있고, 이러한 경향은 위험율의 비(δ)가 커질 때 더욱 두드러진다. 나머지 세 가지 공식도 관측치와 상당히 근접하지만 거의 모든 경우에 관측치 보다 약간 낮은 예상검정력을 주고 있으며, 위험율의 비(δ)가 커진다 해도 그 차이에는 변함이 없다. 이는 정해진 검정력을 가지기 위해 필요한 환자수를 계산할 때 약간 필요

이상으로 많은 환자수를 계산하게 되는 것을 의미하는데, 이는 분석시점인 $T + \tau$ 에서 중도 절단이 되어 그 시점 이후의 정보를 이용할 수 없기 때문에 나타날 수 밖에 없는 예상된 결과라 하겠다. Rubinstein-Gail-Santner 방법의 경우 특히 환자추적기간(τ)이 커짐에 따라서 예상치가 관측치와 비슷한 경향을 보이고 있다. 이는 로그순위검정을 실시할 때 충분한 환자추적기간을 허용하는 것이 중요하다는 것을 보여주고 있다. 앞 절에서는 Lachin 방법과 Freedman 방법의 경우에 환자수가 고정되어 있을 때는 연구참여기간이 긴 경우보다는 환자들이 일찍 연구에 참여해 오랜시간 동안 추적하는 것이 더욱 큰 검정력을 주는 것을 보았는데, 여기에서는 오랜시간 동안 추적한다고 해도 예상치와 관측치가 가까워지지 않는 것을 알 수 있다. 다시 말해서, Rubinstein-Gail-Santner 공식의 경우와는 달리 Lachin 공식이나 Freedman 공식을 사용할 때에는 필요이상으로 많은 환자수를 계산하는 단점을 환자추적기간의 확장 등을 통해서 보완하기가 힘들다고 할 수 있다.

표 3.6: 예상 검정력과 관측된 검정력의 비교 (그룹당 50명, $\delta = 2.0$, $\alpha = 0.05$, 단측검정)

Design	T	$\tau = 0$		$\tau = 1$		$\tau = 2$	
		기대치	관측치	기대치	관측치	기대치	관측치
Pasternack-Gilbert	2		0.961		0.961		0.961
	3	0.945	0.963	0.945	0.963	0.945	0.963
	4		0.957		0.957		0.957
George-Desu	2		0.961		0.961		0.961
	3	0.996	0.963	0.966	0.963	0.966	0.963
	4		0.957		0.957		0.957
Freedman	2	0.735	0.775	0.884	0.905	0.923	0.944
	3	0.814	0.842	0.903	0.929	0.931	0.957
	4	0.856	0.864	0.917	0.930	0.935	0.953
Rubinstein-Gail-Santner	2	0.580	0.603	0.686	0.701	0.734	0.708
	3	0.624	0.626	0.693	0.698	0.705	0.711
	4	0.646	0.663	0.697	0.728	0.706	0.711
Lachin	2	0.732	0.773	0.876	0.908	0.912	0.945
	3	0.807	0.849	0.894	0.937	0.918	0.956
	4	0.845	0.875	0.904	0.923	0.922	0.956

3.5. 각 방법들의 로버스트성

앞에서 소개한 각 방법들이 환자참여분포와 중도절단형태 및 생존분포에 관한 가정을 만족하지 못할 경우에 어느 정도 로버스트한가를 알아보자. 표 3.5와 표 3.6에서 보듯이 중도절단이 없을 때 환자참여분포가 균일분포인 방법(Pasternack-Gilbert 방법)은 포아송분포

에 대해 로버스트하며, 연구의 종료가 중도절단 사유일 때 환자참여분포가 균일분포인 방법(Lachin 방법)은 포아송분포에 대해 로버스트하다. 역으로 George-Desu 방법과 Freedman 방법과 같이 환자참여분포가 포아송분포를 따르는 경우에도 각각 중도절단이 없을 때와 연구의 종료가 중도절단 사유일 때 환자참여분포가 균일분포인 것에 대해 로버스트하다.

Rubinstein-Gail-Santner 방법에 대해 환자참여분포와 중도절단의 형태에 따른 로버스트성을 알아보기 위하여 환자참여분포가 균일분포와 포아송분포이고 중도절단형태가 지수분포, 균일분포인 경우에 대한 별도의 모의실험을 실시하였다. 이 때 중도절단시간의 평균이 같아지도록 지수분포는 $\exp(1)$ 을, 균일분포는 $U(0, 2)$ 를 사용하였다. 결과는 표 3.7과 같다. 환자참여의 분포가 균일분포인 것과 포아송분포인 것과의 차이는 거의 없으며 중도절단형태에 따른 검정력의 변화도 그다지 크지 않아 Rubinstein-Gail-Santner 방법은 환자참여분포와 중도절단의 형태에 대해 로버스트하다고 할 수 있겠다.

표 3.7: 분포와 중도절단의 가정에 따른 검정력의 변화 (그룹당 50명, $\alpha = 0.05$, 단측검정)

Design	T	$\delta = 1.5$			$\delta = 2.0$		
		$\tau = 0$	$\tau = 1$	$\tau = 2$	$\tau = 0$	$\tau = 1$	$\tau = 2$
entry: Poisson censoring: exponential (Rubinstein-Gail-Santner)	2	0.323	0.390	0.395	0.603	0.701	0.708
	3	0.342	0.385	0.379	0.626	0.698	0.711
	4	0.329	0.383	0.377	0.663	0.728	0.711
entry: uniform censoring: exponential	2	0.330	0.363	0.406	0.622	0.673	0.712
	3	0.361	0.374	0.363	0.612	0.697	0.711
	4	0.325	0.376	0.352	0.628	0.703	0.680
entry: Poisson censoring: uniform	2	0.359	0.399	0.412	0.631	0.731	0.749
	3	0.380	0.397	0.417	0.683	0.751	0.757
	4	0.366	0.401	0.431	0.691	0.746	0.760
entry: uniform censoring: uniform	2	0.364	0.401	0.412	0.649	0.745	0.749
	3	0.387	0.396	0.417	0.708	0.752	0.757
	4	0.367	0.408	0.431	0.692	0.758	0.760

위의 다섯 가지 방법들 중 Freedman 방법을 제외한 네 가지 방법은 생존분포가 지수분포를 따른다고 가정하고 있다. 이러한 가정에서 벗어났을 때 각 방법들의 로버스트성을 알아보기 위해 모의실험결과중에서 생존분포가 와이블분포(Weibull distribution)인 경우와 비교하여 보자. 이 때 모든 경우에 대해 두 그룹의 생존분포간의 위험비(hazard ratio)는 일정하다고 가정하였다. 두 그룹의 생존함수는 각각 $S_{1(t)} = \exp(-\delta\gamma_1 t^{\gamma_2})$ 와 $S_{2(t)} = \exp(-\gamma_1 t^{\gamma_2})$ 을 가정하였는데, 여기서 γ_1 은 척도모수(scale parameter)로서 그룹 1의 생존시간 중앙값(median survival time)을 1년이 되게 하기 위하여 $\gamma_1 = \ln 2/\delta$ 로 하였고, γ_2 는 형태모수(shape parameter)로서 $\gamma_2 = 1/3, 2/3, 3/2, 3$ 을 사용하였다. δ 는 두 그룹간 위험비로 앞 절에서와 같이 $\delta = 1.5, 2.0$ 을 사용하였다. 이와 같은 구조하에서 수행한 모의실험 결과는 표 3.8, 표

3.9와 같다. 이 때 각 상황에서의 관측된 검정력을 생존분포가 지수분포인 경우 계산된 예상검정력과 비교하였다. Pasternack-Gilbert 방법과 George-Desu 방법의 경우 모든 γ_2 인 값에 대해 약 7%미만의 적은 검정력 차이를 보였으며, Rubinstein-Gail-Santner 방법도 모든 γ_2 인 값에 대해 약 7%미만의 검정력차이를 보였다. Lachin 방법에서는 연구기간이 4년인 경우에 약 10-15% 정도의 차이를 보이고 있다.

표 3.8: 분포에 따른 검정력의 변화 (그룹당 50명, $\delta = 1.5$, $\alpha = 0.05$, 단측검정)

Design	τ γ_2	$\tau = 0$					$\tau = 1$					$\tau = 2$				
		exp	1/3	2/3	3/2	3	exp	1/3	2/3	3/2	3	exp	1/3	2/3	3/2	3
Pasternack -Gilbert	2		0.576	0.679	0.681	0.681	$\tau = 0$ 일 때와 동일									
	3	0.636	0.574	0.657	0.665	0.665										
	4		0.567	0.653	0.656	0.656										
George -Desu	2		0.567	0.678	0.681	0.681	$\tau = 0$ 일 때와 동일									
	3	0.638	0.575	0.663	0.665	0.665										
	4		0.571	0.652	0.656	0.656										
Freedman	2	0.412	0.391	0.377	0.396	0.409	0.548	0.447	0.494	0.573	0.622	0.598	0.435	0.499	0.631	0.656
	3	0.472	0.383	0.410	0.455	0.480	0.566	0.433	0.493	0.612	0.643	0.608	0.454	0.534	0.641	0.660
	4	0.511	0.407	0.454	0.516	0.555	0.585	0.458	0.533	0.612	0.648	0.614	0.462	0.550	0.652	0.655
Rubinstein -Gail -Santner	2	0.312	0.324	0.305	0.258	0.252	0.373	0.361	0.345	0.329	0.341	0.383	0.351	0.360	0.348	0.356
	3	0.336	0.342	0.314	0.291	0.288	0.377	0.346	0.346	0.346	0.346	0.383	0.350	0.343	0.322	0.360
	4	0.348	0.316	0.293	0.273	0.301	0.379	0.380	0.340	0.322	0.318	0.384	0.357	0.335	0.339	0.344
Lachin	2	0.411	0.385	0.377	0.386	0.412	0.543	0.432	0.497	0.567	0.623	0.591	0.442	0.515	0.623	0.657
	3	0.471	0.407	0.412	0.463	0.506	0.564	0.437	0.501	0.606	0.641	0.599	0.445	0.533	0.644	0.359
	4	0.507	0.399	0.447	0.520	0.537	0.577	0.437	0.525	0.619	0.682	0.456	0.559	0.649	0.654	0.654

4. 토론 및 결론

이제까지 생존함수를 비교하기 위한 다섯 가지 방법들에 대하여 설명하였다. 이들 방법에서의 가정들을 고찰하고, 다양한 상황에서의 표본수와 검정력 등을 비교 제시하였으며 또한 모의실험을 통해서 각 방법들의 이론상의 검정력과 실제로 검정을 실시한 후에 관측된 검정력 등을 비교하였다. 그 결과로서 의학연구자들이 다양한 상황에서 어떠한 방법을 선택할 것인가에 대한 정보를 제시하였다.

Pasternack-Gilbert 방법과 George-Desu 방법에서는 중도절단이 없다고 가정하고 필요한 사망자수를 계산한다. Rubinstein-Gail-Santner 방법은 중도절단, 추적실패(loss to follow-up), 추가추적기간을 고려하므로 Pasternack-Gilbert 방법 및 George-Desu 방법보다 더 현실적으로 적용 가능하다. 추적실패와 추가추적기간이 없다면 George-Desu 방법에 의한 계산결과와 거의 동일하게 된다. 그러나 추가추적기간을 사용하면 필요한 환자수를 크게 줄일 수가 있다. Lachin 방법을 사용하면 각 그룹에서 서로 다른 확률로 환자를 임의로 할당할 수 있으며 다른 방법에 비하여 여러 가지 환자참여형태를 가정할 수가 있다. 또한 Freedman 방법에서는 각 그룹에 다른 수의 환자를 할당할 수 있으며 생존분포에 대한 가정이 필요하지 않다.

표 3.9: 분포에 따른 검정력의 변화 (그룹당 50명, $\delta = 2.0$, $\alpha = 0.05$, 단측검정)

Design	τ τ_2	$\tau = 0$					$\tau = 1$					$\tau = 2$				
		exp	1/3	2/3	3/2	3	exp	1/3	2/3	3/2	3	exp	1/3	2/3	3/2	3
Pasternack -Gilbert	2		0.904	0.963	0.961	0.961	$\tau = 0$ 일 때와 동일									
	3	0.945	0.919	0.964	0.983	0.963										
	4		0.924	0.957	0.957	0.957										
George -Desu	2		0.904	0.963	0.961	0.961	$\tau = 0$ 일 때와 동일									
	3	0.966	0.924	0.963	0.963	0.963										
	4		0.930	0.956	0.957	0.957										
Freedman	2	0.735	0.678	0.681	0.698	0.735	0.884	0.776	0.829	0.901	0.943	0.923	0.805	0.877	0.949	0.957
	3	0.814	0.705	0.738	0.806	0.857	0.903	0.766	0.853	0.927	0.957	0.931	0.814	0.901	0.958	0.972
	4	0.856	0.724	0.770	0.841	0.876	0.917	0.799	0.867	0.930	0.951	0.935	0.804	0.895	0.957	0.964
Rubinstein -Gail -Santner	2	0.580	0.617	0.556	0.469	0.430	0.686	0.652	0.622	0.603	0.639	0.734	0.666	0.639	0.638	0.662
	3	0.624	0.609	0.579	0.532	0.526	0.693	0.649	0.613	0.627	0.649	0.705	0.654	0.653	0.623	0.657
	4	0.646	0.594	0.571	0.521	0.537	0.697	0.653	0.644	0.627	0.644	0.706	0.659	0.639	0.633	0.659
Lachin	2	0.732	0.702	0.681	0.688	0.742	0.876	0.773	0.824	0.899	0.945	0.912	0.804	0.88	0.947	0.956
	3	0.807	0.720	0.750	0.829	0.861	0.894	0.762	0.847	0.932	0.962	0.918	0.824	0.891	0.960	0.972
	4	0.845	0.733	0.777	0.859	0.907	0.904	0.785	0.865	0.932	0.957	0.922	0.798	0.885	0.961	0.964

Pasternack-Gilbert 방법과 George-Desu 방법은 위험을 비(δ)의 최대우도추정량에 근거해서 모수적 검정방법을 실시하는 것을 전제로 제안되었음에도 불구하고, 상당히 관측치에 근접한 예상검정력을 계산해주고 있다. 특히 George-Desu 방법이 Pasternack-Gilbert 방법에 비해서 더욱 관측치에 근접한 검정력을 가지고 있고, 이러한 경향은 환자수가 많이 참여하는 경우나 위험율의 비(δ)가 커질 때 더욱 두드러진다. 나머지 세 공식도 관측치와 상당히 근접하지만 거의 모든 경우에 관측치보다 약간 낮은 예상검정력을 주고 있으며, 위험율의 비(δ)가 커진다 해도 그 차이에는 변함이 없다. Rubinstein-Gail-Santner 방법의 경우 특히 환자추적기간(τ)이 커짐에 따라서 예상치가 관측치와 비슷한 경향을 보이고 있다. 이는 로그순위검정을 실시할 때 충분한 환자추적기간을 허용하는 것이 중요하다는 것을 보여주고 있다. 이와 달리 Lachin 방법이나 Freedman 방법을 사용할 때에는 필요이상으로 많은 환자수를 계산하는 단점을 환자추적기간의 확장 등을 통해서 보완하기가 힘들다고 할 수 있다.

모의실험을 통해 각 방법들의 로버스트성을 살펴본 결과 환자참여분포에 대해 각각 포아송분포인 경우와 균일분포인 경우 모두 로버스트한 결과를 보였으며 Rubinstein-Gail-Santner 방법에서는 중도절단의 형태가 지수분포, 균일분포를 따르는 경우에 모두 로버스트한 결과를 보였다. 또한 생존분포가 지수분포가 아닌 와이블분포를 따르는 경우에 대해서도 대체로 로버스트한 것으로 나타났다.

위의 결과들을 종합해 보면, George-Desu 방법이 가장 검정력이 크며, 관측검정력과 차이도 가장 적고, 또한 가정된 분포에 대해 매우 로버스트하므로 필요한 사망자수를 계산하는 연구에 있어서 가장 적절한 방법으로 사료된다. 중도절단을 고려하는 연구라면, Lachin 방법보다는 Rubinstein-Gail-Santner 방법을 사용하는 것이 유리하다고 생각된다.

임상시험연구에 참여하는 환자들의 나이, 과거병력 또는 질병의 상태가 생존율에 많은

영향을 미치는 경우에는 이러한 요인을 층(strata)으로 고려하여 환자를 확률화 할당하여야 한다. 이와 같이 층을 고려하는 경우에는 필요한 환자수가 층이 없는 경우보다 많아지게 된다. 많은 연구자들이 층을 고려하였을 때 필요한 표본수를 계산하는 방법을 제시한 바 있다(Bernstein and Lagakos, 1978; Palta and Amini, 1985; Lachin and Foulkes, 1986; Moussa; 1988). 다양한 경우의 생존분포와 증도절단의 형태, 층의 수, 연구기간에 따라 필요한 표본수를 비교하고 모의실험을 통해 이들 방법들을 비교하는 연구가 향후 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] Donner, A. (1984). Approaches to sample size estimation in the design of clinical trials - A review, *Statistics in Medicine*, 3, 199-214.
- [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] Freiman, J. A., Chalmers, P. C. , Smith, H. Jr. and Kuebler, R. R. (1978). The importance of beta, the Type II error and sample size in the design and interpretation of the randomized control trial, *New England Journal of Medicine*, 299, 690-694.
- [5] 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.
- [6] Hsieh, F. Y. (1992). Comparing sample size formulae for trials with unbalanced allocation using the logrank test, *Statistics in Medicine*, 11, 1091-1098.
- [7] Lakatos, E. (1988). Sample size based on the log-rank statistic in complex clinical trials, *Biometrics*, 44, 229-241.
- [8] Lakatos, E. and Lan K. K. G. (1992). A comparison of sample size methods for the logrank statistic, *Statistics in medicine*, Vol. 11, 179-191.
- [9] Lachin, J. M. (1981). Introduction to sample size determination and power analysis for clinical trials, *Controlled Clinical Trials*, 2, 93-113.
- [10] Lachin , J. M. and Foulkes, M. A. (1986). Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification, *Biometrics*, 42, 507-519.
- [11] Moussa, M. A. A. (1988). Planning the size of survival time clinical trials with allowance for stratification. *Statistics in Medicine*. 7. 559-569.

- [12] Palta M. and Amini, S. (1985). Consideration of covariates and stratification in sample size determination for survival time studies, *Journal of Chronic Disease*, 38, 801-809.
- [13] 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.
- [14] Rubinstein, L. V., Gail, M. H. and Santner, T. J. (1981). Planning the duration of a comparative clinic trial with loss to follow-up and a period of continued observation, *Journal of Chronic Disease*, 34, 469-479.
- [15] Shuster, J. J. (1990). *Handbook of Sample Size of Guidelines for Clinical Trials*, CRC Press, Inc., West Palm Beach, Florida.

[1997년 8월 접수, 1998년 3월 최종수정]

Sample Size Determination in Survival Studies *

Mira Park¹⁾ Seon Woo Kim²⁾ Jae Won Lee³⁾

ABSTRACT

One of the most important issues in the area of clinical trial research is the determination of the sample size required to insure a specified power in detecting a real or clinically relevant difference of a stated magnitude. Increasingly, medical journals are requiring authors to provide information on the sample size needed to detect a given difference.

We restrict our attention to the designs for comparing two survival distributions. These are concerned with the survival time which is defined as the interval from a baseline (e.g. randomization) to failure (e.g. death, recurrence of disease). Survival times are right censored when patients have not failed by the time of analysis or have been lost to follow-up during the trial.

For different types of clinical trials for comparing survival distributions, there have been many research in sample size determination. We review the existing literature concerning commonly used sample size formulae in the design of randomized clinical trials, and compare the assumption, the power and the sample size calculation of these methods. We also compare by simulation the expected power and observed power of each method under various circumstances. As a result, guidelines in terms of practical usage are provided.

* This work is supported by the SamSung grant, #SBRI, C-96-020.

1) Department of Premedicine, EulJi Medical College, TaeJeon 301-112, Korea.

2) SamSung Biomedical Research Institute, Clinical Research Center, Seoul 135-230, Korea.

3) Department of Statistics, Korea University, Seoul 136-701, Korea.