

## 임상시험에서 치료율 비교를 위한 집단축차검정법에 관한 고찰<sup>1)</sup>

박 경 미<sup>2)</sup> · 이 재 원<sup>3)</sup>

### 요 약

의약학 분야에 있어서 임상시험(clinical trial)을 하는 도중에 약의 독성이 발견되거나 연구 초기에 두 처리군간의 차이가 크게 나타난다면 임상시험의 종료를 신중하게 고려하는 것이 바람직할 것이다. 따라서 요즈음에는 일정시간이 경과할 때마다 실험결과를 중간점검(interim monitoring)하는 것이 점점 요구되어지고있는 추세다. 본 논문에서는 두 치료법의 치료율을 비교하는 경우에 비모수검정통계량에 집단축차검정법을 적용하여 생존분포, 치료율, 중도절단을, 오류분배율함수(error spending rate function)등을 다양하게 변화시키면서 모의실험을 실시하여 Gray-Tsiatis 통계량, logrank 통계량, Gehan 통계량 각각에 대한 검정력과 유의수준을 비교하였다.

### 1. 서론

의약학 분야에 있어서 임상시험(clinical trial)을 하는 도중에 약의 독성이 발견되거나 연구초기에 두 처리군간의 차이가 크게 나타난다면 임상시험의 종료를 신중하게 고려하는 것이 바람직할 것이다. 그러므로 일정시간이 경과할 때마다 실험결과를 중간점검(interim monitoring)하는 것이 점점 요구되어지고있는 추세다. 이런 경우에 고려되어질 수 있는 통계적 검정방법중의 하나가 집단축차검정법(group sequential testing method)이다. 그러나 중간분석마다 일정한 유의수준(예를 들어 5%)을 반복하여 사용하면 실험 전체의 유의수준은 훨씬 커지게 된다. 이러한 문제점을 보완하기 위하여 Pocock(1977), O'Brien and Fleming(1979), Lan and DeMets(1983) 등은 각각의 분석시점에 따른 제1종오류의 적절한 분배방법을 제시하였다.

몇몇 임상시험에서는 환자의 상당수가 완치된후 일정한 기간이 지나도 증상이 재발하지 않는 경우가 있다. 이때 우리의 관심은 주로 두 처리군간의 치료율의 차이를 검정하는데 있을 것이다. Gray and Tsiatis(1989)는 두 처리군간의 비례생존분포함수를 정의한 뒤 비모수적인 방법에 근거하여 검정력을 최대화하는 검정통계량을 제안했다. 그 뒤로 Lee and Sather(1995)는 Farewell(1982)이 제시한 혼합모형(mixture model)과 Gray-Tsiatis 검정법에 근거한 각각의 집단축차검정법을 제시하였다.

본 논문에서 사용한 세 가지 검정법 중에서 비모수적인 Gray-Tsiatis 검정방법은 두 처리군

1) 이 연구는 1994년도 한국과학재단 연구비 지원에 의한 결과임(과제번호:94-1300-01-01-3)  
2) (130-650) 서울 청량우체국 사서함 255호, 과학기술정책관리연구소 위촉연구원  
3) (136-701) 서울시 성북구 안암동 5-1, 고려대학교 통계학과 부교수

에서 완치되지 않은 환자들의 생존분포가 같다고 가정하므로 두 처리군의 생존분포의 동일성 검정이 곧 치료율의 동일성 검정을 의미한다. 그러나 모수적방법인 Farewell(1982)이나 Kuk and Chen(1992)의 검정방법은 분포에 대한 여러가지 가정을 필요로하므로 실제적으로 알려진 정도나 사용의 편의성 때문에 logrank 통계량, Gehan 통계량 등이 더 많이 사용되는 것이 현실이다. 따라서 본 논문에서는 비모수적방법에 근거하여 생존분포, 치료율, 중도절단을, 오류분배율함수(error spending rate function)등을 다양하게 변화시키면서 Gray-Tsiatis 통계량, logrank 통계량, Gehan 통계량 각각에 대한 집단축차검정법의 장단점들을 모의실험을 통해 비교해 보았다.

## 2. 집단축차검정법

### 2.1 일변량 관측치일때의 집단축차검정법

집단축차검정(group sequential testing)이란 일정한 시점마다 그때까지 누적된 자료를 가지고 계획된 분석을 반복 시행하는 것이다. 두 처리군간의 차이를 알기 위해서  $K$ 번의 중간분석(interim analysis)을 고려하고, 이때  $k(k=1, \dots, K)$ 번째 분석의 검정통계량을  $S(k)$ 라고 하자. 그리고 각 환자마다 한 개의 관측치만을 얻는다고 가정하면  $S(k)$ 는 독립적인 임의의 반응변수들의 부분합을 표준화시킨 통계량이 된다. 따라서 각 분석 시점 이후에 추가되는 관측치는 분석시점 이전까지 얻어진 통계량과 독립이 되므로 이러한 경우에 검정통계량  $(S(1), S(2), \dots, S(K))$ 가 독립증분(independent increments)의 성질이 있다고 말한다. 독립증분의 성질을 가지면  $(S(1), S(2), \dots, S(K))$ 는 근사적으로 다변량 정규분포를 따르며 공분산 행렬도 비교적 간단하게 유도될 수 있다.

예를 들어 각각의 독립적인 관측치들이 정규분포를 따르고 그 분산을 알고 있다고 가정하면, 귀무가설은  $H_0: \theta = \mu_1 - \mu_2 = 0$ 으로 표시될 수 있다. 여기서  $\mu_1$ 과  $\mu_2$ 는 각 처리군의 모평균이다. 이때 각 처리군에  $n$ 명씩 총  $2n$ 명의 환자로부터 관측치를 얻을 때마다 중간분석(interim analysis)을 한다면,  $k$ 번째 중간분석의 검정통계량은  $S(k) = \sum_{j=1}^k \sqrt{n}(\bar{X}_{1j} - \bar{X}_{2j}) / \sqrt{2k\sigma^2}$  이 되

고, 여기서  $\bar{X}_{1j}, \bar{X}_{2j}$ 은  $(j-1)$ 번째와  $j$ 번째 분석시점 사이에 관측된 값들의 평균이다. 여기서  $I(k)$ 를 '정보시간'(information time)(cf. Jennison and Turnbull, 1989)이라고 정의하고  $k$ 번째 분석시점까지 관측된 값들의 평균의 Fisher 정보량(information)을  $I(k) = 2nk/\sigma^2$ 로 사용하면 표준화되지않은 통계량은  $S^*(k) = S(k)\sqrt{I(k)} = S(k)\sqrt{2nk/\sigma^2}$ 이 된다. 이때 귀무가설 하에서  $\{S^*(k): k=1, \dots, K\}$ 는 각각 독립증분의 성질을 가지게 되고, 평균이  $Q$  이고 공분산이  $Cov(S^*(k), S^*(l)) = Var(S^*(k)) = 2nk/\sigma^2, k < l$ 인 다변량정규분포를 따른다. 그러므로 각 시

점  $I(k)$  ( $k=1, \dots, K$ )에서 얻어진  $S^*(k)$ 는 표준브라운운동(Standard Brownian Motion)을 따른다고 할 수 있다.

위의 과정에 따라서  $k$ 번째 중간분석에서  $|S(k)| > b_k$  이면 귀무가설을 기각함과 동시에 임상 시험을 종료한다. 여기서 귀무가설이 맞을 때  $k$ 번째 중간분석에서 귀무가설을 기각할 확률을  $\pi_k (1 \leq k \leq K)$ 라고 정의한다. 그러면 귀무가설 하에서

$$P_0 \{ |S(1)| \leq b_1, \dots, |S(k-1)| \leq b_{k-1}, |S(k)| > b_k \} = \pi_k \quad (2.1)$$

이 되고,

$$\pi_1 + \dots + \pi_K = \alpha \quad (\alpha \text{는 제 1종 오류}) \quad (2.2)$$

이 성립한다. 즉,  $P_0 \{ |S(k)| > b_k \text{ for some } 1 \leq k \leq K \} = \alpha$  가 된다.

## 2.2 제1종오류의 분배와 기각치의 결정

각각의 분석시점에서 (2.1)과 (2.2)를 만족하는 기각치를 선택하는 방법은 여러 통계학자에 의해 개발되었다. Pocock(1977)은 각 분석시점 이후에 통계량의 추가분이 독립이고 추가된 관측치의 수가 같다(independent and equal increment)고 가정한 후  $b_1 = \dots = b_K = C_P(K, \alpha)$  을 제안했다. 여기서  $C_P(K, \alpha)$ 은  $K$ 와  $\alpha$ 에 의존하는 상수이다. O'Brien and Fleming (1979)은  $b_k = C_B(K, \alpha) \sqrt{K/k}$  ( $1 \leq k \leq K$ )을 제시하였다( $C_B(K, \alpha)$ 는  $K$ 와  $\alpha$ 에 의존하는 상수). O'Brien-Fleming 방법은 Pocock 방법에 비해서 초반에는 기각치가 아주 크지만 분석시점이 지날수록 작아지므로, 귀무가설을 기각하기가 초반에는 어렵고 후반에는 쉬워진다. 두 방법에서  $C_P$ 와  $C_B$ 는 (2.1)과 (2.2)를 만족하도록 선택한다. Lan and DeMets(1983)는  $\alpha^*(0) = 0$ ,  $\alpha^*(1) = \alpha$ 을 만족하는 비감소(nondecreasing)함수인 "제1종오류분배율 함수"(Type I error spending rate function)  $\alpha^*(t)$ 를 정의하고,  $\pi_k = \alpha^*(t_k) - \alpha^*(t_{k-1})$ 이 성립하도록  $\pi_k$ 를 계산하였다. 여기서  $t_k$ 는  $k$ 번째 중간분석에서의 정보시간의 비율(즉,  $t_k = I_k/I_K$ )이며 이를 정보비율(information fraction)이라고 정의한다.

만일 각 분석시점 사이에 추가된 관측치의 수가 같아서  $t_k = k/K$ 가 성립되는 경우에는  $\alpha_2^*(t) = \alpha \log \{ 1 + (e-1)t \}$ 는 Pocock방법과,  $\alpha_1^*(t) = 2 \{ 1 - \Phi(Z_{\alpha/2}/\sqrt{t}) \}$ 는 O'Brien-Fleming 방법과 비슷한 기각치를 제공한다 ( $\Phi$ 는 표준화된 정규누적분포함수이고  $\Pr(Z > Z_{\alpha/2}) = \alpha/2$ 이다). 따라서 각각의  $\pi_k$ 를 오류분배율함수를 이용해서 계산한 후에 축차적으로  $b_1, b_2, \dots, b_k$ 를 (2.1)과 (2.2)를 이용해서 계산한다. Lan-DeMets 방법은 전체 중간분석 횟수  $K$ 를 사전에 알아야 할 필요가 없고 또한 각 시점 사이에 추가된 관측치의 수가 같지 않아도 된다는 큰 장점

이 있다. 물론  $k$ 번째 중간분석에서 정보비율  $t_k (= I_k/I_K)$ 를 계산하기 위해 최대정보시간 (maximum information time)  $I_K$ 를 미리 추정해야 한다. 하지만 잘 계획된 임상시험에서는 최대정보시간을 미리 추정하는 것은 쉬운 문제이므로 우려할 바가 못된다.

위에서 언급한 여러가지 집단축차검정법의외에도 더욱 확장된 이론들이 많이 제시되어 있고, 또한 자료가 다변량관측치 일때에도 적용할 수 있는 방법들이 제안되어 있다 (cf. Lee and DeMets, 1991, 1992; Lee, 1994a, 1994b).

### 3. 두 처리군간의 치료율 비교

많은 임상시험에서 환자중에 상당수가 완치되는 경우가 있다. 예를 들면 뛰어난 치료법이 많이 개발된 아동암(pediatric cancer)에 관한 임상시험이 그런 경우이다. 이때 연구자는 단순히 생존분포를 비교하는 것보다 두 처리군의 치료율을 비교해보고 싶을 것이다. 처리군  $q(=1,2)$ 의 치료율을  $\theta_q$ 로 표기하면, 생존분포함수는  $S_q(t) = \theta_q + (1-\theta_q)H_q(t)$ 로 나타낼 수 있다. 여기서  $H_q(t)$ 는 각 처리군에서 치료되지 않은 환자들의 생존함수이다. 그리고 귀무가설은  $H_0: \theta_1 = \theta_2$  (즉, 두 처리군의 치료율은 같다)이다.

Farewell (1982)은 모수적 방법에 근거하여 치료율에는 로지스틱모형을, 분포함수에는 와이블 (Weibull)모형을 적용하는 혼합모형을 제안하였는데, 이 모형에서는 두 처리군의 치료율( $\theta_q$ )이 같더라도 생존분포( $S_q$ )는 같지 않을 수 있다. Lee and Sather(1995)는 혼합모형에 근거하여 각 분석시점에서 구해진 검정통계량 ( $L(1), L(2), \dots, L(K)$ )가 근사적으로 다변량 정규분포를 따르는 것을 보였다. 그리고 Lan and DeMets(1983)가 제안한  $\alpha^*(t)$ 를 사용하여 (2.1) 과 (2.2)를 만족하는 기각치를 반복적으로 구하여 집단축차검정법을 제안하였다.

비모수적인 방법에 근거한 검정법으로 Gray and Tsiatis(1989)는 두 처리군의 생존분포가 같다는 귀무가설에 대해서 검정력을 최대화하는 다음과 같은 최적선형순위검정법(optimal linear rank test)을 제안하였다.  $H_q(t)$ 가 치료되지 않은 각 처리군 환자들의 조건부분포( $q=1,2$ )라고 할 때,  $H_1 = H_2 = H$ 라는 가정하에서 두 처리군의 치료율의 차이를 검정하는 것이다. 즉,  $(1-S_2) = \{ (1-\theta_2)/(1-\theta_1) \} (1-S_1)$  가 성립하므로 생존분포가 비례하며 두 처리군의 치료율이 같으면 생존분포함수가 같다는 것을 의미한다. 따라서 일반적으로 많이 쓰이는 선형순위통계량에 근거한 최적선형순위검정통계량(optimal linear rank test statistic)을 고려하였다.  $i$ 번째 ( $i=1, \dots, n$ ) 환자에 대해서  $Y_i$ 는 임상실험에 들어온 시간,  $V_i$ 는 임상실험에 들어와서 환자가 죽거나 병이 재발한 시간,  $C_i$ 는 임상실험에 들어와서 환자가 중도절단된 시간,  $Z_i (=0$  또는 1)는 처리군을 지정한 함수로 나타낸다면,  $n$ 명의 환자에 대해서  $X_i(t)$  및 중도절단함수

$\Delta_i(t)$  ( $i=1, \dots, n$ )를 관측할 수 있다. 즉,  $X_i(t) = \max \{ \min(V_i, t - Y_i, C_i), 0 \}$ ,  $\Delta_i(t)$ 는  $V_i \leq \min(t - Y_i, C_i)$  이면 1 (관측된 경우) 그 이외의 경우에는 0 (중도절단된 경우)을 갖는다. 따라서 시점  $t$ 에서의 일반화선형순위 통계량

$$U(t) = \sum_{i=1}^n \Delta_i(t) Q(t, X_i(t)) \left( Z_i - \frac{\sum_j I\{X_j(t) \geq X_i(t)\} Z_j}{\sum_j I\{X_j(t) \geq X_i(t)\}} \right)$$

을 계산할 수 있다. 여기서  $I(\cdot)$ 는 지시함수(indicator function)이고,  $Q(t, x)$ 는 시점  $t$ 에서 발생하는 두 처리간의 생존분포의 차이에 어떠한 가중치를 부여할 것인가를 나타내는 함수이다.

Gray and Tsiatis(1989)는 각 시점  $t_k$  ( $t_1 < \dots < t_K$ )에서 구해진 가중로그순위검정 통계량 (weighted logrank statistic)인  $(U(t_1), U(t_2), \dots, U(t_K))$ 가 근사적으로 평균이  $\Omega_K$ 인 다변량 정규분포를 따른다는 것을 보였고, Lin(1991)은 일치추정량  $[\sum_{i=1}^n W_i(t_k) W_i(t_l)] / n$ 에 의해 공분산  $Cov(U(t_k), U(t_l))$  를 추정하였다. 여기서

$$W_i(t) = \Delta_i(t) Q(t, X_i(t)) \left[ Z_i - \frac{\sum_j I\{X_j(t) \geq X_i(t)\} Z_j}{\sum_j I\{X_j(t) \geq X_i(t)\}} \right] - \sum_{j=1}^n \left\{ \frac{I\{X_j(t) \geq X_i(t)\} \Delta_j(t) Q(t, X_j(t))}{\sum_j I\{X_j(t) \geq X_i(t)\}} \left[ Z_j - \frac{\sum_j I\{X_j(t) \geq X_i(t)\} Z_j}{\sum_j I\{X_j(t) \geq X_i(t)\}} \right] \right\}$$

으로 나타낼 수 있다. 이때  $Q(t, x)$ 의 값에 따라서 여러가지 통계량을 구할 수 있다(cf. Peto and Peto, 1972; Prentice, 1978). 가령,  $Q(t, x)=1$  이면 logrank검정 통계량,  $Q(t, x)=S(I)$  이면 Gehan 또는 generalized Wilcoxon검정통계량이 된다 (여기서,  $S(I)$ 는 전체 환자에 대한 그때까지 살아있는 환자수의 비율이다). Gray and Tsiatis(1989)가 제안한 최적선형순위검정통계량은  $Q(t, x) = KM^{-1}(x-)$ 를 이용하는데, 여기서  $KM(x-)$ 는 카플란-마이어(Kaplan-Meier) 추정치의 좌방연속(left-continuous)값이다. 따라서 위에서 제시된 생존분포의 동일성을 검정하는 비모수적 통계량에 근거하여 (2.1)과 (2.2)를 만족하는 기각치들을 축차적으로 구하여 집단축차검정을 실시할 수 있다. 또한  $\pi_k, k=1, \dots, K$ 는 Lan and DeMets(1983)가 제안한 오류분배율함수 (error spending rate function)를 이용해서 계산될 수 있다.

이 밖에도 두 처리군에서 구해진 카플란-마이어 추정치의 차이에 근거한 검정방법이나, 또는 최근에 제안된 비모수적 최대우도추정방법(nonparametric maximum likelihood)이 두 처리군의 치료율을 비교하는데 유용하다 (cf. Laska and Meisner, 1992).

### 4. 모의실험 설정 및 결과분석

#### 4.1 모의실험의 상황설정

- ① 최종분석을 포함해서 세번의 중간분석이 있다 (즉,  $K=3$ ).
- ② 표본크기는 한 치료군 당 50명씩 총 100명의 환자를 배당한다.
- ③ 각각의 치료군에서 사용했던 치료되지 않은 각 치료군 환자들의 조건부분포( $H_1$  과  $H_2$ )는 아래 <분포표>와 같이 정리할 수 있다. 비례위험모형을 만족시키지 않는 경우를 만들기 위해서 Kuk and Chen(1992)에서와 같이 지수분포의 혼합형태를 사용하였다.
- ④ 검정통계량 :

GT (Gray-Tsatis statistic), LR (logrank statistic), GW (Gehan statistic)

- ⑤ 치료율(cure rates) ( $\theta_1, \theta_2$ ) :

( 0.1, 0.1 ), ( 0.2, 0.2 ), ( 0.5, 0.5 ), ( 0.1, 0.3 ), ( 0.2, 0.4 ), ( 0.4, 0.6 )

- ⑥ 중도절단을(censoring rates) : 0%, 25%, 50%

- ⑦ 제1종오류분배율함수( Type I error spending rate function ) :

$\alpha_1^*$  ( O'Brien- Fleming형태 ),  $\alpha_2^*$  ( Pocock 형태 ),  $\alpha_3^*$  (=  $\alpha t$ )

<분포1>-<분포7>에 대해서 5% 유의수준을 가지고 양측검정을 200번 시행했다. 그러나 <분포1>에서와 같이 치료되지 않은 각 치료군 환자들의 조건부 분포가 같은 경우에는 유의수준 ( $\alpha$ )을 보다 정확히 비교하기 위해서 1000번을 시행하였다. 각각의 수치는 백분율이다. 위의 상황에서 전체 자료에 대한 치료율은 각 치료군마다 균일분포 ( $0, \theta_i$ ) ( $i=1,2$ )에 따라서 먼저 정해진 후에 나머지 치료되지 않은 환자의 자료에 대해 균일분포 ( $0, C_k$ ) ( $k=1,2,3$ )로 중도절단이 적용되었다. 한 치료군에 대해서 중도절단을 고정시켰으므로 다른 치료군에서의 중도절단은 치료율이 서로 다른 경우에는 주어진 값과 다르다.

<분포표> 각각의 표에서 사용된 두 치료군의 분포 및 중도절단이 고정된 치료군

|       | 치료1                       | 치료2                       | 중도절단이 고정된 치료군 |
|-------|---------------------------|---------------------------|---------------|
| <분포1> | exp(1)                    | exp(1)                    | 치료1           |
| <분포2> | exp(1)                    | 0.5exp(1)+0.5exp(2)       | 치료1           |
| <분포3> | exp(1)                    | exp(2)                    | 치료1           |
| <분포4> | 0.25exp(0.25) +0.75exp(1) | 0.25exp(2) +0.75exp(1)    | 치료1           |
| <분포5> | exp(2)                    | exp(1)                    | 치료2           |
| <분포6> | 0.25exp(2) +0.75exp(1)    | 0.25exp(0.25) +0.75exp(1) | 치료2           |
| <분포7> | 0.5exp(1)+0.5exp(2)       | exp(1)                    | 치료2           |

## 4.2 모의실험의 결과 및 분석

### (1) 두 처리군의 조건부분포가 같은 경우( $H_1 = H_2$ )

<분포1>의 경우에는 두 처리군의 조건부분포가 같으므로 치료율이 같다는 귀무가설은 결국 두 처리군의 생존분포가 같다는 것을 의미한다. 따라서 생존분포의 동일성을 검정하는 GT, LR, GW 통계량들의 유의수준이 5% 근처이다. 검정력은 GT와 LR 통계량의 검정력이 GW 통계량에 비해서 약간씩 크다. 또한 치료율과 중도절단율이 낮아지면 검정력이 커진다. 여기서 Kuk and Chen(1992)에서와 같이 지수분포의 혼합형태를 사용하여도 <분포1>의 결과와 전반적인 양상이 일치한다 (cf. 박경미, 1995).

### (2) 두 처리군의 조건부분포가 다른 경우( $H_1 \neq H_2$ )

<표2>, <표3>의 경우에는 두 처리군의 조건부분포가 다르므로 치료율이 같다는 것이 생존분포의 동일성을 의미하는 것은 아니다. 두 표를 보면 전반적으로 유의수준이 상당히 큰데 (심한 경우는 60%에 이르는 것도 있다) 그 이유는 우리가 설정한 귀무가설이 '두 처리군의 치료율은 같다'라는 것인데 반해 비모수적인 방법에 근거한 통계량 GT, LR, GW들은 '두 처리군의 생존분포의 동일성'을 검정하도록 고안되어진 통계량들이므로 이런 결과가 나타난 것으로 생각할 수 있을 것이다. 따라서 이러한 경우에 GT, LR, GW 통계량들의 검정력은 타당성을 가질 수 없으므로 임상시험에 대한 검정 결과도 신뢰할 수 없다. 다만 두 처리군의 분포가 비슷하고 치료율이 50% 이상이면(예로 <분포2>의 두 처리군의 치료율이 같으면서 높은 경우) 유의수준이 5% 근처이므로 세 가지 통계량 모두 두 처리군의 치료율 동일성 검정에 사용하기에 적절하다.

두 처리군의 생존분포함수가 교차하는 <분포2>, <분포3>, <분포4>의 경우는 아동 암(pediatric cancer)에 관한 임상시험에서 자주 볼 수 있다. 예를 들면 암 환자의 치료시에 골수이식 수술 방법은 초반에 새로 이식된 골수의 부작용 때문에 생존할 확률이 떨어지다가 시간이 지날수록 완치율이 높아지지만 화학적 치료는 초반에 골수이식 수술방법보다 생존할 확률이 높지만 결국에는 완치율이 떨어진다. 따라서 두 치료방법의 생존분포함수는 서로 교차하게 된다. <분포2>, <분포3>, <분포4>에 대해서 중도절단율이 25%인 경우를 비교한 <표2>를 보면 전반적으로 검정력이 낮다. 그리고 GT 통계량의 검정력이 LR, GW 통계량에 비해 좋고 두 처리군의 생존분포함수가 많이 교차할수록(<분포2>, <분포3>, <분포4>의 순서로) 검정력이 높아진다. 여기서 중도절단율이 0%, 50%일때에는 전반적인 양상이 <표2>와 비슷하고 다만 <분포2>와 <분포3>의 검정력은 중도절단율이 낮아지면 좋아진다. 그러나 <분포4>는 검정력의 변화가 거의 없다 (cf. 박경미, 1995).

두 처리군의 생존분포함수가 교차하지 않는 <분포5>, <분포6>, <분포7>에 대해서 중도절단율이 25%인 경우를 비교한 <표3>을 보면 전체적으로 검정력이 높다. 그리고 LR 통계량의 검정력이 GT, GW 통계량에 비해 좋고 두 처리군의 생존분포함수가 비례위험모형에 가까울수록(<분포5>, <분포6>, <분포7>의 순서로) 검정력이 높다. 여기서 <표3>과 비교하여 중도절단율

이 0%, 50%인 경우에는 전반적으로 동일한 결과가 나오며 각 분포의 검정력도 변화가 없다 (cf. 박경미, 1995).

(3) 오류분배율함수의 사용에 따른 경우

오류분배율함수의 사용에 따른 검정력의 차이를 살펴보면 거의 모든 경우에 O'Brien-Fleming 형태의 분배율 함수 ( $\alpha_1^*$ )가 Pocock 형태의 함수 ( $\alpha_2^*$ )와  $\alpha_3^*(=at)$ 를 사용하였을 때에 비해서 좋은 검정력을 주고 있다. 그러나 그 차이는 중도절단이 높아지거나 치료율이 높아지면 작아지는 경향을 보인다. 결론적으로, 통계량의 선택에 관계없이 보수적인 오류분배율 함수 (다시말해서,  $\alpha_1^* > \alpha_3^* > \alpha_2^*$ )를 선택하거나 분석시점을 늦추는 정보비율(information fraction)을 선택한다면 좀더 좋은 검정력을 얻을수 있을 것으로 추측할 수 있다 (참고로, 모의 실험에서 사용되고 있는 정보비율은  $(t_1, t_2, t_3) = (1/3, 2/3, 1)$ 이다).



<표1> 분포1의 경우에 5% 양측검정으로 모의실험한 검정력 비교(단위:%)

(처리1 : exp(1), 처리2 : exp(1))

|       |       | 치료율 ( $\theta_1, \theta_2$ ) |              |              |              |              |              |
|-------|-------|------------------------------|--------------|--------------|--------------|--------------|--------------|
|       |       | (.1,.1)                      | (.2,.2)      | (.5,.5)      | (.1,.3)      | (.2,.4)      | (.4,.6)      |
| 중도질단율 | 검정통계량 | $\alpha_1^*$                 | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ |
|       |       | $\alpha_2^*$                 | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ |
|       |       | $\alpha_3^*$                 | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ |
| 0%    | GT    | 4.7                          | 6            | 3.6          | 61.5         | 50           | 41           |
|       |       | 4.8                          | 4.8          | 4.4          | 46.5         | 36           | 31           |
|       |       | 4.7                          | 5.1          | 4.4          | 49.5         | 39           | 36.5         |
|       | LR    | 5.4                          | 5.3          | 4.4          | 51.5         | 46.5         | 43           |
|       |       | 4.7                          | 4.6          | 4.6          | 38           | 35           | 32.5         |
|       |       | 5                            | 4.6          | 4.5          | 41.5         | 39           | 36           |
|       | GW    | 4.7                          | 4.5          | 4.8          | 37.5         | 39           | 39           |
|       |       | 4.1                          | 4.7          | 4.6          | 29.5         | 27           | 31.5         |
|       |       | 4.3                          | 4.6          | 4.6          | 31.5         | 31           | 35           |
| 25%   | GT    | 5.3                          | 5.3          | 5.1          | 50.5         | 42           | 36           |
|       |       | 4.4                          | 4.6          | 4.8          | 35           | 30           | 30           |
|       |       | 4.3                          | 4.6          | 5.2          | 39.5         | 33.5         | 30           |
|       | LR    | 4.6                          | 5.9          | 5.4          | 46           | 38           | 33.5         |
|       |       | 4.4                          | 4.5          | 5.2          | 34.5         | 28.5         | 29           |
|       |       | 4.7                          | 4.9          | 5.6          | 39           | 31.5         | 29.5         |
|       | GW    | 4.2                          | 4.9          | 5.3          | 33.5         | 31           | 31.5         |
|       |       | 4.3                          | 4.5          | 5.3          | 24.5         | 23.5         | 27.5         |
|       |       | 4.6                          | 4.8          | 5            | 29.5         | 25.5         | 29           |
| 50%   | GT    | 5.3                          | 5.3          | 5.8          | 35.5         | 29           | 30           |
|       |       | 4.8                          | 4.3          | 5.5          | 24           | 20           | 25.5         |
|       |       | 4.7                          | 4.5          | 5.7          | 28.5         | 23           | 25.5         |
|       | LR    | 5.6                          | 4.6          | 5.8          | 32           | 28.5         | 29.5         |
|       |       | 5.2                          | 4.7          | 5.4          | 25.5         | 19.5         | 25.5         |
|       |       | 4.8                          | 4.6          | 5.5          | 29.5         | 22           | 26.5         |
|       | GW    | 5.1                          | 4.3          | 5.5          | 23.5         | 20           | 27           |
|       |       | 5.3                          | 4.3          | 5.5          | 24.5         | 17.5         | 25           |
|       |       | 4.9                          | 4.5          | 4.9          | 21.5         | 18.5         | 27           |

<표2> 5% 양측검정일때 모의실험에 의한 검정력 비교(단위:%)

( $H_1 \neq H_2$ , 교차하는 경우)

|              |       | 치료율 ( $\theta_1, \theta_2$ ) |              |              |              |              |              |
|--------------|-------|------------------------------|--------------|--------------|--------------|--------------|--------------|
|              |       | (.1,.1)                      | (.2,.2)      | (.5,.5)      | (.1,.3)      | (.2,.4)      | (.4,.6)      |
| 중도절단을<br>25% | 검정통계량 | $\alpha_1^*$                 | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ |
|              |       | $\alpha_2^*$                 | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ |
|              |       | $\alpha_3^*$                 | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ |
| <분포2>        | GT    | 11.5                         | 8.5          | 6.5          | 34           | 29.5         | 21.5         |
|              |       | 15                           | 12.5         | 6.5          | 22.5         | 15           | 15           |
|              |       | 14.5                         | 11           | 6.5          | 27           | 20           | 17.5         |
|              | LR    | 14.5                         | 12.5         | 5            | 20.5         | 18.5         | 20.5         |
|              |       | 14                           | 15           | 5.5          | 9            | 13.5         | 17           |
|              |       | 14                           | 14.5         | 5.5          | 17.5         | 18.5         | 17           |
|              | GW    | 16                           | 14.5         | 5.5          | 9.5          | 11           | 17.5         |
|              |       | 13.5                         | 16           | 4.5          | 9            | 10.5         | 11           |
|              |       | 13.5                         | 14.5         | 5            | 9            | 11           | 13           |
| <분포3>        | GT    | 52.5                         | 40           | 18           | 22           | 16.5         | 17.5         |
|              |       | 58.5                         | 46.5         | 26.5         | 23.5         | 16           | 13           |
|              |       | 57                           | 45           | 23           | 21.5         | 15.5         | 15           |
|              | LR    | 61.5                         | 26           | 20           | 12           | 8.5          | 15           |
|              |       | 61                           | 31.5         | 25.5         | 17           | 12.5         | 11           |
|              |       | 61                           | 31           | 25           | 16           | 11.5         | 13           |
|              | GW    | 65                           | 32           | 23           | 9            | 5            | 10           |
|              |       | 62.5                         | 34.5         | 26.5         | 16.5         | 11           | 10.5         |
|              |       | 64                           | 33.5         | 25           | 14           | 9            | 10           |
| <분포4>        | GT    | 30.5                         | 21.5         | 14.5         | 14.5         | 13.5         | 20           |
|              |       | 27                           | 20           | 45.5         | 10.5         | 8.5          | 14.5         |
|              |       | 27.5                         | 19.5         | 14           | 12           | 10           | 17.5         |
|              | LR    | 32.5                         | 17           | 12.5         | 7.5          | 11           | 17.5         |
|              |       | 27                           | 17           | 11           | 7.5          | 8            | 15           |
|              |       | 29                           | 16           | 11.5         | 7            | 8            | 15           |
|              | GW    | 31                           | 19           | 13.5         | 7.5          | 8            | 16           |
|              |       | 27                           | 18           | 11.5         | 6.5          | 7            | 12           |
|              |       | 27.5                         | 18.5         | 11           | 7.5          | 7            | 13.5         |

<표3> 5% 양측검정일때 모의실험에 의한 검정력 비교(단위:%)

( $H_1 \neq H_2$ , 교차하지 않는 경우)

|              |       | 치료율 ( $\theta_1, \theta_2$ ) |              |              |              |              |              |
|--------------|-------|------------------------------|--------------|--------------|--------------|--------------|--------------|
|              |       | (.1,.1)                      | (.2,.2)      | (.5,.5)      | (.1,.3)      | (.2,.4)      | (.4,.6)      |
| 중도절단을<br>25% | 검정통계량 | $\alpha_1^*$                 | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ | $\alpha_1^*$ |
|              |       | $\alpha_2^*$                 | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ | $\alpha_2^*$ |
|              |       | $\alpha_3^*$                 | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ | $\alpha_3^*$ |
| <분포5>        | GT    | 52.5                         | 40           | 18           | 92.5         | 85.5         | 73           |
|              |       | 58.5                         | 46.5         | 25.5         | 90.5         | 81.5         | 73.5         |
|              |       | 57                           | 45           | 23           | 91           | 84.5         | 75           |
|              | LR    | 61.5                         | 26           | 20           | 95           | 88           | 75.5         |
|              |       | 61                           | 31.5         | 25.5         | 92.5         | 85.5         | 76           |
|              |       | 61                           | 31           | 25           | 95           | 87.5         | 77.5         |
|              | GW    | 65                           | 32           | 23           | 93           | 92           | 77.5         |
|              |       | 62.5                         | 34.5         | 6.5          | 92.5         | 86.5         | 75.5         |
|              |       | 64                           | 33.5         | 25           | 92.5         | 90.5         | 77.5         |
| <분포6>        | GT    | 30.5                         | 21.5         | 14.5         | 86           | 74           | 66           |
|              |       | 27                           | 20           | 15.5         | 77.5         | 66.5         | 62           |
|              |       | 27.5                         | 19.5         | 14           | 80           | 68           | 63           |
|              | LR    | 32.5                         | 17           | 12.5         | 86           | 77.5         | 68           |
|              |       | 27                           | 17           | 11           | 78           | 69           | 62.5         |
|              |       | 29                           | 18           | 11.5         | 81           | 73           | 65.5         |
|              | GW    | 31                           | 19           | 13.5         | 76           | 73.5         | 69           |
|              |       | 27                           | 18           | 11.5         | 67.5         | 67.5         | 62.5         |
|              |       | 27.5                         | 18.5         | 11           | 73           | 70           | 65.5         |
| <분포7>        | GT    | 11.5                         | 8.5          | 6.5          | 69           | 65.5         | 46.5         |
|              |       | 15                           | 12.5         | 6.5          | 64.5         | 65.5         | 44.5         |
|              |       | 14.5                         | 11           | 6.5          | 66           | 61           | 44           |
|              | LR    | 14.5                         | 12.5         | 5            | 74           | 69           | 50           |
|              |       | 14                           | 15           | 5.5          | 68.5         | 59           | 48           |
|              |       | 14                           | 14.5         | 5.5          | 70           | 61           | 44.5         |
|              | GW    | 16                           | 14.5         | 5.5          | 70           | 65           | 50           |
|              |       | 13.5                         | 16           | 4.5          | 64           | 55           | 48.5         |
|              |       | 13.5                         | 14.5         | 5            | 68           | 58.5         | 50           |

## 5. 결론

임상시험은 환자들을 대상으로 하는만큼 사전에 충분한 시간과 노력을 기울여 도중에 발생할 수 있는 오류들을 최소화 해야한다. 이런 관점에서 볼 때 미리 분석시점, 분석 횟수, 제1종오류의 분배등을 계획한 후에 자료를 관찰하고 여러번의 중간분석을 실시하는 집단축차검정법은 바람직한 분석방법이라 할 수 있다.

집단축차검정법을 적용한 비모수 통계량을 사용하여 치료율을 비교한 본문의 모의실험 결과를 참조해 보면 임상시험의 다양한 상황에 맞는 적절한 통계량을 선택할수 있을 것이다. 또한 오류분배율함수와 정보비율의 선택도 검정력에 영향을 미친다는 것을 추측할 수 있다. 그러나 일반적으로 사용되는 Gray-Tsiatis, Logrank, Gehan 통계량 등은 두 처리군의 조건부분포가 서로 다를때 유의수준이 너무 커지므로 검정결과를 신뢰할 수 없는 경우가 있다. 그러므로 본문의 <표2>와 <표3>을 참고하여 적절한 통계량을 신중하게 고려하여 선택해야한다.

## 참고문헌

- [1] 박경미 (1995). 임상실험에서 치료율 비교를 위한 집단축차검정법에 관한 고찰, 석사학위논문, 고려대학교 통계학과.
- [2] Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors, *Biometrics* 38, 1041-10467.
- [3] Gray, R. J., and Tsiatis, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates, *Biometrics* 45, 899-904.
- [4] Jennison, C., and Turnbull, B. W. B. (1989). Interim analyses : the repeated confidence interval approach, *Journal of royal statistical society* 51, 305-361.
- [5] Kuk, A. Y. C., and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression, *Biometrika* 79, 531-41.
- [6] Lan, K. K. G., and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials, *Biometrika* 70, 659-663.
- [7] Laska, E. M., and Meisner, M. J. (1992). Nonparametric estimation and testing in a cure model, *Biometrics* 48, 1223-1234.
- [8] Lee, J. W. (1994-a). Group sequential testing in clinical trials with multivariate observations: A review. *Statistic in Medicine* 13, 101-111.
- [9] Lee, J. W. (1994-b). An overview of group sequential procedures, *The Korean Journal of Applied Statistics* 7, No. 2, 35-51.
- [10] Lee, J. W., and DeMets, D. L. (1991). Sequential comparison of changes with repeated measurements data, *Journal of the American Statistical Association* 86, 757-762.
- [11] Lee, J. W., and DeMets, D. L. (1992). Sequential rank tests with repeated measurements in clinical trials, *Journal of the American Statistical Association*

87, 136-142.

- [12] Lee, J. W., and Sather, H. N. (1995). Group sequential models for comparison of cure rates in clinical trials, *Biometrics* 51, 756-763.
- [13] Lin, D. Y. (1991). Nonparametric sequential testing in clinical trials with incomplete multivariate observation, *Biometrika* 78, 123-131.
- [14] O'Brien, P. C., and Fleming, T. R. (1979). A multiple testing procedure for clinical trials, *Biometrics* 35, 549-556.
- [15] Peto, R., and Peto, J. (1972). Asymptotically efficient rank invariant procedures(with discussion), *Journal of the Royal Statistical Society, Ser.A* 135, 185-206.
- [16] Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials, *Biometrika* 64, 191-199.
- [17] Prentice, R. L. (1978). Linear rank tests with right censored data, *Biometrika* 65, 167-179.

## Group Sequential Testing Methods for Comparing Cure Rates<sup>4)</sup>

Kyung Me Park<sup>5)</sup> , Jae Won Lee<sup>6)</sup>

### Abstract

There are many clinical trials where a large portion of the patients are cured of disease. In such a case, one might be more interested in testing the differences in cure rates rather than other types of differences in failure distribution. For ethical and economic reasons, clinical trials must be repeatedly monitored for evidence of treatment benefit or harm. In this article, we examined by simulation the properties of nonparametric group sequential methods for comparing the cure rates between two treatment groups during the trial in a wide range of alternatives, censoring rates and cure rates.

- 
- 4) This research was supported by grant 94-1300-01-01-3 from the Korea Science and Engineering Foundation
- 5) Researcher, Science and Technology Policy Institute, P.O.Box 255, Cheongryang, Seoul 130-650, Korea
- 6) Associate Professor, Department of Statistics, Korea University, 5-1 Anam-dong Sungbuk-gu, Seoul 136-701, Korea