

Confidence Bounds following Adaptive Group Sequential Tests with Repeated Measures in Clinical Trials

Sook Jung Jo^a · Jae Won Lee^{b,1}

^aGlobal Biostatistics, ICON Clinical Research; ^bDepartment of Statistics, Korea University

(Received February 19, 2013; Revised July 2, 2013; Accepted August 5, 2013)

Abstract

A group sequential design can end a clinical trial early if a confirmed efficacy or a futility of study medication is found during clinical trials. Adaptation can adjust the design of clinical trials based on accumulated data. The key to this methodology is considered to control the overall type 1 error rate while maintaining the integrity of clinical trials. The estimation would be more complex and the sample size calculation will be more difficult if the clinical trials have repeated measurement data. Lee *et al.* (2002) suggested a repeated observation case by using the independent increments properties of the interim test statistics and investigated the properties of the proposed confidence interval based on the stage-wise ordering. This study extend Lee *et al.* (2002) to adaptive group sequential design. We suggest test statistics for the adaptation as redesigning the second stage of clinical trials and induce the stage-wise confidence interval of parameter of interests. The simulation will help to confirm the suggested method.

Keywords: Confidence bounds, Adaptive group sequential test, stage-wise ordering.

1. 서론

인간에게 새로운 치료제나 치료법을 적용하여 그 효과를 입증하는 연구 과정을 임상시험이다. 일반적으로 임상시험 단계에 있는 치료제나 치료법은 선행된 세포시험과 동물시험의 비임상시험기간 동안 치료효과에 긍정적인 결과를 보인 것들이다. 임상시험은 새로운 방법을 인간에게 최초로 적용하는 단계인 만큼 안전성을 우선적으로 관찰할 수 있도록 하는 것이 최근의 추세이다.

또한 불확실한 가정에 근거한 설계로 인한 임상시험의 실패확률이 높아지는 단점을 보완하기 위하여, 임상시험 도중에 중간점검을 통하여 최초 설계를 수정하거나 임상시험을 조기에 종료하는 적응적 설계법(Adaptive Design)이 제안되어 왔다. 적응적 설계법은 일정시점까지 누적된 자료에 근거하여, 임상시험을 조기에 종료하거나, 연구계획서를 변경하는 등 다양한 설계법을 포함한다. 예를 들면, 임상시험에 등재할 피험자의 수를 재계산할 수도 있고, 효과가 없는 투여군을 제외할 수도 있다. 중간에 누적되

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2012R1A1A2008686).

¹Corresponding author: Professor, Department of Statistics, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 136-710, Korea. E-mail: jael@korea.ac.kr

자료로 중간분석(interim analysis)을 실시하여 유효성/무용성 또는 독성이 조기에 입증되는 경우 임상 시험을 중단할 수 있는 집단축차설계법(group sequential design)도 대표적인 적응설계법이라고 할 수 있다.

적응설계에서 가장 중요하게 고려되는 요소 중 하나는 연구 전체의 제1종 오류율(experimental error rate)을 잘 통제하는 것이다. Bauer와 Kohne (1994)과 Proschan과 Hunsberger (1995)는 2단계 임상 시험을 제안하였고, 이 후 Fisher (1998), Shen와 Fisher (1999), Cui 등 (1999), Lehman과 Wassmer (1999), Denne (2001), Posch 등 (2003), Cheng와 Shen (2004) 등은 2단계 임상시험에서 확장된 다단계 임상시험 설계법을 제안하였다. Muller와 Shafer (2001)는 Bauer와 Kohne (1994)의 방법에 집단축차검정법(group sequential test)을 도입하여 2단계 이상의 임상시험을 할 수 있도록 하였고, Muller와 Shafer (2004)는 그 방법을 일반화하였다. Brannath 등 (2007)은 적응적 집단축차검정 후, 관심모수에 대한 단계적 순서관계를 고려한(stage-wise)신뢰구간 추정방법을 제안한 바 있다.

또한, 최근 임상시험에서 환자는 반복적으로 관측치를 남기는 것이 일반적이다. 이처럼 장기간에 걸쳐 반복적인 측정치를 가지는 경우에 관심모수의 통계적 검정 및 추정 시에는 반복측정치 내의 변이를 포함한 통계적 모형을 고려하여야 할 것이다.

본 논문에서는 반복측정자료를 가지는 적응적 집단축차설계 임상시험에서, 임상시험 종료 후 신뢰구간을 추정하는 방법을 제안하였다. 반복측정치를 가지는 경우에 적응적집단축차설계법에서 적용할 수 있는 검정통계량과 임상시험이 종료된 후에 단계적 순서관계를 고려한 신뢰구간의 추정법을 유도하였고, 모의실험을 통하여 신뢰구간의 안정성을 증명하였다.

2. 단일관측치의 임상시험 종료 후 신뢰구간 추정

Brannath 등 (2009)은 두 모집단의 모평균을 비교하는 적응적집단축차설계법을 적용한 임상시험에서 임상시험이 종료된 후의 신뢰구간을 추정하는 방법을 제안하였다. Muller와 Schafer (2001, 2004)가 제안한 적응적 집단축차설계법에 Tsiatis 등 (1984)의 단계별 순서관계(stage-wise ordering)의 개념을 도입하였다.

최대분석횟수가 K 이고 두집단의 모평균인 μ_t, μ_c 을 비교하는 집단축차검정법을 고려하여보자. 관심모수가 $\delta = \mu_t - \mu_c$ 이면, k ($k = 1, \dots, K$)번째 중간분석에서의 검정통계량은 $S_k = \hat{\delta}\sqrt{I_k}$ 이고, 이때 $\hat{\delta}$ 는 모수 δ 의 최대우도추정량이고 정보량 $I_k = \{se(\hat{\delta})\}^{-2}$ 이 된다. 이 때 (S_1, \dots, S_K) 는 평균이 $E[S_k] = \delta\sqrt{I_k}$, $k = 1, \dots, K$ 인 다변량정규분포를 하고, $k < l$ 에 대하여 $Cov(S_k, S_l) = \sqrt{I_k/I_l}$ 이 된다. (S_1, \dots, S_K) 는 Jennison과 Turnbull (2000)에서 제안한 세가지 조건을 만족하여 표준결합분포(canonical joint distribution)를 따르게 되므로 k 번째 중간분석에서의 기각치 $a_k, k = 1, \dots, K$ 는 귀무가설이 $H_0 : \delta < 0$ 하에서 $\sum_{k=1}^K \pi_k = \alpha$ 일 때 $P_0\{S_1 \leq a_1, \dots, S_k > a_k\} = \pi_k$ 를 만족하도록 정할 수 있다. 따라서, 집단축차설계에서 계산된 기각치 a_1, \dots, a_K 는 귀무가설하에서 $P_0\{\cup_{k=1}^K (S_k \geq a_k)\} = \alpha$ 를 만족한다.

일반적인 귀무가설 $H_h : \delta < h$ 를 고려하여 보자. 만약 $T (< K)$ 단계의 중간분석결과 임상시험이 종료되었고 T 시점에서 검정통계량 S_T 의 관측값을 s_T 라고 한다면 T 시점에서의 단측유의확률은

$$p(h) = P_h \left\{ \bigcup_{k=1}^{T-1} (S_k \geq a_k) \cup (S_T \geq s_T) \right\} \quad (2.1)$$

와 같이 표현할 수 있으며, 여기서 $P_h(\cdot)$ 는 $\delta = h$ 에서의 확률이다. 본 시험의 유의수준을 α 라고 하면 유의확률 $p(h) < \alpha$ 일 때 귀무가설은 기각된다. 반대로 $C_\alpha = \{h : p(h) > \alpha\}$ 에 포함되는 h 에 대해서

는 귀무가설을 기각하지 못하므로 C_α 는 δ 에 대한 $100 \times (1 - \alpha)\%$ 신뢰구간의 의미를 가진다. 이 때, $p(h) = \alpha$ 를 만족하는 h 를 신뢰구간의 하한 δ^L 로 정하면, (δ^L, ∞) 는 모수 δ 에 대한 단측 $100 \times (1 - \alpha)\%$ 신뢰구간이 된다. 이는 표현은 다르지만 Tsiatis 등 (1984)와 Kim과 DeMets (1987)이 제안한 신뢰구간의 하한과 의미상 동일한 결과이다.

$L (< T \leq K)$ 번째 중간분석에서 적응법이 시행된 임상시험에서, 임의의 모수 h 에 대하여 종료시점 $\tau = \inf\{t : S_t > a_t\}$ 에서 적당한 기각치 $a_\tau(h)$ 가 존재한다고 가정하면, 기각역 $R(h)$ 는

$$R(h) = \left\{ \bigcup_{j=1}^{\tau-1} (S_j > a_j) \cup (S_\tau > a_\tau(h)) \right\} \quad (2.2)$$

와 같이 표현할 수 있다. 따라서, $P_h\{R(h)\} = \alpha$ 를 만족하는 기각치 $a_\tau(h)$ 를 찾으면 해당 임상시험의 기각역 $R(h)$ 가 정의된다.

L 번째 중간분석에서 적응법을 시행한 경우를 가정하여 보자. 귀무가설이 $\delta = 0$ 일 때 k 번째 검정통계량이 S_k ($k = 1, \dots, K$)이고 각 중간분석 단계에 대하여 미리 계산된 기각치가 a_k ($k = 1, \dots, K$)일 때, L 번째 중간분석이 종료된 후의 조건부 기각확률은

$$\epsilon = P_0 \left\{ \bigcup_{j=1}^K (S_j \geq a_j) \mid S_j = s_j, j \leq L \right\}$$

로 표현할 수 있다. 여기서 s_j 는 j 번째 중간분석에서의 관측값이다. 종료시점을 τ 라고 가정하면, 관심 모수의 참 값 h 는 구간 $\delta_t \leq h < \delta_{t-1}$ 에 있음을 예상할 수 있고, 종료시점은 $t-1 < \tau < t$ 에 존재할 것이다. 이 때의 t 를 $\tau(h)$ 로 두면 각각의 $H_h : \delta \leq h$ 에 대한 유의수준 α 검정에 대하여

$$\begin{aligned} P_h\{R(h)\} &= P_h \left\{ \bigcup_{j=1}^{\tau(h)-1} (S_j > a_j) \cup (S_{\tau(h)} > a_\tau(h)) \right\} \\ &= \alpha \end{aligned}$$

를 만족하도록 $a_\tau(h)$ 를 계산할 수 있다. 여기서, L 번째 중간분석에서 적응법을 적용한 것을 가정한 조건부 기각확률 $\epsilon(h)$ 는

$$\epsilon(h) = \begin{cases} 0, & \text{if } h \geq \delta_L, \\ P_h\{R(h, L) \mid S_L = s_L\}, & \text{if } h < \delta_L \end{cases} \quad (2.3)$$

으로 표현되며, L 번째 중간분석에서 적응법으로 2단계 임상시험을 재설계할 때 제1종오류로 사용할 수 있다. 여기서, $R(h, L) = \bigcup_{j=L+1}^{\tau(h)-1} (S_j > a_j) \cup (S_{\tau(h)} > a_\tau(h))$ 이다. 또한, 재설계된 2단계 임상시험이 종료된 후의 유의확률은

$$p^{(2)}(h) = P_h \left\{ \bigcup_{k^{(2)}=1}^{T^{(2)}-1} (S_{k^{(2)}}^{(2)} \geq a_{k^{(2)}}^{(2)}) \cup (S_{T^{(2)}}^{(2)} \geq s_{T^{(2)}}^{(2)}) \right\} \quad (2.4)$$

으로 계산된다. 따라서, 1단계 임상시험을 조건으로 하고 확률적으로 독립인 2단계 임상시험이 종료된 후의 δ 에 대한 신뢰구간의 하한 δ^L 은

$$\delta^L = \inf \left\{ h : p^{(2)}(h) > \epsilon(h) \right\}$$

로 추정할 수 있다.

참고로, 수식에 사용된 위첨자는 2단계 임상시험을 의미한다. 예를 들면, 2단계 임상시험에 대하여 최대 분석 회수는 $K^{(2)}$ 회이고, $k^{(2)}(k^{(2)} = 1^{(2)}, 2^{(2)}, \dots, K^{(2)})$ 번째 중간분석에서의 검정통계량은 $S_{k^{(2)}}^{(2)}$, 기각치는 $a_{k^{(2)}}^{(2)}$ 이며, $T^{(2)}$ 번째 중간분석에서의 검정통계량이 기각치 $a_{T^{(2)}}^{(2)}$ 보다 크면 임상시험을 종료하게 되는 식이다.

3. 반복측정치를 가지는 임상시험 종료 후 신뢰구간 추정

본 절에서는 Brannath 등 (2009)의 방법을 반복측정자료 혹은 경시적자료를 가지는 모형으로 확장하여, 적응적 집단축차검정법 종료 후 관심모수에 대한 신뢰구간 계산방법을 유도하고자 한다.

Laird와 Ware (1982)의 선형혼합모형을 $m_{1k} + m_{2k}$ 개체 중 $q (= 1, 2)$ 치리를 받은 i 번째 개체 y_i 를

$$y_i = X_i\beta_q + Z_i b_i + e_i, \quad b_i \sim N(0, D), \quad e_i \sim N(0, \sigma^2 I), \quad i = 1, \dots, m_{1k} + m_{2k}$$

로 정의한다. 여기서, 미지의 임의효과(개체효과) b_i 의 공분산행렬 D 와 e_i 의 분산 σ^2 은 알려져 있으며,

$$X_i = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in} \end{pmatrix}, \quad i = 1, \dots, m_{1k} + m_{2k}; \quad \beta_1 = \begin{pmatrix} \beta_{01} \\ \beta_1 \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} \beta_{02} \\ \beta_2 \end{pmatrix}$$

로 가정한다. 또한 b_i 는 $r \times 1$ 벡터, Z_i 는 b_i 와 반응벡터 y_i 를 연결하는 $n_i \times r$ ($r \leq p$) 차원을 가지는 기지의 계획행렬이다. x_{ij} 는 i 번째 개체의 최초 방문시점부터 j 번째 방문한 시점까지의 시간이다.

일반적으로 선형혼합모형에서 귀무가설은 '두 집단의 기울기는 차이가 없다', 즉, $H_0 : \delta (= \beta_1 - \beta_2) = 0$ 이 된다. Lee와 DeMets (1991)에서는 k 번째 중간분석까지 누적된 자료로 일반화 최소제곱 추정량 $\hat{\beta}_1(k)$ 과 $\hat{\beta}_2(k)$ 를 추정하고, k 번째 중간분석에서의 검정통계량 $\{S(k) = \hat{\delta}(k)\sqrt{I_k}, k = 1, \dots, K\}$ 가 근사적으로 다변량정규분포를 따름을 증명하였다. 또한, 귀무가설 하에서 k 번째 중간분석에서의 기각치 a_k 는 $P_0\{|S_1| \leq a_1, \dots, |S_{k-1}| \leq a_{k-1}, |S_k| > a_k\} = \pi_k$ 를 만족하도록 축차적인 적분을 통하여 계산할 수 있음을 보였다.

만약 임상시험의 중간분석결과가 적응적 집단축차설계에서 사전에 적응법을 실시하기 위한 조건과 부합하지 않아 적응법을 시행하지 않고 다음 단계로 진행되었다고 가정하자. 귀무가설 $H_h : \delta < h$ 이고 최대 분석회수가 K 인 집단축차검정에서 임상시험이 종료된 $T(\leq K)$ 단계 이후 계산된 유의확률은 식 (2.1)이다. 유의수준이 α 라면 모수 δ 에 대한 단측 $100 \times (1 - \alpha)\%$ 신뢰구간의 하한 δ^L 은 $\inf\{h : p(h) > \alpha\}$ 를 만족하는 h 로 구할 수 있으며, 이는 Lee 등 (2002)에서 제안한 신뢰구간의 하한과 동일하다.

이번엔, $L(< T \leq K)$ 단계에서 적응법이 시행된 경우를 가정하면, 1단계 임상시험의 결과를 조건으로 한 조건부 기각역인 $R(h)$ 를 적절한 기각치 $a_r(h)$ 는 (2.2)와 같이 정의할 수 있다. 또한, L 번째 중간분석에서의 관측값이 s_L 인 경우 조건부 기각확률 $\epsilon(h)$ 는 (2.3) 처럼 단일관측치의 경우와 동일하게 유도할 수 있다. Brannath 등 (2009)은 단일관측치의 경우 계산된 조건부 기각확률 $\epsilon(h)$ 와 2단계 임상시험에서 계산된 유의확률을 이용하여 신뢰구간의 하한을 추정하였다.

그러나, 반복측정치를 가지는 임상시험의 경우는 단일관측치의 자료구조와는 달리 개체내 변이를 고려하여야 한다. 적응법을 고려한 L 번째 중간분석 시점까지 누적된 자료를 살펴보면, 어떤 환자는 계획된 임상시험 기간을 모두 채우고 임상시험을 종료하였을 수 있지만, 어떤 환자들에서는 임상시험에 등재되었으나 계획된 추적기간을 채우지 못하여 진행되고 있는 중일 수도 있다. 다시 말하면 일부 환자들에

서는 1단계 임상시험에서의 정보 (S_1, \dots, S_L) 와 2단계 임상시험에서의 정보 $(S_{L+1}, \dots, S_{L+K^{(2)}}) = (S_{1^{(2)}}^{(2)}, \dots, S_{K^{(2)}}^{(2)})$ 에 중복되어 포함되므로 1단계 임상시험과 2단계 임상시험은 확률적으로 독립이 아니다. 따라서 2단계 임상시험에서의 유의확률은 (2.4)와 같이 계산할 수 없다. 그러므로, 2단계 임상시험에서의 유의확률을 계산하기 위해서 1단계 임상시험과 확률적으로 독립인 확률변수의 유도가 필요하게 되었다.

Lee 등 (2002)에서 유도된 확률변수 $T_k = S_k \sqrt{I_k}$ 는 평균이 $\delta \sqrt{I_k}$ 이고 분산이 1인 다변량 정규분포를 하며 각 시점간에 독립증분을 가짐을 보였다. 이와 같은 성질을 이용하여 확률변수 $V_{k^{(2)}}$ 를 아래와 같이 정의할 때 다음의 정리가 성립된다.

정리 3.1 확률변수 $V_{k^{(2)}}, k^{(2)} = 1, \dots, K^{(2)}$ 를 다음과 같이 정의하면,

$$\begin{aligned} V_{k^{(2)}} &= \frac{T_{L+k^{(2)}} - T_L}{\sqrt{I_{L+k^{(2)}} - I_L}} \\ &= \frac{\sqrt{I_{L+k^{(2)}}} S_{L+k^{(2)}} - \sqrt{I_L} S_L}{\sqrt{I_{L+k^{(2)}} - I_L}} \end{aligned}$$

와 같이 정의하면, $V_{k^{(2)}}$ 는 1단계 임상시험과 확률적으로 독립이며, JENNISON과 TURNBULL (2000)이 제시한 세 가지 조건을 만족하여 표준결합분포를 따른다.

1. $(V_{1^{(2)}}, V_{2^{(2)}}, \dots, V_{K^{(2)}})$ 는 다변량 정규분포를 따르며,
2. $E[V_{k^{(2)}}] = \delta \sqrt{I_{k^{(2)}}^{(2)}}$,
3. $k^{(2)} < l^{(2)}$ 일 때 $\text{Cov}(V_{k^{(2)}}, V_{l^{(2)}}) = \sqrt{I_{k^{(2)}}^{(2)}/I_{l^{(2)}}^{(2)}}$.

증명: $\text{Cov}(T_L, T_{L+k^{(2)}} - T_L) = 0$ 이므로 1단계 임상시험으로부터의 확률변수인 T_L 과 $T_{L+k^{(2)}} - T_L$ 은 서로 독립이다. 따라서, $T_{L+k^{(2)}} - T_L$ 의 선형결합으로 표현된 $V_{k^{(2)}}$ 와 T_L 의 선형결합으로 표현된 S_L 은 서로 독립인 확률변수가 된다.

또한,

$$T_{L+k^{(2)}} = (T_{L+k^{(2)}} - T_L) + T_L$$

로 표현하면, 각 시점간에서 평가된 모수 δ 에 대한 정보량은

$$\begin{aligned} I_{T_{L+k^{(2)}}}(\delta) &= I_{T_{L+k^{(2)}} - T_L}(\delta) + I_{T_L}(\delta) \\ &= I_{V_{k^{(2)}}}(\delta) + I_{T_L}(\delta) \end{aligned}$$

의 관계에 있으므로, $V_{k^{(2)}}$ 에 대한 정보량을 $I_{k^{(2)}}^{(2)}$ 로 두면, $I_{k^{(2)}}^{(2)} = I_{L+k^{(2)}} - I_L$ 이 됨을 쉽게 알 수 있다. $V_{k^{(2)}}$ 에 대한 기대값은

$$\begin{aligned} E[V_{k^{(2)}}] &= E \left[\frac{T_{L+k^{(2)}} - T_L}{\sqrt{I_{L+k^{(2)}} - I_L}} \right] \\ &= \frac{\delta I_{L+k^{(2)}} - \delta I_L}{\sqrt{I_{L+k^{(2)}} - I_L}} \\ &= \delta \sqrt{I_{k^{(2)}}^{(2)}}, \quad k^{(2)} = 1^{(2)}, \dots, K^{(2)} \end{aligned}$$

이다.

$V_{k^{(2)}}$ 에 대한 분산은

$$\begin{aligned}\text{Var}[V_{k^{(2)}}] &= \text{Var}\left[\frac{T_{L+k^{(2)}} - T_L}{\sqrt{I_{L+k^{(2)}} - I_L}}\right] \\ &= \frac{\text{Var}(T_{L+k^{(2)}}) + \text{Var}(T_L) - 2\text{Cov}(T_{L+k^{(2)}}, T_L)}{I_{L+k^{(2)}} - I_L} \\ &= \frac{\text{Var}(T_{L+k^{(2)}}) - \text{Var}(T_L)}{I_{L+k^{(2)}} - I_L} \\ &= 1\end{aligned}$$

이다. 따라서, $V_{k^{(2)}}$ 는 평균이 $\delta\sqrt{I_{k^{(2)}}^{(2)}}$ 이고 분산이 1인 정규분포를 함을 알 수 있다.

Lee 등 (2002)에서 유도하였듯이 $k < l$ 일 때 $\text{Cov}(T_k, T_l) = \text{Var}(T_k) = I_k$ 를 이용하면, $k^{(2)} < l^{(2)}$ 에서 $V_{k^{(2)}}$ 와 $V_{l^{(2)}}$ 의 공분산은

$$\begin{aligned}\text{Cov}(V_{k^{(2)}}, V_{l^{(2)}}) &= \text{Cov}\left(\frac{T_{L+k^{(2)}} - T_L}{\sqrt{I_{L+k^{(2)}} - I_L}}, \frac{T_{L+l^{(2)}} - T_L}{\sqrt{I_{L+l^{(2)}} - I_L}}\right) \\ &= \frac{\text{Cov}(T_{L+k^{(2)}} - T_L, T_{L+l^{(2)}} - T_L)}{\sqrt{I_{L+k^{(2)}} - I_L}\sqrt{I_{L+l^{(2)}} - I_L}} \\ &= \frac{\text{Cov}(T_{L+k^{(2)}}, T_{L+l^{(2)}}) - \text{Cov}(T_{L+k^{(2)}}, T_L) - \text{Cov}(T_L, T_{L+l^{(2)}}) + \text{Var}(T_L)}{\sqrt{I_{L+k^{(2)}} - I_L}\sqrt{I_{L+l^{(2)}} - I_L}} \\ &= \frac{I_{L+k^{(2)}} - I_L}{\sqrt{I_{L+k^{(2)}} - I_L}\sqrt{I_{L+l^{(2)}} - I_L}} \\ &= \sqrt{\frac{I_{k^{(2)}}^{(2)}}{I_{l^{(2)}}^{(2)}}}, \quad 1 \leq k^{(2)} < l^{(2)} \leq K^{(2)}\end{aligned}$$

가 된다. □

이로써, $V_{k^{(2)}}$ 를 사용하면 1단계 검정통계량 S_L 과는 독립이고, 따라서 2단계 임상시험은 조건부 제 1종 오류 $\epsilon(h)$ 하에서 재설계 할 수 있다. 또한, $V_{k^{(2)}}$ 는 Jennison과 Turnbull (2000)이 제시한 세가지 조건을 만족하여 표준결합분포를 따르므로 2단계 임상시험의 기각치, $a_{k^{(2)}}^{(2)}$ ($k^{(2)} = 1^{(2)}, \dots, K^{(2)}$)는 $\pi_{1^{(2)}}^{(2)} + \dots + \pi_{K^{(2)}}^{(2)} = \epsilon(h)$ 를 만족하는 $\pi_{k^{(2)}}^{(2)}$ ($k = 1^{(2)}, \dots, K^{(2)}$)에 대하여 $P_0\{V_{1^{(2)}} \leq a_{1^{(2)}}^{(2)}, \dots, V_{k^{(2)}} > a_{k^{(2)}}^{(2)}\} = \pi_{k^{(2)}}^{(2)}$ 를 만족하도록 계산할 수 있다. 그리고, 2단계 임상시험이 종료된 후의 유의확률 $p^{(2)}$ 는

$$p^{(2)}(h) = P_h \left\{ \bigcup_{k^{(2)}=1^{(2)}}^{T^{(2)}-1} \left(V_{k^{(2)}} \geq a_{k^{(2)}}^{(2)} \right) \cup \left(V_{T^{(2)}} \geq v_{T^{(2)}}^{(2)} \right) \right\} \quad (3.1)$$

이 된다. 이제, $V_{k^{(2)}}$ 를 이용하여 (3.1)에서 계산된 2단계 임상시험의 유의확률과 조건부기각확률 $\epsilon(h)$ 의 관계로부터, 반복측정치를 가지는 임상시험에서 1단계 임상시험을 조건으로 하고 확률적으로 독립인 2단계 임상시험이 종료된 후의 δ 에 대한 신뢰구간의 하한, δ^L 은

$$\delta^L = \inf \left\{ h : p^{(2)}(h) > \epsilon(h) \right\}$$

로 추정할 수 있게 되었다.

본 절에서는 선형혼합효과모형의 결과로 전체를 나타냈지만, 반복측정치를 가지는 모형에서 검정통계량 S_k 로부터 확률적으로 독립증분을 가지는 확률변수 T_k 를 정의할 수 있는 모든 모형에는 본 논문에서 제안된 방법으로 적응적 집단축차검정을 실시하고 임상시험이 종료된 후의 단계적 순서관계를 고려한 단측 신뢰구간의 하한을 추정할 수 있다.

4. 모의실험

모의실험을 통해 단계별 순서관계를 고려한 신뢰구간의 성질을 알아보고, naive 신뢰구간, Mehta 등 (2007)의 반복신뢰구간(repeated confidence interval)과 비교하여 보았다. 2단계 임상시험이 종료된 후 naive 신뢰구간으로는 1단계 임상시험 종료 후 계산된 조건부기각확률과 초기 임상시험 설계 시 고려한 제 1종오류를 고려하여 각각

$$\begin{aligned} \text{naive1} : \hat{\delta}_k^{(2)} - Z_{\epsilon(h)} \sqrt{I_k^{(2)}} \\ \text{naive2} : \hat{\delta}_k^{(2)} - Z_{0.05} \sqrt{I_k^{(2)}} \end{aligned}$$

로 정하였다. 여기서, $\Phi(Z_\alpha) = \alpha$ 이고, $\Phi(\cdot)$ 는 표준정규분포의 누적확률분포함수이다.

Mehta 등 (2007)에서는 1단계 임상시험에서 집단축차검정의 j ($j = 1, \dots, L$)번째 단계의 중간분석 통계량 S_j 를 $h\sqrt{I_j}$ 만큼 위치이동시킨 검정통계량을 기각치 a_j 와 비교하는 검정법을 제안하였고, 기각역의 여집합을 모수에 대한 반복신뢰구간으로 정의하였다. 적응설계 후 2단계 임상시험에서는 $j^{(2)}$, $j^{(2)} = 1, \dots, K^{(2)}$ 번째 중간분석 통계량 $s_{j^{(2)}}^{(2)}$ 를 $h\sqrt{I_{j^{(2)}}}$ 만큼 위치이동시킨 검정통계량을 기각치 $a_{j^{(2)}}^{(2)}$ 와 비교하고, 이 후의 신뢰구간 추정은 1단계 임상시험과 동일한 방법으로 기각역의 여집합을 반복신뢰구간으로 정의한다. 반복신뢰구간은 계산법이 간단하지만 다소 과대추정하는 경향이 있는 것으로 알려져 있다.

모의실험에 사용된 자료는 폐경기에 있는 여성들에게 칼슘의 공급이 뼈의 밀도변화에 미치는 영향을 알아보는 시험결과이다. 시험군의 여성들에게는 칼슘을 공급하고, 대조군의 여성들에게는 위약(偽藥; placebo)을 공급하였다. 선형혼합모형에서는 칼슘을 공급한 처리군은 1년에 1%의 감소가 있는 것으로, 위약을 공급한 처리군은 $(1 + \delta)\%$ 의 감소가 있는 것으로 고려하였다. 양 투여군에는 1 : 1로 피험자가 배정되며 각 환자는 1년에 4회씩 뼈밀도를 측정하는 것을 3년간 지속하게 된다. 따라서, 모든 환자들로부터 최대 12번의 관측치를 가진다.

본 모의실험에서 환자는 매년 초에 계획된 환자의 50%를 모집하여 2년간 등재를 마치는 것으로 가정하였다. 따라서, 1단계 임상시험 설계에서 환자모집이 2년, 마지막으로 등재된 환자의 추적기간이 3년임을 감안하면 예상되는 전체 임상시험 기간은 4년이 된다. 매 연말에 중간분석을 시행하여, 최종분석을 포함하면 임상시험 기간동안 최대 4번의 분석이 계획된다. 1단계 임상시험에서 오류분배율 함수는 가장 보수적인 O'Brien and Fleming의 오류함수 $\alpha_1^*(t) = 2 - 2\Phi(z_{\alpha/2}/t)$ 를 선택하여 각 중간분석 단계에서의 기각치를 계산하였다.

이것을 수리적 모형으로 다시 써보면,

$$\begin{aligned} y_{ij} &= 100 + b_{i0} + (-1.0 + b_{i1})t_{ij} + e_{ij}, \quad i = 1, \dots, m_1; j = 1, \dots, 12, \\ y_{ij} &= 100 + b_{i0} + (-1.0 - \delta + b_{i1})t_{ij} + e_{ij}, \quad i = m_1 + 1, \dots, m_1 + m_2; j = 1, \dots, 12 \end{aligned}$$

의 선형혼합효과 모형으로 표현할 수 있다. 여기서, i 번째 개체에 대한 임의효과(random effect) 벡터 b_i 는

$$\begin{pmatrix} b_{i0} \\ b_{i1} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \right), \quad e_{ij} \sim N(0, \sigma^2)$$

의 분포를 하고, 오차항 e_i 는 $N(0, \sigma^2 I_{n_i})$ 를 따른다.

모의실험에서 개체 내 효과인 임의효과 벡터의 공분산행렬($D = (d_{11}, d_{12}, d_{21}, d_{22})$)은 (3, 1, 1, 4)으로 고정하고, 개체 간 분산인 σ^2 은 3, 0.3, 30의 세가지 경우를 고려하였다. 양 투여군의 기율기 차이인 δ 는 0.75%로 가정하였고 5%를 유의수준으로 정하였다. 선형혼합효과모형의 경우 피험자 수를 계산하는 방법이 닫힌 형태로 존재하지 않으므로 모의실험을 통하여 90%의 검정력을 얻을 수 있는 표본의 크기를 예측하였다. 자료의 분산구조와 오류분배를 함수에 따라 예측된 표본의 크기는 σ^2 이 3, 0.3, 30일 때 각각 160, 140, 200명으로 계산되었다. 참고로 피험자 수 추정을 위한 모의실험은 각 100회의 반복을 실시하였다. 2단계 임상시험에서 고려된 적응법(adaptation)으로는 피험자 수 재계산과 오류분배를 함수의 변경이다. 피험자 수 재계산은 중간분석결과에서 계산된 조건부기각확률과 조정된 기율기를 사용하였고, 재계산된 피험자 수는 최초 계획한 크기에서 감소하지는 않고 최대 2배까지 증가하는 것을 허용하였다. 조정된 기율기로는 0.75%와 중간분석 시점에서 추정된 기율기의 산술평균을 사용하였다. 그리고, 피험자 수 재계산은 역시 모의실험을 통하여 추정하였는데, 1단계 임상시험 종료 후 계산된 조건부기각확률 하에서 90% 검정력을 가질 수 있도록 정하였다. 2단계에서 고려된 오류분배를 함수는 Pocock의 형태(PK)와 Hwang 등 (1990)이 제안한 오류분배함수족(Hwang-Shih-DeCani family)에서 $\gamma = -4$ 인 경우(H-DC(-4))로 고려하였다. 참고로 언급된 각 오류분배함수는 다음과 같다.

$$\begin{aligned} \text{O'Brien and Fleming type : } \alpha_1^*(t) &= 2 - 2\Phi\left(\frac{z_{\frac{\alpha}{2}}}{t}\right), \\ \text{Pocock type : } \alpha_2^*(t) &= \alpha \ln\{1|(\exp - 1)t\}, \\ \text{Hwang-Shih-DeCani family : } \alpha_3^*(t) &= \alpha \frac{1 - \exp(-\gamma t)}{1 - \exp(-\gamma)}. \end{aligned} \quad (4.1)$$

유의수준이 5% 일때 최초실제에서 가정한 상기 3가지 방법의 오류분배를 함수는 Figure 4.1와 같다. Figure 4.1에서 확인할 수 있듯이 H-DC(-4) 형태는 O'Brien & Fleming 방법(OB-F)과 비교하여 연구 초반에서 비교적 덜 보수적이지만 Pocock의 방법과 비교하면 보수적인 성향을 가지고 있다.

집단축차검정에서 각 경우의 기각치는 정규성에 근거하여 계산하는 MULNOR (Schervish, 1984)를 이용하였고, 두 처리군의 기율기와 공분산 구조의 추정은 Lindstorm과 Bates (1988)의 알고리즘을 적용하였다. 적응법 이후의 단계적 순서관계를 고려한 신뢰하한의 추정에는 Brannath 등 (2009)이 제안한 α 흡수상수를 도입한 알고리즘을 사용하였다.

Table 4.1, Table 4.2, Table 4.3는 모의실험을 통하여 제안된 단계별 순서관계를 이용한 신뢰구간(stage-wise C.I.)과 Mehta 등 (2007)의 반복신뢰구간(repeated C.I.), 1단계 임상시험 종료 후 계산된 조건부 기각확률을 이용하여 쉽게 계산한 신뢰구간(naive1 C.I.)과 최초 임상시험에서 고려된 유의수준인 0.05를 이용하여 쉽게 계산한 신뢰구간(naive2 C.I.)의 4가지 방법으로 추정된 신뢰구간의 포함확률을 비교한 결과이다. 참고로, 각 모의실험에서 1000회 독립표본에서 각각 추정된 95% 단축 신뢰구간이 실제 모수를 포함한 확률인 포함확률에 대한 표준오차는 0.0148 이상이었다. 성공확률이 0.95인 베르누이 시행을 독립적으로 1000회 실시한다면 모수에 대한 95% 신뢰구간은 (0.936, 0.964)이므로 모의실험으로부터 계산된 포함확률이 해당 구간 내에 포함되면 95%의 신뢰도를 가진다고 볼 수 있다.

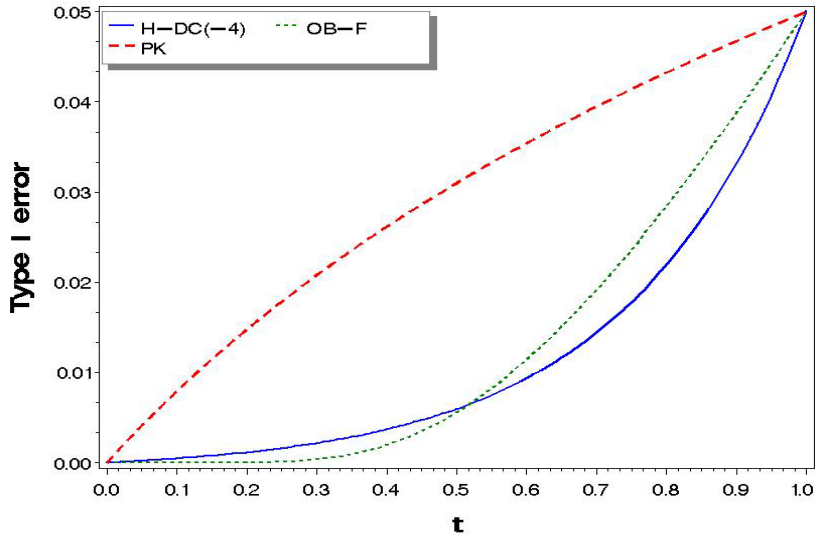


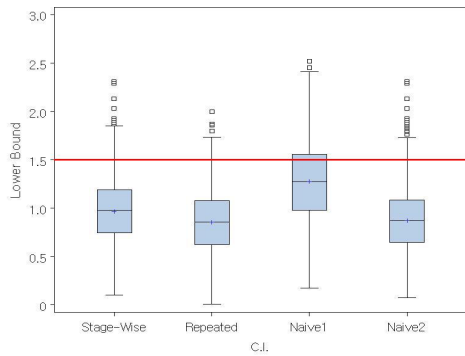
Figure 4.1. Type I error spending rate functions used in the simulation

Table 4.1. Coverage probability for 95% 1-sided confidence interval where $D = (3, 1, 1, 4)$ and $\sigma^2 = 3$

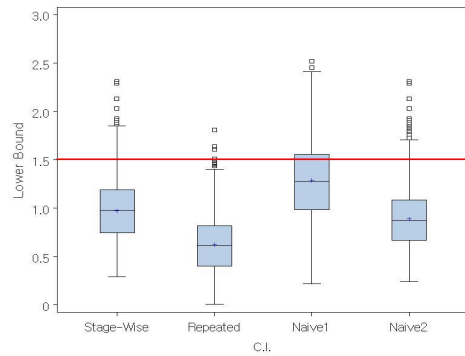
δ	PK				H-DC(-4)				
	stage-wise C.I.	repeated C.I.	naivel C.I.	naive2 C.I.	stage-wise C.I.	repeated C.I.	naivel C.I.	naive2 C.I.	
$L = 1$	1.5	0.952	0.972	0.702	0.974	0.946	0.996	0.683	0.978
	1	0.956	0.975	0.813	0.980	0.956	0.995	0.796	0.971
	0.75	0.954	0.974	0.851	0.971	0.954	0.992	0.829	0.964
	0.5	0.949	0.970	0.877	0.971	0.954	0.991	0.870	0.955
	0.25	0.950	0.966	0.892	0.956	0.948	0.988	0.880	0.942
$L = 2$	1.5	0.947	0.980	0.674	0.984	0.961	0.996	0.689	0.983
	1	0.953	0.991	0.949	0.987	0.943	0.994	0.934	0.999
	0.75	0.956	0.989	0.886	1.000	0.955	0.988	0.897	0.999
	0.5	0.933	0.994	0.834	0.999	0.934	0.983	0.814	1.000
	0.25	0.941	0.974	0.804	0.999	0.956	0.989	0.836	0.999

개체내 효과인 임의효과벡터의 공분산행렬 $D = (3, 1, 1, 4)$ 과 개체간 효과의 분산 σ^2 의 크기에 따른 모의실험 결과는 다음과 같다.

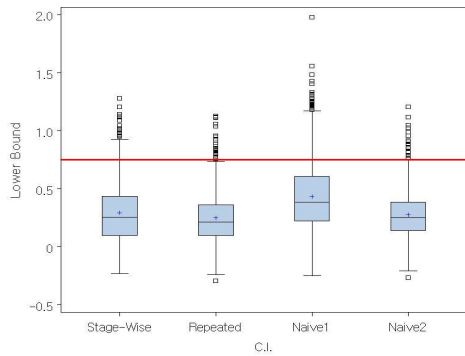
Table 4.1의 $d_{11}/\sigma^2 = 1$ 이고 $\sigma^2 = 3$ 인 경우의 결과로 개체 내 변동과 개체 간 변동의 크기가 유사한 경우이다. 첫번째 중간분석 시점에서 적응법을 실시한 경우($L = 1$)와 2번째 중간분석 시점에서 적응법을 실시한 경우($L = 2$) 두가지 결과이다. 적응법의 시점, 모수의 크기 및 2단계 임상시험에서 사용된 오류분배율 함수의 종류와 상관없이 단계적 순서를 고려한 신뢰구간의 포함확률은 신뢰계수인 0.95에 가까운 매우 안정적인 포함확률을 보이고 있다. 적응법의 시기를 비교하면 $L = 2$ 일 때 δ 가 작을수록 다소 과소추정하는 경향을 보였다. 반복신뢰구간의 포함확률은 과대추정하는 경향을 보였는데, 이는 Mehta 등 (2007)의 결과와 일치한다. 조건부 기각확률을 사용한 naive1 신뢰구간의 포함확률은 전체적으로 과소추정되었으며, 1단계 임상시험의 유의수준을 이용하여 naive2 신뢰구간의 포함확률은 과대추정되는 경향을 보였다.



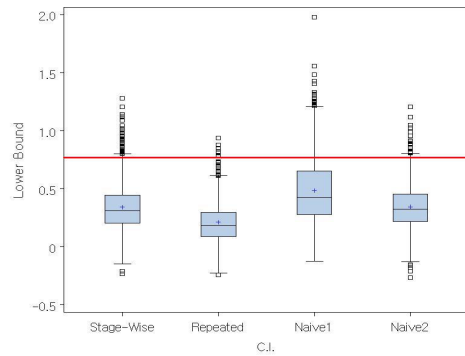
(a) $\sigma^2 = 3, \delta = 1.5$, PK type



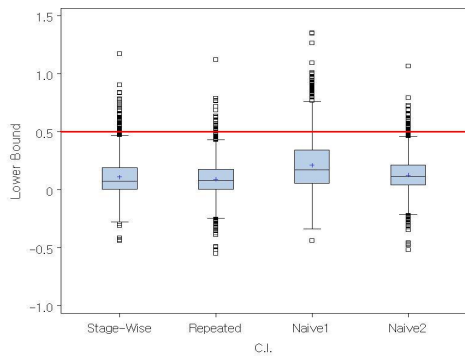
(b) $\sigma^2 = 3, \delta = 1.5$, H-DC(-4) type



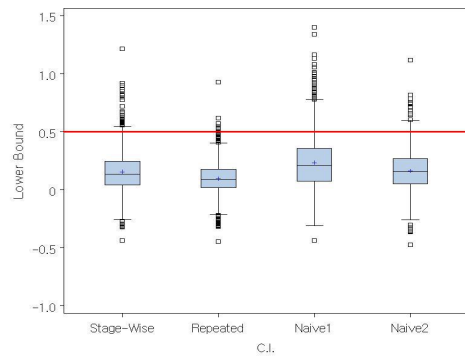
(c) $\sigma^2 = 3, \delta = 0.75$, PK type



(d) $\sigma^2 = 3, \delta = 0.75$, H-DC(-4) type



(e) $\sigma^2 = 3, \delta = 0.5$, PK type



(f) $\sigma^2 = 3, \delta = 0.5$, H-DC(-4) type

Figure 4.2. Box plot for lower limit of 95% 1-sided confidence interval where $D = (3, 1, 1, 4)$ and $\sigma^2 = 3$

Table 4.2에서는 첫번째 중간분석에서 적응법을 실시한 경우($L = 1$)의 모의실험으로부터 추정된 신뢰 구간하한에 대한 상자-수염 그림이다. 반복신뢰구간의 경우 신뢰구간하한이 상대적으로 작은 값을 가지는 것을 볼 수 있는데, 이는 반복신뢰구간의 포함확률이 과소추정되는 경향이 있는 것과 상호 연관성이 있는 것으로 보인다. naive1 신뢰구간의 신뢰구간하한은 산포도가 큰 경향이 있는 것이 확인되는데, 이

Table 4.2. Coverage probability for 95% 1-sided confidence interval where $D = (3, 1, 1, 4)$ and $\sigma^2 = 0.3$

δ	PK				H-DC(-4)			
	stage-wise	repeated	naive1	naive2	stage-wise	repeated	naive1	naive2
	C.I.	C.I.	C.I.	C.I.	C.I.	C.I.	C.I.	C.I.
1.5	0.950	0.998	0.948	1.000	0.950	0.998	0.948	1.000
1	0.950	0.998	0.833	1.000	0.950	0.998	0.948	1.000
0.75	0.941	0.995	0.757	0.998	0.950	0.994	0.769	1.000
0.5	0.961	0.980	0.772	0.998	0.966	0.996	0.774	1.000
0.25	0.963	0.982	0.780	0.999	0.965	0.995	0.776	1.000

Table 4.3. Coverage probability for 95% 1-sided confidence interval where $D = (3, 1, 1, 4)$ and $\sigma^2 = 30$

δ	PK				H-DC(-4)			
	stage-wise	repeated	naive1	naive2	stage-wise	repeated	naive1	naive2
	C.I.	C.I.	C.I.	C.I.	C.I.	C.I.	C.I.	C.I.
1.5	0.939	0.979	0.903	0.942	0.944	0.998	0.920	0.952
1	0.952	0.978	0.900	0.950	0.944	0.998	0.894	0.920
0.75	0.944	0.982	0.912	0.934	0.970	0.996	0.926	0.932
0.5	0.941	0.972	0.895	0.912	0.946	0.967	0.900	0.899
0.25	0.946	0.967	0.900	0.899	0.934	0.974	0.900	0.908

는 조건부 기각확률이 1단계 임상시험에서 축적된 자료에 따라 제 1종 오류인 0.05보다 크거나 작게 계산될 수 있는 것을 고려하면 쉽게 이해할 수 있다.

Table 4.2의 $d_{11}/\sigma^2 = 10$ 이고 $\sigma^2 = 0.3$ 인 경우의 결과로 개체 내 변동이 개체 간 변동에 비하여 큰 경우이다. 결과를 보면, 2단계 임상시험에서 사용된 오류분배율 함수가 Pocock 형식이거나 Hwang과 DeCani ($\gamma = -4$)의 경우 모두 단계적 순서를 고려한 신뢰구간의 포함확률은 모수의 값이 작아질수록 과대추정되는 경향이 있기는 하였으나 다른 추정방법과 비교하여 보면 매우 안정적인 포함확률을 보이고 있었다. 이는 검정력의 반복신뢰구간과 naive2 신뢰구간의 포함확률은 전체적으로 과대추정되었으며, naive1 신뢰구간의 포함확률은 과소추정 하는 경향을 보였다.

Table 4.3은 $d_{11}/\sigma^2 = 0.1$ 이고 $\sigma^2 = 30$ 인 경우의 결과로 개체내 변동이 개체 간 변동에 비하여 작으며 다른 모의실험 가정들과 비교하면 개체간 변동이 큰 경우이다. 결과를 보면, 단계적 순서를 고려한 신뢰구간의 포함확률은 Table 4.1와 Table 4.2의 단계적 순서를 고려한 신뢰구간과 비교할때는 다소 과소추정되는 경향을 보이나, Table 4.3의 다른 신뢰구간과 비교하여 보면 매우 안정적인 포함확률을 보이고 있었다.

Table 4.1, Table 4.2, Table 4.3의 결과를 종합하면, 단계적 순서를 고려한 신뢰구간의 경우 매우 안정적으로 신뢰계수인 0.95에 비교적 근사하는 결과를 보였다. 반면 Mehta 등 (2007)에서 제안된 반복신뢰구간은 과대추정하는 결과를 보였고, 이는 Mehta 등 (2007)의 결과와 일치한다. 또한, 개체 간 변동이 클수록 포함확률은 원하는 수준에서 멀어지는 결과를 보였는데, 이는 Lee 등 (2002)의 반복추정치를 가지는 집단축차설계 임상시험에서 추정된 신뢰구간에 대한 모의실험 결과에서도 보인바 있다. 반복추정치를 가지는 임상시험의 경우에는 단일관측치의 결과와 비교하여 과대추정하는 경향을 보였고, 결과적으로 자료구조와 설계방법이 복잡해질수록 신중한 추정법을 적용해야함을 알 수 있다. 또한, 집단축차검정을 고려하지 않고 적용할 수 있는 naive 신뢰구간은 경우에 따라 과소추정 또는 과대추정하여 일관성을 보이지 못하였다. 이는 예전 연구에서 논의되었던 적응법을 제외한 집단축차설계의 예에서도 나타난 바와 같이 적응적 집단축차설계의 임상시험에 적용하는 것은 추천되지 않는다.

일반적으로 임상시험이 종료된 후에 관심모수에 대한 신뢰구간을 계산해야 한다. 만약, 적응적 집단축차설계를 한 임상시험이 반복측정자료를 가지는 경우라면, 본 논문에서 제안한 단계적 순서관계를 고려한 신뢰구간을 사용할 때 가장 안정적인 추정결과를 얻을 수 있음을 모의실험을 통하여 확인하였다.

5. 결론

본 논문에서는 적응적 집단축차설계의 임상시험에서 자료가 반복측정자료를 가지는 경우의 신뢰구간 추정법에 대해서 논의하였다. Lee 등 (2002)은 집단축차설계법에서 선형혼합효과모형, 일반화추정방정식 등의 반복측정자료를 가지는 임상시험에 대하여 단계적 순서관계를 고려하여 신뢰구간을 추정하는 방법을 제안하였다.

본 연구에서는 Lee 등 (2002)을 적응적 집단축차설계의 경우로 확장하여 반복측정자료를 가지는 임상시험에서 적응법 실시 이후의 2단계 임상시험을 새롭게 설계할 수 있는 검정통계량과 임상시험이 종료된 후 관심모수에 대한 신뢰구간을 추정하는 방법을 제안하였다. 먼저, 자료가 반복측정치를 가지는 경우, 중간분석 시점까지 누적된 자료는 개체들 중 임상시험을 종료하지 못한 경우 임상시험에서 필요로 하는 측정 횟수의 일부가 될 수 있다. 그러므로, 적응법을 적용하여 2단계 임상시험을 고려하는 경우 1단계와 2단계 임상시험의 자료가 확률적으로 독립성을 유지할 수 없다. 따라서 본 논문에서는 독립증분을 하는 확률변수를 고려하여 2단계 임상시험을 1단계 임상시험과는 독립적으로 재설계 할 수 있게 하였고, 임상시험이 종료된 후 단계적 순서관계를 가지는 신뢰구간의 추정방법을 제안하였다. 또한, 신뢰구간의 하한 추정에는 Brannath 등 (2009)이 제안한 알고리즘을 적용하였다. 모의실험에서 Mehta 등 (2007)이 제안한 반복신뢰구간과 naive 신뢰구간의 포함확률과 본 연구에서 제안한 단계적 순서관계를 이용한 신뢰구간의 포함확률을 비교하였다. 반복신뢰구간의 포함확률은 알려진대로 과대추정하는 경향을 보였고, 이와 비교하여 단계적 순서관계를 이용한 신뢰구간은 안정적인 포함확률을 보였다. naive 방법을 적응적 집단축차설계의 임상시험에 적용하는 것은 추천되지 않는다. 모의실험은 선형혼합효과모형을 적용하였으나, 반복측정치를 가지는 모형에서 검정통계량으로부터 확률적으로 독립증분을 가지는 확률변수를 정의할 수 있는 모든 모형에서 적용이 가능하다.

본 논문은 1회의 적응법을 고려하여 적응적 설계를 기준으로 2단계 형태를 가정하였으나, 이는 3단계 이상으로 확장해볼 수 있을 것이다. 또한, 본 논문에서는 2집단을 비교하는 확증적 3상 임상시험에 초점이 맞추어져 있으나, 3집단 이상을 비교하여 치료효과가 적은 집단을 순차적으로 제외하는 형태의 임상시험에서 반복측정치를 적절한 형태로 적용하는 검정방법을 고찰하고 검정결과에 상관없이 임상시험 종료 후의 단계적 신뢰구간을 적용할 수 있는 등의 후속연구를 기대할 수 있을 것이다.

References

- Bauer, P. and Kohne, K. (1994). Evaluation of experiments with adaptive interim analyses, *Biometrics*, **50**, 1029–1041. Correction *Biometrics*, **52**, (1996), 380.
- Brannath, W., Mehta, C. R. and Posch, M. (2009). Exact confidence bounds following adaptive group sequential tests, *Biometrics*, **65**, 539–546.
- Cheng, Y. and Shen, Y. (2004). Estimation of a parameter and its exact confidence interval following sequential sample size reestimation trials, *Biometrics*, **60**, 910–918.
- Cui, L., Hung, H. M. J. and Wang, S. (1999). Modification of sample size in group sequential clinical trials, *Biometrics*, **55**, 853–857.
- Denne, J. S. (2001). Sample size recalculation using conditional power, *Statistics in Medicine*, **20**, 2645–2660.
- Fisher, L. D. (1998). Self-designing clinical trials, *Statistics in Medicine*, **17**, 1551–1562.

- Hwang, I. K., Shih, W. J. and DeCani, J. S. (1990). Group sequential design using a family of type I error probability spending functions, *Statistics in Medicine*, **9**, 1439–1445.
- Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*, Boca Raton, Chapman and Hall/CRC, Florida.
- Kim, K. and DeMets, D. L. (1987). Confidence intervals following group sequential tests in clinical trials, *Biometrics*, **43**, 857–864.
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics*, **38**, 963–974.
- 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.
- Lee, J. W., Jo, S. J., DeMets, D. L. and Kim, K. (2002). Confidence intervals following group sequential tests in clinical trials with multivariate observations, *Journal of Statistical Computation and Simulations*, **72**, 247–259.
- Lehmacher, W. and Wassmer, G. (1999). Adaptive sample size calculations in group sequential trials, *Biometrics*, **55**, 1286–1290.
- Lindstorm, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data, *Journal of the American Statistical Association*, **83**, 1014–1022.
- Mehta, R. C., Bauer, P., Posch, M. and Brannath, W. (2007). Repeated confidence intervals for adaptive group sequential trials, *Statistics in Medicine*, **26**, 5422–5433
- Muller, H. H. and Schafer, H. (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches, *Biometrics*, **57**, 886–891.
- Muller, H. H. and Schafer, H. (2004). A general statistical principle for changing a design any time during the course of a trials, *Statistics in Medicine*, **23**, 2497–2508.
- Posch, M., Bauer, P. and Brannath, W. (2003). Issues in designing flexible trials, *Statistics in Medicine*, **23**, 953–969.
- Proschan, M. A. and Hunsberger, S. A. (1995). Designed extension of studies based on conditional power, *Biometrics*, **51**, 1315–1324.
- Schervish, M. J. (1984). Multivariate normal probabilities with error bound (with corrections in 1985), *Applied Statistics*, **33**, 81–94.
- Shen, Y. and Fisher, L. D. (1999). Statistical inference for self-designing clinical trials with a one sided hypothesis, *Biometrics*, **55**, 190–197.
- Tsiatis, A. A., Rosner, G. L. and Mehta, C. R. (1984). Exact confidence intervals following a group sequential test, *Biometrics*, **40**, 797–803.

반복측정자료를 가지는 적응적 집단축차검정에서의 신뢰구간 추정

조숙정^a · 이재원^{b,1}

^a아이콘클리니컬 리서치 임상통계팀, ^b고려대학교 통계학과

(2013년 2월 19일 접수, 2013년 7월 2일 수정, 2013년 8월 5일 채택)

요약

집단축차설계법은 중간분석을 실시하여 임상시험용 의약품의 유효성 또는 무용성이 조기에 발견되면 임상시험을 조기에 종료할 수 있는 시험설계법이다. 적응적 설계법은 중간분석 결과를 이용하여 시험설계를 변경하거나, 확률적으로 독립인 두개의 임상시험 결과를 결합하는 등 다양한 적응법으로 임상시험의 설계를 수정할 수 있는 시험설계법이다. 집단축차설계법과 적응적 설계법에서 주요하게 고려할 점은, 시험 전체적으로 제1종 오류를 적절히 분배하고 통제하여 임상시험 전체의 일관성을 유지하도록 하는 것이다. 반복측정자료 또는 경시적자료의 통계적 모형이 고려되는 경우에는 통계적 추론이 더욱 복잡하고 어려워진다. Lee 등 (2002)에서는 반복측정치를 가지는 임상시험에서 집단축차설계에서 미리 정한 기준에 의하여 임상시험이 종료된 후, 독립증분과 단계적 순서관계를 고려한 신뢰구간 추정법을 제안한 바 있다. 본 연구는 Lee 등 (2002)를 적응적 설계로 확장하였다. 적응법을 실시한 전과 후의 임상시험을 확률적으로 독립인 관계로 정의하는 검정통계량을 유도하여 적응적 집단축차검정법이 가능하게 하였다. 또한, 임상시험이 종료된 후 단계적 순서관계를 고려한 신뢰구간 추정법을 제안하였으며, 모의실험을 통하여 그 안정성을 확인하였다.

주요용어: 신뢰구간추정, 적응적 설계법, 적응적 집단축차설계, 단계적 순서관계.

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2012R1A1A2008686).

¹교신저자: (136-710) 서울시 성북구 안암로 145, 고려대학교 통계학과, 교수. E-mail: jael@korea.ac.kr