

이변량 효능과 안전성 이항변수의 표본수 결정방법

이현학¹ · 송혜향²

¹가톨릭대학교 의학통계학과, ²가톨릭대학교 의학통계학과

(2008년 12월 접수, 2009년 1월 채택)

요약

두 군의 처리를 비교하는 임상시험에서 효능(efficacy)과 안전성(safety)이 동일하게 중요한 변수로 취급되는 경우에 이변량(bivariate) 반응변수로서 분석되고 연구계획의 단계에서도 이변량 표본수 결정방법이 사용되어야 한다. Thall과 Cheng (1999)은 효능과 안전성의 반응값이 이변량 이항(bivariate binary) 변수인 경우의 표본수 결정방법을 제시하였으며, 본 연구에서는 목표모수 설정과정은 기존의 연구와 같으나 윌콕슨-만-휘트니(Wilcoxon-Mann-Whitney: WMW) 통계량에 근거한 검정법과 표본수 결정방법을 제시한다. Thall과 Cheng (1999)의 검정통계량은 변수 변환시킨 비율의 근사 정규성에 근거하는 반면에, WMW 통계량은 확률에 근거한 비모수적 방법으로 이변량 이항변수 뿐만 아니라 이변량 순위변수로 측정된 반응값에도 적용시킬 수 있다. Thall과 Cheng (1999)에 제시한 항암치료 임상연구의 두 예제에 위의 두 다른 방법으로 계산된 표본수를 비교한 결과, Thall과 Cheng (1999)의 첫째 예제에서는 이변량 WMW 방법에 의한 표본수가 더욱 작았으나 둘째 예제에서는 더욱 큰 것으로 나타났다.

주요용어: 표본수 계산, 이변량 WMW, 효능과 안전성.

1. 서론

두 치료군의 처리 비교는 임상시험에서 가장 많이 쓰이는 계획으로 신약을 기존의 약제와 비교하여 효능(efficacy)의 차이를 비교하는 것이 한 가지 예이며, 이러한 효능의 비교에 대한 임상시험 연구가 현재 끊임없이 진행되고 있다. 그러나 독성이 강한 화학 치료제의 임상시험에서는 효능과 독성(toxicity)을 동일하게 중요한 변수로 취급하여 이변량(bivariate) 반응변수로 분석해야 한다는 의견이 최근 제기되고 있다 (Jennison과 Turnbull, 1993; Cook, 1996; Cook과 Farewell, 1994; Bryant와 Day, 1995; Conaway와 Petroni, 1995, 1996). 효능과 독성은 서로 긴밀한 관계에 있으면서 약제에 대한 다른 양상의 반응을 나타낸다. 예를 들어 연부조직육종(soft tissue sarcoma)에 대한 화학 치료제의 용량을 증가시키면 종양이 수축되지만 한편으로는 신장 및 신경 계통에 역영향을 미치는 위험도 높아진다. 또한 급성백혈병의 치료에서 중요한 목표는 완전관해(complete remission)를 얻는 것인데, 골수이식 치료 시에 흔히 이식편대 숙주반응(graft-versus-host-disease)의 위험이 따르므로 효능과 독성을 동시에 감안하여 분석해야 한다.

효능과 독성의 이변량적 관점에서 임상시험의 목적은 어떤 구체적인 효능의 증가를 얻기 위해 ‘어느 정도의 부작용을 감수해야 하는가’ 또는 독성을 줄이기 위해서는 ‘어느 정도까지 효능을 떨어뜨릴 수 있는가’라는 trade-off의 문제로 표현될 수가 있다 (Conaway와 Petri, 1996; Thall과 Cheng, 1999). 따라서 효능과 독성의 이변량 검정에서 대립가설은 여러 trade-off 목표모수(target parameter)의 최소 블록집

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

E-mail: hhsong@catholic.ac.kr

합(convex hull)으로 표현된다. 임상시험 설계에서는 이와 같은 효능과 독성을 이변량 반응변수로 동시에 검정하는 것이, 일차적으로 효능에 대해 검정한 후 다음으로 독성에 대해 검정하는 것보다 바람직하는데 그 이유는 Pocock 등 (1987)이 언급하였듯이 효능에 대해 우선 검정함으로써 인위적으로 유의한 결론을 얻을 수 있고, 자료의 부분적인 결론을 제시하게 되는 단점 때문이다. 이변량 반응변수에 대한 검정 및 표본수 결정방법이 아직 많이 알려져 있지 않다. 그러므로 이변량의 문제를 단순화시켜 효능과 독성의 이변량 관측값을 단일변수로 변환하여 검정한 경우도 있으나 이는 효능과 독성 각각에 대한 특성을 무시하는 단점이 있다 (Tubert-Bitter 등, 1995; Thall과 Russell, 1998).

본 논문에서는 환자의 효능과 안전성(독성의 반대 개념)의 반응이 이항자료인 비교 임상시험에서 윌콕슨-만-휘트니(Wilcoxon-Mann-Whitney: WMW) 검정통계량에 근거한 표본수 결정방법을 제안한다. 제안하는 비모수적 WMW 검정통계량은 $\Pr(Y_C \leq Y_T)$ 를 각 반응변수의 처리효과 모수로 사용하며, 다변량 대표본 U -통계량(multivariate multi-sample U -statistics)의 공분산을 이용하여 검정하고 표본수를 결정하는 방법이다 (Lehmann, 1975, pp. 362-371). 위에서 $\Pr(Y_C \leq Y_T)$ 는 Y_C 와 Y_T 가 서로 독립인 두 반응변수 모집단으로부터 각각 추출된 랜덤변수일 때 Y_T 가 Y_C 보다 클 확률이며, WMW 통계량과 직접적으로 관련되어 있음을 Birnbaum과 Klose (1957)이 확률의 추정량으로서 $U/(n_C n_T)$ (여기서 n_C 와 n_T 는 독립된 두 군의 표본수이다)를 제시한 것으로 알 수 있고, Noether (1987)는 이 확률을 사용하여 단변량 순위변수 표본수 결정 공식을 제시하였다. 의학에서 효능 또는 안전성에 대한 환자반응이 순위변수로 기록된 경우를 쉽게 찾아볼 수 있고 (Levine 등, 1988; Norwood와 Sampson, 1988; Petri 등, 1990; Thall과 Russell, 1998), 이변량 순위변수의 예로는 Molenberghs와 Lesaffre (1994)가 제시한 정신과 연구에서 수집된 자료를 들 수 있다. 뇌의 세로토닌의 분비가 제대로 되지 않는 환자를 플루보사민(flvoxamine) 약제로 치료한 결과는 정신과 증상이 없음을 표현한 1 단계부터 증상이 심한 7 단계까지의 순위변수이며, 부작용은 4 단계로서 부작용 없음(0), 일상생활을 방해하지 않는 정도(1), 일상생활을 방해하는 정도(2), 치료효과를 능가하는 정도의 심한 부작용(3)으로 측정된다. 그러나 순위변수로 측정된 자료의 분석 시에는 이항변수화(dichotomization)하여 편재하는 이항변수 검정통계량으로 결론내리고 있음이 현실이다. 본 논문에서는 이변량 순위변수의 표본수 결정방법인 WMW 방법을 Thall과 Cheng (1999)가 제안한 이변량 이항변수의 예제에 적용하여 표본수를 비교한다. WMW 방법은 이변량 순위변수에 적용 가능한 방법이므로 Thall과 Cheng (1999)이 제안한 방법으로 구한 표본수와 WMW 통계량을 이변량 이항변수에 적용하여 구한 표본수에서 차이를 보일 것이다. 본 논문에서는 이변량 이항변수의 표본수 계산결과만을 제시하며 이변량 순위변수의 경우에 대한 표본수 계산결과는 차후 논문에 제시될 것이다. 본 논문의 WMW 검정통계량과 표본수 결정방법은 제 II상 또는 제 III상 시험에 적용될 수 있다.

2. WMW 통계량에 근거한 표본수 결정

본 논문에서 제안하는 WMW 통계량에 근거한 표본수 결정은 목표모수 설정과정과 표본수 계산과정이 Thall과 Cheng (1999)에 제시된 것과 같다. 논문의 내용상 요구되는 Thall과 Cheng (1999)의 가설 설정방법은 2.3절에, 표본수 계산과정은 2.4절에 간략히 제시할 것이며 저자들이 제안하는 WMW 통계량은 자세히 설명한다.

이제 WMW 통계량과 통계량의 분포를 순위변수에 대해 설명하며, 이항변수의 경우에는 두 범주만을 고려하면 된다. C 와 T 가 서로 독립인 기존(control) 치료군과 실험(test) 치료군을 각각 나타낼 때, Y_k^E 과 Y_k^S ($k = C, T$)는 효능(efficacy: E)과 안전성(safety: S)을 나타내는 순위 반응변수로서 각각 1부터 R 과 1부터 C 까지의 값을 가지므로 각 군의 효능과 안전성의 순위 자료는 $R \times C$ 분할표로 표현된다. 단순히 세 범주인 경우($R = C = 3$)가 표 2.1에 제시되었다.

표 2.1. The 3 × 3 contingency table of efficacy and safety probabilities of control and treatment group

(a) Control group					(b) Treatment group				
Safety \ Efficacy	1	2	3	Total	Safety \ Efficacy	1	2	3	Total
1	$p_{C,11}$	$p_{C,12}$	$p_{C,13}$	$p_{C,1\cdot}$	1	$p_{T,11}$	$p_{T,12}$	$p_{T,13}$	$p_{T,1\cdot}$
2	$p_{C,21}$	$p_{C,22}$	$p_{C,23}$	$p_{C,2\cdot}$	2	$p_{T,21}$	$p_{T,22}$	$p_{T,23}$	$p_{T,2\cdot}$
3	$p_{C,31}$	$p_{C,32}$	$p_{C,33}$	$p_{C,3\cdot}$	3	$p_{T,31}$	$p_{T,32}$	$p_{T,33}$	$p_{T,3\cdot}$
Total	$p_{C,\cdot 1}$	$p_{C,\cdot 2}$	$p_{C,\cdot 3}$	1	Total	$p_{T,\cdot 1}$	$p_{T,\cdot 2}$	$p_{T,\cdot 3}$	1

우선 단변량인 경우, 즉 효능에 대한 치료효과의 비교에서 Y_C 와 Y_T 가 각각 기존 치료와 실험 치료의 순위 반응변수일 때, Thall과 Lachin (1988)이 두 독립된 군의 Y_C 분포와 Y_T 분포가 겹치는 정도의 척도라고 언급한 $\Pr(Y_C \leq Y_T) - \Pr(Y_T \leq Y_C)$, 또는 Simonoff 등 (1986)이 제안한 $\Delta = \Pr(Y_C \leq Y_T)$ 를 사용한다. 확률 $\Pr(Y_C \leq Y_T)$ 는 의료진에게 쉽게 이해되는 척도로 구체적으로 기존 치료에 비해 실험 치료가 더 좋게 반응할 확률이며, 치료효과가 없음은 $\Pr(Y_C \leq Y_T) = 0.5$ 로 표현된다. 표 2.1에서 기존 치료군의 총 환자수를 n_C , 실험 치료군의 총 환자수를 n_T 라 하면, 효능의 WMW 통계량은 다음과 같다.

$$\hat{\Delta}^E = \widehat{\Pr}(Y_C^E \leq Y_T^E) = \frac{1}{n_C n_T} \sum_{k=1}^{n_C} \sum_{l=1}^{n_T} \Psi(Y_{Ck}^E, Y_{Tl}^E) \tag{2.1}$$

$$\text{단, } \Psi(Y_{Ck}^E, Y_{Tl}^E) = \begin{cases} 1, & \text{if } Y_{Ck}^E \leq Y_{Tl}^E, \\ 0.5, & \text{if } Y_{Ck}^E = Y_{Tl}^E, \\ 0, & \text{if } Y_{Ck}^E \geq Y_{Tl}^E. \end{cases}$$

이러한 단변량 WMW 통계량의 분산 추정은 Hochberg (1981) 및 여러 논문에 언급되었다.

2.1. 이변량 WMW 통계량과 결합분포

효능과 안전성의 이변량 치료효과 벡터 $\Delta = (\Delta^E, \Delta^S)$ 의 각 원소인 $\Delta^E = \Pr(Y_C^E \leq Y_T^E)$ 와 $\Delta^S = \Pr(Y_C^S \leq Y_T^S)$ 는 $R \times C$ 분할표의 주변 확률을 이용하여 정의되며, 각 주변확률의 추정량(frequency estimators)을 대입하여 WMW 통계량 $\hat{\Delta} = (\hat{\Delta}^E, \hat{\Delta}^S)$ 을 얻는다.

$$\begin{aligned} \Delta^E &= \sum_{i=1}^{R-1} p_{C,i} \cdot \sum_{l=i+1}^R p_{T,l} + \frac{1}{2} \sum_{i=1}^R p_{C,i} p_{T,i} \\ &= \sum_{i=1}^R p_{C,i} \cdot \left(\sum_{l=i}^R p_{T,l} - \frac{p_{T,i}}{2} \right), \\ \Delta^S &= \sum_{j=1}^{C-1} p_{C,j} \cdot \sum_{l=j+1}^C p_{T,l} + \frac{1}{2} \sum_{j=1}^C p_{C,j} p_{T,j} \\ &= \sum_{j=1}^C p_{C,j} \cdot \left(\sum_{l=j}^C p_{T,l} - \frac{p_{T,j}}{2} \right). \end{aligned}$$

한편, 이변량 WMW 통계량의 공분산은 셀 확률, 주변확률 그리고 표본수 n_C 와 n_T 를 이용하여 추정한다 (DeLong 등, 1988). 구체적으로 이변량 WMW 벡터 $\hat{\Delta} = (\hat{\Delta}^E, \hat{\Delta}^S)$ 에 대한 공분산 행렬은

$\lambda = n_C/n_T$ 일 때 다음과 같다.

$$\begin{aligned}\Sigma &= \frac{1}{n_C} \Sigma_{10} + \frac{1}{n_T} \Sigma_{01} \\ &= \frac{1}{n_C} \begin{pmatrix} \sigma_{10}^{EE} & \sigma_{10}^{ES} \\ \sigma_{10}^{ES} & \sigma_{10}^{SS} \end{pmatrix} + \frac{1}{n_T} \begin{pmatrix} \sigma_{01}^{EE} & \sigma_{01}^{ES} \\ \sigma_{01}^{ES} & \sigma_{01}^{SS} \end{pmatrix} \\ &= \frac{1}{n_C} \left[\begin{pmatrix} \sigma_{10}^{EE} & \sigma_{10}^{ES} \\ \sigma_{10}^{ES} & \sigma_{10}^{SS} \end{pmatrix} + \lambda \begin{pmatrix} \sigma_{01}^{EE} & \sigma_{01}^{ES} \\ \sigma_{01}^{ES} & \sigma_{01}^{SS} \end{pmatrix} \right].\end{aligned}\quad (2.2)$$

$$\begin{aligned}\sigma_{10}^{EE} &= \sum_{i=1}^R p_{C,i} \left(\sum_{l=i}^R p_{T,l} - p_{T,i}/2 - \Delta^E \right)^2, \\ \sigma_{01}^{EE} &= \sum_{i=1}^R p_{T,i} \left(\sum_{l=1}^i p_{C,l} - p_{C,i}/2 - \Delta^E \right)^2, \\ \sigma_{10}^{SS} &= \sum_{j=1}^C p_{C,j} \left(\sum_{l=j}^C p_{T,l} - p_{T,j}/2 - \Delta^S \right)^2, \\ \sigma_{01}^{SS} &= \sum_{j=1}^C p_{T,j} \left(\sum_{l=1}^j p_{C,l} - p_{C,j}/2 - \Delta^S \right)^2, \\ \sigma_{10}^{ES} &= \sum_{i=1}^R \sum_{j=1}^C p_{C,ij} \left(\sum_{l=i}^R p_{T,l} - p_{T,i}/2 - \Delta^E \right) \left(\sum_{l=j}^C p_{T,l} - p_{T,j}/2 - \Delta^S \right), \\ \sigma_{01}^{ES} &= \sum_{i=1}^R \sum_{j=1}^C p_{T,ij} \left(\sum_{l=1}^i p_{C,l} - p_{C,i}/2 - \Delta^E \right) \left(\sum_{l=1}^j p_{C,l} - p_{C,j}/2 - \Delta^S \right).\end{aligned}$$

검정은 다변량 WMW 통계량의 불편일치 추정량은 점근적으로(asymptotically) 다변량 정규분포함을 이용하여 검정한다 (Lehmann, 1975, pp. 362-371). 그러므로 Δ 의 일치 추정량이 $\hat{\Delta}$ 이고 공분산 행렬을 Σ_{Δ} 라 할 때 두 군이 동일 표본수 ($n_T = n_C = n$)인 경우 n 이 무한대로 간다면 중심극한정리(Central Limit Theorem)에 의해서,

$$\sqrt{n}(\hat{\Delta} - \Delta) \approx \text{BVN}(\mathbf{0}, \Sigma_{\Delta}), \quad \text{여기서 } \Sigma_{\Delta} = \left[\begin{pmatrix} \sigma_{10}^{EE} & \sigma_{10}^{ES} \\ \sigma_{10}^{ES} & \sigma_{10}^{SS} \end{pmatrix} + \begin{pmatrix} \sigma_{01}^{EE} & \sigma_{01}^{ES} \\ \sigma_{01}^{ES} & \sigma_{01}^{SS} \end{pmatrix} \right] \quad (2.3)$$

을 따른다. 표본수가 다른 경우 n_C 와 n_T 가 무한대로 가고 $n_C/n_T \rightarrow \lambda$ 일 때 식 (2.3) 대신에 다음의 식 (2.4)를 이용한다.

$$\sqrt{n_C}(\hat{\Delta} - \Delta) \approx \text{BVN}(\mathbf{0}, \Sigma_{\Delta}), \quad \text{여기서 } \Sigma_{\Delta} = \left[\begin{pmatrix} \sigma_{10}^{EE} & \sigma_{10}^{ES} \\ \sigma_{10}^{ES} & \sigma_{10}^{SS} \end{pmatrix} + \lambda \begin{pmatrix} \sigma_{01}^{EE} & \sigma_{01}^{ES} \\ \sigma_{01}^{ES} & \sigma_{01}^{SS} \end{pmatrix} \right]. \quad (2.4)$$

식 (2.3)과 (2.4)에서 Σ_{Δ} 은 일치추정량인 $\hat{\Sigma}_{\Delta}$ 로 대체하며 식 (2.4)에서 효능과 안전성 사이의 상관계수는 다음과 같다.

$$\rho_{\Delta} = \frac{\sigma_{10}^{ES} + \lambda \sigma_{01}^{ES}}{\sqrt{\sigma_{10}^{EE} + \lambda \sigma_{01}^{EE}} \sqrt{\sigma_{10}^{SS} + \lambda \sigma_{01}^{SS}}}. \quad (2.5)$$

2.2. Global 오즈비로 표현한 셀 확률

공분산 행렬 (2.2)의 추정량을 구하기 위해서, 더욱 구체적으로 σ_{10}^{ES} 와 σ_{01}^{ES} 를 구하기 위해서는 T 와 C 의 순위 반응변수의 결합분포, 즉 결합확률을 알아야 한다. 이 결합확률은 표 2.1의 T 와 C 의 효능과 안전성의 주변 확률을 알고, 효능과 안전성 사이의 연관성에 대한 척도를 가정할 때 구해진다. Agresti (1984, pp. 15-22)는 $R \times C$ 분할표 순위 반응변수의 연관성 척도로서 세 가지 오즈비(odds ratio), 즉 local, local-global, global 오즈비를 제시하고 있으며, 주변확률과 오즈비 ψ_{ij} 를 이용하여 결합확률을 구할 수 있음을 제시하였다. 본 논문에서는 여러 오즈비 중에서 global 오즈비를 이용하여 결합확률을

구한다. 즉, 표 2.1의 3×3 분할표의 순위 반응변수에서 기존 치료군의 경우 주변확률로 표현한 global 오즈비 ψ_{ij} 는 다음과 같다.

$$\psi_{C,ij} = \frac{\sum_{a \leq i} \sum_{b \leq j} p_{C,ab} \sum_{a > i} \sum_{b > j} p_{C,ab}}{\sum_{a \leq i} \sum_{b > j} p_{C,ab} \sum_{a > i} \sum_{b \leq j} p_{C,ab}}, \quad \text{단, } i, j = 1, 2, 3; a, b = 1, 2, 3.$$

이를 이용하여 역으로 결합확률을 구하면 다음과 같다 (Lipsitz 등, 1990).

$$\begin{aligned} p_{C,11} &= \begin{cases} \frac{\omega_{11} - \sqrt{\omega_{11}^2 - 4\psi_{C,11}(\psi_{C,11} - 1)p_{C,1}p_{C,1}}}{2(\psi_{C,11} - 1)}, & \text{for } \psi_{C,11} \neq 1, \\ p_{C,1}p_{C,1}, & \text{for } \psi_{C,11} = 1, \end{cases} \\ &\quad \text{단, } \omega_{11} = (\psi_{C,11} - 1)(p_{C,1} + p_{C,1} + 1), \\ p_{C,13} &= \begin{cases} \frac{\omega_{12} - \sqrt{\omega_{12}^2 - 4\psi_{C,12}(\psi_{C,11} - 1)p_{C,1}p_{C,3}}}{2(\psi_{C,12} - 1)}, & \text{for } \psi_{C,12} \neq 1, \\ p_{C,1}p_{C,3}, & \text{for } \psi_{C,12} = 1, \end{cases} \\ &\quad \text{단, } \omega_{12} = (\psi_{C,12} - 1)(p_{C,1} + p_{C,3}) - \psi_{C,12}, \\ p_{C,31} &= \begin{cases} \frac{\omega_{21} - \sqrt{\omega_{21}^2 - 4\psi_{C,21}(\psi_{C,21} - 1)p_{C,3}p_{C,1}}}{2(\psi_{C,21} - 1)}, & \text{for } \psi_{C,21} \neq 1, \\ p_{C,3}p_{C,1}, & \text{for } \psi_{C,21} = 1, \end{cases} \\ &\quad \text{단, } \omega_{21} = (\psi_{C,21} - 1)(p_{C,3} + p_{C,1}) - \psi_{C,21}, \\ p_{C,33} &= \begin{cases} \frac{\omega_{22} - \sqrt{\omega_{22}^2 - 4\psi_{C,22}(\psi_{C,22} - 1)p_{C,3}p_{C,3}}}{2(\psi_{C,22} - 1)}, & \text{for } \psi_{C,22} \neq 1, \\ p_{C,3}p_{C,3}, & \text{for } \psi_{C,22} = 1, \end{cases} \\ &\quad \text{단, } \omega_{22} = (\psi_{C,22} - 1)(p_{C,3} + p_{C,3} + 1). \end{aligned} \tag{2.6}$$

나머지 결합확률은 $\sum_{j=1}^3 p_{C,ij} = p_{C,i}$, $\sum_{i=1}^3 p_{C,ij} = p_{C,j}$, $\sum_{i=1}^3 \sum_{j=1}^3 p_{C,ij} = 1$ 을 이용하여 구한다. 마찬가지로 실험 치료군의 순위변수 결합확률은 위의 식에서 C 를 T 로 바꾸어 구한다. 표본수 결정에서는 기존 치료군과 실험 치료군의 global 오즈비가 동일하다는 가정 하에서, 즉 공통 global 오즈비 $\psi_{ij} = \psi_{C,ij} = \psi_{T,ij}$ 를 이용하여 구한다.

2.3. 가설 및 검정통계량

항암 기존 치료에 대한 실험 치료의 효능과 안전성을 검증하고자 하는 임상시험에서 임상적 효능의 표본수 결정에서 요구되는 효능과 안전성의 trade-off 관계를 대립가설에 해당하는 목표모수로서 제시한다. 구체적으로 이항변수인 효능과 안전성의 차이가 없음에 해당하는 모수벡터 $(0, 0)$ 에 대해 trade-off 관계를 표현한 목표모수 벡터 $\xi = (\xi^E, \xi^S)$ 의 예는 다음과 같다. 한 원소만이 양의 실수를 가지게 되는, 즉 목표모수가 $\xi^E > 0 > \xi^S$ 또는 $\xi^E < 0 < \xi^S$ 인 경우는 안전성의 ξ^S 만큼 감소에 효능이 ξ^E 만큼 증가한 것과 효능의 ξ^E 만큼 감소에 안전성의 ξ^S 만큼 증가한 trade-off 관계를 표현한 것이다. 이보다 최적의 목표모수는 $\xi^E > 0 = \xi^S$ 인데 이는 안전성이 떨어지지 않으면서 효능이 증가하는 것이며, 이와 반대로 $\xi^E = 0 < \xi^S$ 이면 효능이 떨어지지 않으면서 안전성이 증가하는 것이다. 일반적으로 효능과 안전성이 둘 다 증가하는 경우인 $\xi^E > 0$ 과 $\xi^S > 0$ 의 동시 달성은 임상에서 불가능하기에 고려하지 않는다. WMW 통계량을 이용하는 경우에는 귀무가설을 표현한 모수벡터는 $(0, 0)$ 이 아니라 $(0.5, 0.5)$ 가 된다.

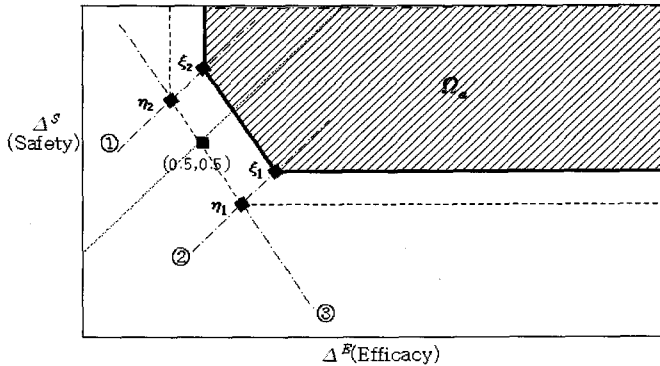


그림 2.1. Null point, alternative region and test boundary for the WMW statistic

임상의가 위에서와 같이 제시한 여러 개의 목표모수 ξ_1, \dots, ξ_K 로서 기각영역 $\Omega_a \subset R^2$ 가 정해진다. 다시 말하면 기각영역 Ω_a 는 적어도 $\xi_k (k = 1, \dots, K)$ 보다 더욱 바람직한 점의 집합인 $A(\xi_k) = \{\Delta : \Delta^E \geq \xi_k^E \text{ and } \Delta^S \geq \xi_k^S\}$ 를 이용하여 $\bigcup_{k=1}^K A(\xi_k)$ 의 최소 볼록집합(convex hull)으로 정의한다. $\Delta = (\Delta^E, \Delta^S)$ 가 바람직한 기각영역에 속한다면 최소 볼록집합은 $\gamma \Delta^E + (1-\gamma) \Delta^S (0 < \gamma < 1)$ 이고 귀무가설은 기각영역 Ω_a 밖에 속하게 된다. 그림 2.1에서 기각영역 Ω_a 은 굵은 실선(solid line)으로 제시된 다각형 영역이다.

WMW 검정통계량은 Δ 의 일치추정량 $\hat{\Delta}$ 의 분포, 즉 $\sqrt{n}(\hat{\Delta} - \Delta)$ 가 근사적으로 평균이 $(0, 0)$ 이고 공분산행렬이 Σ_Δ 인 이변량 정규분포함을 이용하며, 여기서 Σ_Δ 의 일치추정량인 $\hat{\Sigma}_\Delta$ 로 대체한다. 기각영역 Ω_a 의 결정은 Ω_a 를 45도 실선인 $\{\Delta : \Delta^E = \Delta^S\}$ 을 따라 귀무가설인 $(0.5, 0.5)$ 쪽으로 이동하면서 제 1종 오류(Type I error)를 만족할 때 멈추게 된다(그림 2.1). 즉, 임의의 c_α 가 $\Pr\{\hat{\Delta} \in R(c_\alpha) | \Delta = \mathbf{0}\} \leq \alpha$ 를 만족하는 기각역 $R(c_\alpha) = \Omega_a - (c_\alpha, c_\alpha)$ 지점을 찾는데 이 $R(c_\alpha)$ 는 Ω_a 에서 최종적으로 제 1종 오류 α 가 정해진 (c_α, c_α) 를 제외한 기각영역인 것이다. 이러한 구분에 크게 구애받지 않고 Ω_a 와 $R(c_\alpha)$ 모두를 기각영역이라 부른다. 효능과 안전성에 동일 중요성을 둔다면 기각영역을 45도 선을 따라 움직이며 만약 중요성을 달리 두고자 한다면 45도가 아닌 각도로 움직일 수 있다.

2.4. 표본수 결정

표본수를 결정하기 위해서는 검정력 함수 $\phi(\Delta) = \Pr\{\hat{\Delta} \in R(c_\alpha) | \Delta = (\Delta^E, \Delta^S)\}$ 가 요구된다. α 와 β 를 각각 제 1종과 제 2종 오류라 할 때 기각영역과 검정력은 다음에 의해 정해진다.

$$\Pr\{\hat{\Delta} \in R(c_\alpha) | \Delta = (0, 0)\} = \alpha,$$

$$\Pr\{\hat{\Delta} \in R(c_\alpha) | \xi_i = (\xi_i^E, \xi_i^S)\} = 1 - \beta, \quad (i = 1, \dots, K).$$

이변량 정규분포를 기각역에 대해 이중적분(double integration)으로 구체적인 기각영역 $R(c_\alpha)$ 와 표본 크기 $n (= n_C = n_T)$ 을 구한다. 이변량 정규분포에서 검정력 함수는 두 모수 Δ^E, Δ^S 각각에 대해 모두 증가하므로 여러 $\phi(\xi_1), \dots, \phi(\xi_K)$ 에 대해서 표본크기 n 은

$$\min_{1 \leq k \leq K} \phi(\xi_k) \geq 1 - \beta$$

을 만족하는 표본수로 결정한다. 이는 각 목표모수에서의 최소 검정력이 정해진 검정력보다 크게 된다 면 모든 목표모수에서의 검정력은 정해진 검정력보다 크기 때문이다.

표 3.1. Contingency tables of the number of patients and its probabilities($i = T, C$)

(a) The number of patients				(b) Probabilities			
Safety ($Y_{i,2}$)	Yes(1)	No(0)	Total	Safety ($Y_{i,2}$)	Yes(1)	No(0)	Total
Efficacy ($Y_{i,1}$)				Efficacy ($Y_{i,1}$)			
Yes(1)	$W_{i,11}$	$W_{i,10}$	$X_{i,1}$	Yes(1)	$\pi_{i,11}$	$\pi_{i,10}$	$\theta_{i,1}$
No(0)	$W_{i,01}$	$W_{i,00}$	$n_i - X_{i,1}$	No(0)	$\pi_{i,01}$	$\pi_{i,00}$	$1 - \theta_{i,1}$
Total	$X_{i,2}$	$n_i - X_{i,2}$	n_i	Total	$\theta_{i,2}$	$1 - \theta_{i,2}$	1

구체적인 계산과정은 다음과 같다. 그림 2.1에서 목표모수가 두 점 $\xi_1 = (\xi_1^E, \xi_1^S)$ 와 $\xi_2 = (\xi_2^E, \xi_2^S)$ 일 때 기각영역 Ω_a 를 45도 선을 따라 (0.5, 0.5)쪽으로 평행 이동한 지점을 $\eta_1 = (\eta_1^E, \eta_1^S)$ 와 $\eta_2 = (\eta_2^E, \eta_2^S)$ 라 한다. 기울기가 1인 직선 ①과 ②, 또한 두 점 ξ_1, ξ_2 을 지나는 직선과 평행인 직선 ③을 이용하여 $\eta_2 = (\eta_2^E, \eta_2^S) = (c, c + \xi_2^S - \xi_2^E)$ 로부터 제 1종 오류 α 를 만족토록 표본수 n 과 상수 c 를 구하면 된다. 본 논문에서는 프로그램 Mathematica의 FINDROOT 함수를 이용하여 이변량 정규분포의 이중적분을 시행해 가면서 n 과 c 를 뉴튼-랩슨(Newton-Raphson) 방법으로 구하였다. 이 과정에서 각 목표변수에 따라 공분산 행렬이 다르게 추정되므로 각 목표모수에 따라 변하는 상관관계를 사용하여 검정력을 계산한다.

3. Thall과 Cheng (1999)의 표본수 결정

이제 Thall과 Cheng (1999)이 제시한 이변량 이항변수의 경우의 검정통계량과 표본수 결정방법을 간략히 설명한다. Thall과 Cheng (1999)이 제시한 기호를 그대로 사용하여 앞 절에서 효능을 E , 안전성을 S 로 첨자에서 제시된 것과는 다르게 본 절에서는 효능은 1, 안전성은 2로 제시된다.

3.1. 확률분포

기존 치료약제군과 실험 치료약제군의 각 $n_i (i = T, C)$ 명 환자에게서 수집된 효능과 안전성의 이변량 이항자료는 표 3.1과 같이 요약된다. 각 셀의 환자수를 $W_{i,xy}$ 라 할 때, $\mathbf{W} = (W_{i,00}, W_{i,10}, W_{i,01}, W_{i,11})$ 은 확률이 $(\pi_{i,00}, \pi_{i,10}, \pi_{i,01}, \pi_{i,11})$ 인 다항분포에 따른다. 그러므로 효능과 안전성이 모두 1, 즉 Yes인 환자수 $X_{i,1} = W_{i,11} + W_{i,10}$ 와 $X_{i,2} = W_{i,11} + W_{i,01}$ 은 모수가 각각 $(n_i, \theta_{i,1})$ 과 $(n_i, \theta_{i,2})$ 인 이항분포하며, 이때 공분산은 $\text{Cov}(X_{i,1}, X_{i,2}) = n_i(\pi_{i,11} - \theta_{i,1}\theta_{i,2}) = n_i(\pi_{i,11}\pi_{i,00} - \pi_{i,10}\pi_{i,01})$ 이다.

기존 치료약제군 또는 실험 치료약제군의 효능을 $Y_{i,1}$, 안전성을 $Y_{i,2}$ 라 하자($i = T, C$). 효능과 안전성의 이변량 반응이 각각 x 와 y 일 때의 확률을 $\text{Pr}(Y_{i,1} = x \text{ and } Y_{i,2} = y) = \pi_{i,xy}$ 이며, 효능과 안전성의 주변확률은 각각 $\theta_{i,1} = \text{Pr}(Y_{i,1} = 1) = \pi_{i,11} + \pi_{i,10}$ 와 $\theta_{i,2} = \text{Pr}(Y_{i,2} = 1) = \pi_{i,11} + \pi_{i,01}$ 로 표현된다($i = T, C$). 또한 효능과 안전성의 연관성, 즉 $Y_{i,1}$ 과 $Y_{i,2}$ 의 연관성은 상관계수, 오즈비로 표현될 수 있으며 각각 다음과 같다.

$$\text{Corr}(Y_{i,1}, Y_{i,2}) = \frac{n_i(\pi_{i,11} - \theta_{i,1}\theta_{i,2})}{\sqrt{\theta_{i,1}(1 - \theta_{i,1})\theta_{i,2}(1 - \theta_{i,2})}}$$

$$\psi_i = \frac{\pi_{i,11}\pi_{i,00}}{\pi_{i,10}\pi_{i,01}} \tag{3.1}$$

기존 치료약제와 비교하여 실험 치료약제의 효능과 안전성에 차이가 있는지를 알아보는 임상시험에서 Thall과 Cheng (1999)은 효능 또는 안전성의 확률 $\theta_{i,j} = \text{Pr}(Y_{i,j} = 1)$ 의 함수인 $g(\theta_{i,j}) = \sin^{-1}(\theta_{i,j})^{1/2}$ 를 제한하여, 두 모집단의 비교는 처리효과와 차인 $\Delta^j = g(\theta_{T,j}) - g(\theta_{C,j}) (j = 1, 2)$ 로서 검정한다. 여기서 모수는 $\Delta = (\Delta^1, \Delta^2)$ 이며 Δ^1 과 Δ^2 는 각각 두 치료약제군의 효능과 안전성의 차를 표현한다. 두 치료약제군에 차이가 없는 경우에 $\Delta = (0, 0)$ 인 값을 가지게 되고, Δ^1 (또는 Δ^2)가 양의

값을 가지면 T 군이 C 군에 비해 더 높은 효능(또는 안전성)을 가지고 있음을 뜻하며, 음의 값을 가지면 C 군이 T 군에 비해 더 높은 효능(또는 안전성)을 가짐을 뜻한다.

3.2. 검정통계량

검정가설은 앞의 2.3절의 내용과 같다. Thall과 Cheng (1999)은 분산 안정화 변환 $g(\theta) = \sin^{-1}(\theta)^{1/2}$ 을 이용하여 효능과 안전성의 확률 $\theta = (\theta_1, \theta_2)$ 의 일치추정량이 $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ 일 때, 이변량 검정통계량인 $\hat{\Delta} = (\hat{\Delta}^1, \hat{\Delta}^2) = (g(\hat{\theta}_1) - g(\theta_1), g(\hat{\theta}_2) - g(\theta_2))$ 의 분포를 다음과 같이 제시하였다.

$$\sqrt{n} \begin{pmatrix} g(\hat{\theta}_1) - g(\theta_1) \\ g(\hat{\theta}_2) - g(\theta_2) \end{pmatrix} \sim AN(\mathbf{0}, \Sigma_g),$$

$$\text{단, } \Sigma_g = \begin{pmatrix} g'(\theta_1)^2 \theta_1(1-\theta_1) & g'(\theta_1)g'(\theta_2)(\pi_{11} - \theta_1\theta_2) \\ g'(\theta_1)g'(\theta_2)(\pi_{11} - \theta_1\theta_2) & g'(\theta_2)^2 \theta_2(1-\theta_2) \end{pmatrix} = \begin{pmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{12} & \sigma^{22} \end{pmatrix}$$

그러므로 각 군에 n 명의 표본수가 할당된 독립된 두 표본의 검정통계량은

$$\sqrt{2n}(\hat{\Delta} - \Delta) \sim AN(\mathbf{0}, \Sigma_\Delta), \quad \text{단, } \Sigma_\Delta = 2 \begin{pmatrix} \sigma_C^{11} + \sigma_T^{11} & \sigma_C^{12} + \sigma_T^{12} \\ \sigma_C^{12} + \sigma_T^{12} & \sigma_C^{22} + \sigma_T^{22} \end{pmatrix}$$

이다. 여기서 σ_i^{ab} 는 Σ_g 의 (a, b) 지점의 원소이다. $g(\theta) = \sin^{-1}(\theta)^{1/2}$ 의 경우 $g'(\theta) = 4\theta(1-\theta)^{-1/2}$ 이 되므로 검정통계량의 공분산은 매우 단순한 형태가 된다.

$$\sqrt{2n}(\hat{\Delta} - \Delta) \sim AN(\mathbf{0}, \Sigma_\Delta), \quad \text{단, } \Sigma_\Delta = \begin{pmatrix} 1 & \rho_\Delta \\ \rho_\Delta & 1 \end{pmatrix}. \quad (3.2)$$

여기서 ρ_Δ 로서 기존 치료군과 실험 치료군의 공통 상관계수 $\rho_\Delta = (\rho_C + \rho_T)/2$ 를 사용한다. 각 치료군의 $\rho_i (i = T, C)$ 는 앞의 2.2절에서와 마찬가지로 $\pi_{i,11}$ 의 확률을 구한 후에 주변확률 $\theta_{i,1}$, $\theta_{i,2}$ 와 오즈비 ψ_C 가 정해지면 식 (3.1)에 의해 ρ_i 를 구한다.

3.3. 표본수 결정

표본수를 결정하기 위해서는 식 (3.2)에 제시된 검정통계량을 이용하여 검정력 함수

$$\phi(\Delta) = \Pr \left\{ \hat{\Delta} \in R(c_\alpha) \mid \Delta = (\Delta^1, \Delta^2) \right\}$$

를 구하며 $\phi(\Delta)$ 는 Δ , n 그리고 상관계수 ρ_Δ 에 의해 결정된다. 따라서 $R(c_\alpha)$ 와 n 이 주어졌을 때 각 목표모수점의 검정력 $\phi(\xi_k) (k = 1, \dots, K)$ 를 계산하기 위해, 우선 각 ξ_k 점에 따라 ρ_Δ 의 값이 다르므로 각 목표모수점마다 개별적으로 상관계수 ρ_Δ 를 추정한다. 다시 말하면 공통 오즈비가 $\psi = \psi_C = \psi_T$ 일 때 다항분포에 의해 $\pi_{i,11}$ 을 구한 후에 각 ξ_k 에 대해 ρ_Δ 값을 추정한다. 임상시험에서 요구되는 표본크기 $2n$ 은 각 목표모수점의 검정력 $\phi(\xi_1), \dots, \phi(\xi_K)$ 이 $\min_{1 \leq k \leq K} \phi(\xi_k) \geq 1 - \beta$ 를 만족하는 표본수로 정한다.

4. 결과

이제 여러 실험 계획에 대해 WMW 검정통계량으로 계산한 표본수를 Thall과 Cheng (1999)의 표본수와 비교한다. Thall과 Cheng (1999)이 제시한 임상시험 첫 번째 예제의 계획 1은 전이된 연부조직육종에 대한 화학요법 이포스파마이드(ifosfamide) 용량 $10\text{g}/\text{m}^2$ 이 기존 치료이며 이 기존 치료약제와 비교하고자 하는 실험 치료약제는 화학요법 이포스파마이드(ifosfamide) 용량 $16\text{g}/\text{m}^2$ 이다. 기존 치

표 4.1. Sample sizes of Thall and Cheng (1999) and WMW methods for the ifosfamide trial with $\psi = 1.0$

Design	$\delta_k = (\delta_{k,1}, \delta_{k,2})$	ξ_k		Sample size	
		Thall and Cheng	Bivariate WMW	Thall and Cheng	Bivariate WMW
1	\mathbf{p}_1 (0.30, -0.10)	(0.322, -0.172)	(0.650, 0.450)	226	190
	\mathbf{p}_2 (0.20, -0.05)	(0.221, -0.096)	(0.600, 0.475)		
	\mathbf{p}_3 (0.15, 0.00)	(0.169, 0.000)	(0.575, 0.500)		
2	\mathbf{p}_1 (0.30, -0.15)	(0.322, -0.238)	(0.650, 0.425)	232	192
	\mathbf{p}_2 (0.20, -0.10)	(0.221, -0.172)	(0.600, 0.450)		
	\mathbf{p}_3 (0.15, -0.05)	(0.169, -0.096)	(0.575, 0.475)		
3	\mathbf{p}_1 (0.30, -0.10)	(0.322, -0.172)	(0.650, 0.450)	486	422
	\mathbf{p}_2 (0.20, -0.05)	(0.221, -0.096)	(0.600, 0.475)		
	\mathbf{p}_3 (0.10, 0.00)	(0.116, 0.000)	(0.550, 0.500)		

$\dagger(\theta_C = \theta_{C,1}, \theta_{C,2}) = (0.20, 0.95)$ and $\xi_{k,j} = \sin^{-1}(\theta_{C,j} + \delta_j)^{1/2} - \sin^{-1}(\theta_{C,j})^{1/2}$ for the k^{th} target, for $j = 1, 2$.

료의 효능과 안전성의 확률은 $\theta_C = (\theta_{C,1}, \theta_{C,2}) = (0.2, 0.95)$ 이며, 두 이항변수의 오즈비는 이 예제에서 $\psi_C = 1$ 이라 두었다. 임상이가 제시한 목표모수는 세 점으로서 각각 $\mathbf{p}_1 = (0.5, 0.85)$, $\mathbf{p}_2 = (0.4, 0.9)$, $\mathbf{p}_3 = (0.35, 0.95)$ 이며, 목표모수 세 점 \mathbf{p}_1 , \mathbf{p}_2 , \mathbf{p}_3 은 기존치료제에 비해 실험 치료약제에 의해 효능이 각각 0.3, 0.2, 0.15 만큼 증가하기 위해서 안전성이 각각 0.1, 0.05, 0 정도 낮아짐을 감수해야 하는 trade-off를 표현하고 있다. 표 4.1에서 $\delta_k = (\delta_{k,1}, \delta_{k,2})$ 는 이러한 목표모수와 기존 치료약제 사이의 확률의 차이이며, 결합확률은 오즈비 $\psi = \psi_C = \psi_T$ 에 의해 정해진다. 제 1종 오류를 $\alpha = 0.05$ 로, 검정력은 $1 - \beta = 0.8$ 로 정한 후 각 치료약제군에 동일 표본수가 할당되는 경우로 계산된 총 표본수는 표 4.1과 같다.

표 4.1의 계획 1의 경우 Thall과 Cheng (1999)의 표본수 결정과정은 다음과 같다. 이변량 정규분포의 이중적분법의 시행으로 뉴턴-랩슨 방법으로 구한 n 과 c 는 $c = 0.108298$ 인 경우 $n = 111$ 로서 두 군의 표본수는 222명이 되며, 이 때 각 목표지점에서의 검정력은 구체적으로 ξ_1 에서는 0.793, ξ_2 에서는 0.854, ξ_3 에서는 0.801이다. 그러나 ξ_1 에서 검정력이 0.80을 넘지 못하므로 계속 다른 수치 c 를 찾으며 $c = 0.107332$ 의 경우에 $n = 113$ 로서 두 군의 표본수는 226명이 되며 이 때 각 목표지점에서의 검정력은 ξ_1 에서는 0.801, ξ_2 에서는 0.861, ξ_3 에서는 0.808로서 모두 0.80이상이므로 표본수 226명으로 정한다.

표 4.1에 제시된 계획 1, 2, 3에 대한 표본수를 살펴보면 기존 치료와 실험 치료의 효능과 안전성의 차이에 대해 계획 2는 계획 1에 비해 각 점에서 목표로 하는 효능의 증가에 대한 확률은 서로 같으나 안전성의 확률은 0.05 정도 더욱 낮아짐을 견뎌야 하기 때문에 두 설계의 표본수는 큰 차이가 없다. 그러나 계획 3은 계획 1에 비해 목표로 하는 \mathbf{p}_3 의 효능에 대한 확률이 매우 작은 폭인 0.05가 더 낮아졌으나 계산된 표본수에서는 2배 이상의 차이가 난다. 그 이유는 이변량 모수공간에서 낮아지는 확률의 크기가 비록 같다 하더라도 효능에 대한 차이가 안전성의 차이에 대한 경우보다도 이항변수인 경우 귀무가설 (0, 0)의 방향으로 더욱 가까워지기 때문에 동일 검정력으로 더욱 작은 차이를 검정하기 위해서는 표본수가 커야 하는 것이다. 결론적으로, 표 4.1에 제시된 각 계획에서 Thall과 Cheng (1999)의 방법보다도 WMW 방법에 의한 표본수가 더욱 작게 계산되었다.

두 번째 예제 자료는 급성 골수성 백혈병(acute myelogenous leukemia: AML)의 화학치료법으로서 기존 치료는 사이토신 아라비노사이드(cytosine arabinoside: ara-C)이며 실험 치료약제는 겐시타빈(gemcitabin)과 사이클로포스파마이드(cyclophosphamide)의 조합이다. 기존 치료약제의 안전성과

표 4.2. Sample sizes of Thall and Cheng (1999) and WMW methods for the AML trial with $\psi = 3.05$

Design	$\delta_k = (\delta_{k,1}, \delta_{k,2})$		ξ_k		Sample size	
			Thall and Cheng	Bivariate WMW	Thall and Cheng	Bivariate WMW
1	\mathbf{p}_1	(0.20, -0.05)	(0.258, -0.051)	(0.600, 0.475)	334	462
	\mathbf{p}_2	(0.00, 0.25)	(0.000, -0.295)	(0.500, 0.625)		
2	\mathbf{p}_1	(0.20, -0.05)	(0.258, -0.051)	(0.600, 0.475)	436	576
	\mathbf{p}_2	(0.00, 0.20)	(0.000, 0.226)	(0.500, 0.600)		
3	\mathbf{p}_1	(0.20, -0.05)	(0.258, -0.051)	(0.600, 0.475)	744	890
	\mathbf{p}_2	(0.10, 0.00)	(0.116, 0.000)	(0.550, 0.500)		
	\mathbf{p}_3	(0.00, 0.25)	(0.000, 0.295)	(0.500, 0.625)		
4	\mathbf{p}_1	(0.20, 0.05)	(0.258, 0.052)	(0.600, 0.525)	240	282
	\mathbf{p}_2	(0.05, 0.20)	(0.056, 0.226)	(0.525, 0.600)		

$\dagger(\theta_C = \theta_{C,1}, \theta_{C,2}) = (0.70, 0.62)$ and $\xi_{k,j} = \sin^{-1}(\theta_{C,j} + \delta_j)^{1/2} - \sin^{-1}(\theta_{C,j})^{1/2}$ for the k^{th} target, for $j = 1, 2$.

표 4.3. Effect of association on sample sizes in Design 1 for the ifosfamide trial

$\pi_{C,11}$	0.2	0.195	0.19	0.18	0.175	0.17	0.16
ψ_C	∞	2.239	1	0.351	0.146	0.051	0
Thall and Cheng	230	228	226	220	216	208	202
Bivariate WMW	190	190	190	188	188	186	184

표 4.4. Effect of association on sample sizes in Design 1 for the AML trial

$\pi_{C,11}$	0.62	0.57	0.53	0.49	0.45	0.41	0.37	0.32
ψ_C	∞	21.90	7.27	3.05	1.38	0.606	0.224	0
Thall and Cheng	412	386	360	334	306	276	244	200
Bivariate WMW	568	532	498	462	424	384	342	280

효능의 확률은 $\theta_C = (\theta_{C,1}, \theta_{C,2}) = (0.7, 0.62)$ 이고, 과거 연구로부터 알게된 $\psi = 3.05$ 를 그대로 채택하여 표본수를 결정한다. 임상에 의해 정해지는 목표모수에 대한 효능과 안전성의 확률을 이용하여 계산한 표본수는 표 4.2와 같다.

표 4.2의 계획 1에서 목표모수 \mathbf{p}_1 은 효능이 0.2 정도 증가하게 되면 안전성은 0.05 정도 떨어짐을 감수해야 하고, \mathbf{p}_2 는 실험 치료약제로서 효능이 동일하게 유지되면서 단지 안전성이 0.25 정도 증가함을 표현한다. 반면에 계획 2는 위의 \mathbf{p}_2 에서 안전성이 0.25 증가하는 대신에 0.20 정도 증가함을 목표로 두었기에 계획 1에 비해 \mathbf{p}_2 의 안전성에 대한 확률이 0.05 낮기 때문에 Thall과 Cheng (1999)의 표본수는 334에서 436으로, 그러나 WMW 방법의 표본수는 462에서 576으로 증가하였다. 계획 3은 계획 1의 목표모수에 $\mathbf{p}_2 = (0.1, 0)$ 을 하나 더 추가한 설계로서 귀무가설에 더욱 가까운 점이므로 Thall과 Cheng (1999)의 표본수는 334에서 744로, WMW 방법의 표본수는 462에서 890으로 크게 증가하였다. 계획 4는 효능과 안전성에 대한 차이의 확률 모두 증가하는 경우를 표현하여 실제로 임상시험에서는 불가능한 경우이며 계산된 표본수가 작음을 볼 수 있다. 표 4.2에서 계획 2의 경우 계획 1의 대립가설을 포함하기 때문에 표본수가 더욱 커졌고, 계획 3의 경우 계획 2의 대립가설을 포함하기 때문에 표본수가 더욱 크다.

이제 연관성의 척도인 오즈비 ψ_C 에 따라서 표본수의 크기가 어떻게 변하는지를 알아보기 위해서 표 4.1과 4.2에서 각각 계획 1에 대해서 오즈비를 여러 가지로 변화시켜 표본수를 계산하여 표 4.3과 4.4에

제시하였다. 기존 치료약제군의 셀의 확률 $\pi_{C,11} = \Pr(Y_{C,1} = 1 \text{ and } Y_{C,2} = 1)$ 은 오즈비에 의존하기 때문에 표 4.3과 4.4에 표본수와 함께 이 확률도 제시하였다. 표 4.3에서 오즈비가 증가할수록 표본수가 커지는데 오즈비가 무한대로 커져도 표본수의 변화폭은 그리 크지 않다. 표 4.4에서도 표 4.3과 비슷한 양상을 보이지만 표본수의 변화폭이 크다.

5. 결론

임상시험의 표본수 결정과정에 대해 많은 연구가 진행되었지만, 이변량 변수에 대한 표본수 계산 과정은 그리 많이 알려져 있지 않다. 본 논문에서 제안하는 효능과 안전성의 이변량 순위변수의 표본수 결정방법인 이변량 WMW 검정법은 분포의 가정없이 사용할 수 있으며 여러 범주 $R \times C$ 로의 확장이 용이하다. 표본수 계산에 사용된 예제에서 보면 효능과 안전성의 분포는 일반적으로 그 양상이 매우 다르며 이러한 사실이 표본수 결정에 영향을 미치고 있다. 한편 Thall과 Cheng (1999)은 비율에 분산 안정화 변환을 적용하여 표본수를 결정하는 방법으로 기각역을 정하기 위해 효능과 안전성의 모수를 $\{\Delta : \Delta^1 = \Delta^2\}$ 방향으로 움직임에 있어서 변환 $\Delta^j = g(\theta_{T,j}) - g(\theta_{C,j})$ ($j = 1, 2$)을 사용하고 여기서 $g(\theta_{i,j}) = \sin^{-1}(\theta_{i,j})^{1/2}$ 이므로 그 결과로서 변환된 척도에서의 45도 각도의 이동이지만 비율, 즉 $\theta_{i,j}$ 의 척도에서는 45도가 아닌 각도의 이동이다. 그러나 WMW 통계량을 이용한 방법은 확률을 그대로 사용하므로 45도 각도 이동으로 효능과 안전성의 대등함을 유지한 상태의 표본수 결정방법이다.

계산된 표본수는 표 4.1의 경우 Thall과 Cheng (1999)의 방법보다 WMW 방법의 표본수가 적어서 매우 고무적이며, 만약 이항변수의 경우 더욱 적은 표본수를 계산해 줄 수 있다면 순위변수로 확장했을 때도 역시 성립되리라고 짐작할 수 있다. 특히 WMW 방법을 이항변수의 경우에 적용하는 것은 동점 자료가 많아지는 결과를 초래하여 WMW 검정통계량의 효율성이 떨어질 것인데도 표본수가 더욱 적었다. 그러나 표 4.2의 경우에는 더욱 크게 계산되었다. 이러한 결과의 한 가지 이유는 표 4.1의 기존 치료약제의 효능과 안전성의 대립가설 하에서의 확률은 $\theta_C = (\theta_{C,1}, \theta_{C,2}) = (0.2, 0.95)$ 이고 목표모수점의 비율도 이와 비슷한 값이어서 모두 극한값인 0과 1에 가까운 비율이므로 이항비율의 변환 후에도 근사 정규성이 성립된다고 볼 수 없는 경우이고 따라서 특정 분포에 대한 가정이 없이 제안된 WMW 방법이 더욱 적절한 경우이다. 표 4.2는 경우 기존 치료약제의 효능과 안전성의 대립가설 하에서의 확률은 $\theta_C = (\theta_{C,1}, \theta_{C,2}) = (0.7, 0.62)$ 이고 또한 이에 가까운 목표모수점의 비율도 모두 근사 정규성이 성립되는 비율이기 때문에 Thall과 Cheng (1999)의 방법에 의한 표본수가 더욱 적었다.

표본수 결정은 검정을 전제로 한다. 따라서 여러 trade-off 목표모수의 최소 불복집합으로 표현된 기각 영역의 검정법을 언급해야 하는데 Thall과 Cheng (1999)의 논문에는 각 목표모수에 대한 검정법이 본 논문의 3.2절에서와 같이 설명되었을 뿐 구체적인 언급을 찾아볼 수 없다. 실제로 표본수 결정은 여러 목표모수가 있을 때 각 목표모수점에서의 검정력이 모두 미리 정해진 수준보다 큰 경우의 표본수로 결정한다. 이에 대응되는 검정은 귀무가설과 가장 가까운 거리에 있는 기각영역에 속한 대립가설에 대한 검정이 되며 이를 모의실험 자료의 검정력으로 확인하였다.

참고문헌

- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*, John Wiley & Sons, New York.
- Birnbaum, Z. W. and Klose, O. M. (1957). Bounds for the variance of the Mann-Whitney statistic, *The Annals of Mathematical Statistics*, **28**, 933-945.
- Bryant, J. and Day, R. (1995). Incorporating toxicity considerations into the design of two-stage phase II clinical trials, *Biometrics*, **51**, 1372-1383.
- Conaway, M. R. and Petroni, G. R. (1995). Bivariate sequential designs for phase II trials, *Biometrics*, **51**,

- 656-664.
- Conaway, M. R. and Petroni, G. R. (1996). Designs for phase II trials allowing for a trade-off between response and toxicity, *Biometrics*, **52**, 1375-1386.
- Cook, R. J. and Farewell, V. T. (1994). Guidelines for monitoring efficacy and toxicity responses in clinical trials, *Biometrics*, **50**, 1146-1152.
- Cook, R. J. (1996). Coupled error spending functions for parallel bivariate sequential tests, *Biometrics*, **52**, 442-450.
- DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach, *Biometrics*, **44**, 837-845.
- Hochberg, Y. (1981). On the variance estimate of a Wilcoxon-Mann-Whitney statistic for group ordered data, *Communications in Statistics - Theory and Method*, **A10**, 1719-1732.
- Jennison, C. and Turnbull, B. W. (1993). Group sequential tests for bivariate response: Interim analysis of clinical trials with both efficacy and safety endpoints, *Biometrics*, **49**, 741-752.
- Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco.
- Levine, M. N., Guyatt, G. H., Gent, M., De Pauw, S., Goodyear, M. D., Hryniuk, W. M., Arnold, A., Findlay, B., Skillings, J. R. and Bramwell, V. H. (1988). Quality of life in stage II breast cancer: An instrument for clinical trials, *Journal of Clinical Oncology*, **6**, 1798-1810.
- Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1990). Maximum likelihood regression methods for paired binary data, *Statistics in Medicine*, **9**, 1517-1525.
- Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate plackett distribution, *Journal of the American Statistical Association*, **89**, 633-644.
- Noether, G. E. (1987). Sample size determination for some common nonparametric tests, *Journal of the American Statistical Association*, **82**, 645-647.
- Norwood, P. K. and Sampson, A. R. (1988). A statistical methodology for postmarketing surveillance of adverse drug reaction reports, *Statistics in Medicine*, **7**, 1023-1030.
- Petri, H., Leufkens, H., Naus, J., Silkens, R., van Hessen, P. and Urquhart, J. (1990). Rapid method for estimating the risk of acutely controversial side effects of prescription drugs, *Journal of Clinical Epidemiology*, **43**, 433-439.
- Pocock, S. J., Geller, N. L. and Tsiatis, A. A. (1987). The analysis of multiple endpoints in clinical trials, *Biometrics*, **43**, 487-498.
- Simonoff, J. S., Hochberg, Y. and Reiser, B. (1986). Alternative estimation procedures for $\Pr(X < Y)$ in categorized data, *Biometrics*, **42**, 895-907.
- Thall, P. F. and Cheng, S. C. (1999). Treatment comparisons based on two-dimensional safety and efficacy alternatives in oncology trials, *Biometrics*, **55**, 746-753.
- Thall, P. F. and Lachin, J. M. (1988). Analysis of recurrent events: Nonparametric methods for random-interval count data, *Journal of the American Statistical Association*, **83**, 339-347.
- Thall, P. F. and Russell, K. E. (1998). A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in phase I/II clinical trials, *Biometrics*, **54**, 251-264.
- Tubert-Bitter, P., Bloch, D. A. and Raynauld, J. (1995). Comparing the bivariate effects of toxicity and efficacy of treatments, *Statistics in Medicine*, **14**, 1129-1141.

Determination of Sample Sizes of Bivariate Efficacy and Safety Outcomes

Hyun-Hak Lee¹ · Hae-Hiang Song²

¹Department of Biostatistics, Medical College, The Catholic University;

²Department of Biostatistics, Medical College, The Catholic University

(Received December 2008; accepted January 2009)

Abstract

We consider sample-size determination problem motivated by comparative clinical trials where patient outcomes are characterized by a bivariate outcome of efficacy and safety. Thall and Cheng (1999) presented a sample size methodology for the case of bivariate binary outcomes. We propose a bivariate Wilcoxon-Mann-Whitney(WMW) statistics for sample-size determination for binary outcomes, and this nonparametric method can be equally used to determine sample sizes of ordinal outcomes. The two methods of sample size determination rely on the same testing strategy for the target parameters but differs in the test statistics, an asymptotic bivariate normal statistic of the transformed proportions in Thall and Cheng (1999) and nonparametric bivariate WMW statistic in the other method. Sample sizes are calculated for the two experimental oncology trials, described in Thall and Cheng (1999), and for the first trial example the sample sizes of a bivariate WMW statistic are smaller than those of Thall and Cheng (1999), while for the second trial example the reverse is true.

Keywords: Sample sizes, bivariate WMW, efficacy and safety.

²Corresponding author: Professor, Department of Biostatistics, Medical College, The Catholic University of Korea, Seoul 137-701, Korea. Email: hhsong@catholic.ac.kr