

Estimation of Maximal Tolerated Dose in Sequential Phase I Clinical Trials¹⁾

In-Hye Park²⁾ and Hae-Hiang Song³⁾

Abstract

The principal aim of a sequential phase I clinical trial, in which the toxicity responses of a group of patient(s) determine the dose level of the next patient(s) group, is to estimate the maximal tolerated dose(MTD) of a new drug. In this paper we compared with a simulation study the performance of the MTD estimates that are determined by a stopping rule in a design and also those that are determined by analyzing the data after a clinical trial is terminated. To the latter belong the mean, median, mode and maximum likelihood estimates.

For the Standard Methods the stopping rule MTD is quite inefficient but the median MTD has a best efficiency and is robust with respect to the three different toxicity curves. The problem of non-convergence of MLE MTD is severe. A more improved MTD estimate is produced by combining the advantages of the various MTD estimates and its efficiency is better than the single median MTD estimate especially for the toxicity curve of an unlucky choice of dose levels. The simulation results suggest that simple types of phase I designs can be combined with relatively standard analytic techniques to provide a more efficient MTD estimate.

I. 서 론

제 1상 임상시험은 새로이 개발된 약제를 사람에게 처음으로 적용하는 단계로서 안전한 1회 복용량 범위의 추정을 목적으로 한다. 즉, 사람에게 허용할 수 있는 독성수준에서 최대 효능을 가진 용량 범위의 추정이며(O'Quigley & Chevret, 1991), 또는 목표로 정한 독성 반응율을 초과하지 않는 복용량의 추정인데(Chevret, 1993) 이를 최대허용용량(maximal tolerated dose; MTD)의 추정이라 한다. 여기서 MTD는 독성이 나타나지 않는 용량 범위 중에서 최대 복용량을 뜻하거나, 또는 독성반응이 나타나는 용량 범위 중에서 최소 복용량을 뜻하며(Korn et al., 1994; Goodman et al., 1995), 좀 더 구체적으로는 용량에 따른 독성 반응율을 표현한 백분위수에서 대략 33분위수를 MTD로 정하고 있다(Storer, 1989). 실험의 특징에 따라서 50%의 환자에게서 독성반응이 나타난 복용량을 중앙효과용량(median effective dose; MED)이라고 구분하여 이를 추정하기도 한다(O'Quigley & Chevret, 1991). 이와 같이 추정된 복용량은 제 2상과 제 3상의 임상시험에서 사용

1) 이 연구는 가톨릭대학교 연구비(CMC)의 지원으로 연구되었음

2) Department of Biostatistics, Catholic University, Medical College, Seoul, 137-701, Korea

3) Professor, Department of Biostatistics, Catholic University, Medical College, Seoul, 137-701, Korea

될 것이다.

MTD 추정을 위해 제시된 기존의 제 1상 임상시험 계획법으로는 비모수적 방법인 Up-and-Down 방법, 이의 변형인 그룹 Up-and-Down 방법, 두 단계 Up-and-Down 방법, 또한 베이지안 방법을 적용시킨 연속재평가방법(continual reassessment method; CRM)이 주류를 이루고 있다(O'Quigley & Chevret, 1991). Up-and-Down 방법은 계획이 단순하고 환자수를 결정하는 것으로 실험 전 준비가 끝나지만, 이 방법은 목적하는 독성 반응율에 해당하는 MTD를 제대로 제시하지 못하거나 일부 환자는 치료효과가 없는 낮은 용량수준에 할당되고 있는 문제점이 지적되었다(Mick & Ratain, 1993). 이와 달리 CRM은 때로는 기존의 계획법보다 적은 환자수로 추정된 MTD가 높은 정도(precision)를 가지며(O'Quigley & Chevret, 1991), 단점을 계속 보완한 CRM 방법이 제시되고 있고(Goodman et al., 1995), 최근에는 우도적(likelihood) 접근도 제시되었다(O'Quigley & Shen, 1996). 그러나 CRM은 매 용량 단계마다 다음 환자의 복용량을 결정하기 위해 적분계산을 해야하기 때문에 현실 적용에의 어려움을 부인할 수가 없다. 이러한 점에 감안하여 본 논문에서는 현실 적용이 용이한 Up-and-Down 방법들을 개선하여 제 1상 계획에서 좀 더 정확한 MTD추정의 가능성을 알아보고자 한다. 또한 계획 내에서 추정되는 MTD와 계획이 종결된 후 MTD를 추정하는, 계획과 분리된 추정의 효율성도 알아본다.

II. 제 1상 임상시험의 계획법과 MTD 추정법

제 1상 임상시험은 환자들을 대상으로 독성 및 부작용의 제한된 반응만을 관찰하는 과정이며, 여기서 중요한 점은 사람이 연구 대상이므로 안전을 고려해야 한다는 것이다. 또한 제 1상 임상시험의 MTD를 찾는 방법에 있어서도 안전성(safety)을 고려해야 하므로 다음의 세가지 대책이 제시되고 있다(O'Quigley & Chevret, 1991). 첫째는 잠재 독성(potential toxicity)을 짐작할 수 있는 동물실험의 결과를 임상시험에 이용하는 것이며, 둘째는 건강한 지원자를 대상으로 실험하는 것이고, 셋째는 용량을 서서히 증가시키는(slow dose escalation) 방법이다. 만약 홍역과 같은 사람에게만 발생하는 질병의 경우에는 첫번째 대책은 부적절하고, 치료제의 효과를 평가하고자 할 경우 또는 건강한 사람에게 치명적인 해를 끼치는 약제(예: 인슐린분비 촉진제)의 경우에는 두번째 대책이 부적절하다. 세번째 대책은 매우 안전한 방법일 수 있으나 많은 시간이 걸리며 생명이 위급한 환자에게 치료가 즉시 적용되어져야 할 경우에 위험할 수 있다. 따라서 안전책은 연구하고자 하는 약제와 환자의 특성에 따라 적절하게 적용되어야 한다.

본 논문에서는 항암 또는 AIDS 치료제에 대한 신약의 제 1상 임상시험을 중점으로 다루며, 적은 환자수로서 동물실험 결과를 이용하여 낮은 용량에서부터 출발하여 목표로 정한 독성반응율에 도달할 때까지 서서히 용량을 증가시킨다. 신약의 낮은 용량에서는 무효능일 것이며 높은 용량에서는 심각한 독성반응이 있을 것이라는 가정 하에서 진행하며, 또한 복용량과 독성반응율의 관계에 대해서 어느 정도 짐작하고 있는 상태에서 출발한다.

2.1 제 1상 임상시험의 계획법

매 용량수준에 한명의 환자를 배정하는 계획은 Dixon과 Mood(1948)에 의해 제시되었다. 그 후

Wetherill(1963)은 각 용량수준에 2명 또는 그 이상을 실험단위로 정하는 계획을 제안하였으며, Storer(1989)와 Korn et al.(1994)은 표준방법(Standard Method; SM)이라 알려진 각 용량수준에 3명을 배정하는 여러 계획들을 제시하였다. 한편 베이지안 방법은 Gatsonis와 Greenhouse(1992)에 의해 제시되어 O'Quigley와 Chevret(1991), O'Quigley(1992), Piantadosi와 Liu(1996)에 의해 CRM 방법이 적극 장려되었다. 그리고 제 1상 임상시험에 대한 최근의 연구들 대부분이 여러 표준방법들과 CRM 방법의 비교에 대한 것이다. 한편 Korn et al.(1994)은 CRM 방법의 문제점을 지적하면서 표준방법을 수정할 것과 또한 앞으로도 표준방법을 계속 사용할 것을 주장하고 있다. 따라서 본 연구는 표준방법의 수정의 가능성과 알아보는 방향으로 진행하게 되며 본 연구에서 비교할 계획들을 아래에 간략히 설명한다.

2.1.1 단순 Up-and-Down 방법 (또는 DM 방법)

제 1상 임상시험에서 복용량 단계는 일반적으로 로그 단위로 표현된 복용량 수준에서 동일 간격으로 정하며 이와 같은 복용량 수준을 x_i ($i = 1, 2, \dots, k$)로 표시한다. 각 환자에게 주는 신약의 복용량이 다르며, $x(j)$ 는 j 번째 환자의 복용량 수준을 나타내고, Y_j 는 j 번째 환자에게 나타난 독성반응으로서 1(반응) 또는 0(무반응)의 값을 갖는 이항 확률변수이다 ($j = 1, 2, \dots, n$). 여러 계획에서 환자(군)에게 주는 용량은 일반적으로 최근 환자(군)의 독성반응의 유무에 따라 결정된다.

독성반응을 곡선에서 목표로 하는 확률을 θ 라 할 때, 제 1상 임상시험의 각 계획들은 정해진 θ 에 해당하는 용량수준을 추정토록 계획되었다. 단순 Up-and-Down 계획은 MED, 즉 $\theta = 0.5$ 를 추정하기 위해 Dixon과 Mood에 의해 제안되어 DM 방법이라고도 불리우며, 계획의 시행 절차는 다음과 같다. 먼저 동일 용량수준 간격을 가진 복용량 (x_1, x_2, \dots, x_k)을 정한다. 첫번째 환자는 첫 용량 수준 $x(1) = x_1$ 에서 시작하여 다음 환자의 용량수준은 아래와 같이 정한다.

$$x(j+1) = \begin{cases} x(j) + dI(x(j) < x_k), & Y_j = 0 \text{ (독성반응이 없을 때)} \\ x(j) - dI(x(j) > x_1), & Y_j = 1 \text{ (독성반응 일 때)} \end{cases} \quad (1)$$

즉, 독성반응이면 용량을 감소시키고 아닌 경우에는 용량을 증가시킨다. 여기서 $d = x_{i+1} - x_i$ ($i = 1, 2, \dots, k-1$)로 용량수준 간격이며, 함수 I 는 지시함수(indicator function)로 참일 때 1이며 거짓일 때 0의 값을 가진다.

Dixon과 Mood는 미리 정해진 표본수에서 시행을 종료한 후 수집된 자료에 프로빗(probit) 모형과 점근 최대우도 추정량을 이용하여 MED를 구하였다. 그러나 DM 방법에서는 모형을 이용하지 않고 시행이 종료된 시점의 멈춤용량수준(stopping dose)을 MED의 추정값으로 제시하는 방법을 사용하고 있다.

2.1.2 그룹 Up-and-Down 방법

Wetherill(1963)은 다수 환자의 반응에 기초한 여러 Up-and-Down 방법들을 제시하였다. 그 한 예가 2명 환자 반응이 (0, 0) 또는 (1, 1)로 서로 같으면 DM 방법에서와 같이, 전자의 경우 다음 환자군의 복용량을 증가시키고 후자의 경우 복용량을 감소시킨다. 그러나 2명 환자의 반응결과가 (0, 1)과 (1, 0)으로 서로 다르면, 다시 한명 환자를 동일 복용량에 추가로 적용시켜 3명중 2명의 반응에 따라 복용량을 증가 또는 감소시킨다.

현재 가장 많이 사용되고 있는 표준방법(Standard Method; SM) 또는 전통적인 계획(traditional design)이라 일컬어지는 계획들은 33% 독성반응율에 해당하는 MTD를 추정하기 위해 각 용량수준에 기본적으로 3명의 환자가 배정되는 계획이며, 연구자에 따라서 조금씩 다른 계획으로 제시되었다. 여기서 다른 계획이란 우선 용량의 증가, 감소 또는 동일 용량에 머무르는 결정 규칙이 다르다는 뜻이며, 그리고 MTD를 결정하는 방법이 다르다는 뜻이다. MTD 추정치는 계획에 따라 다른 것은 물론이며 동일한 계획법이라 하더라도 MTD의 결정 방법에 따라 다르게 된다(Storer, 1989; Goodman et al., 1995). 곧 설명하게 되는 Storer(1989)의 Design A, Korn et al.(1994)의 SM3 계획과 SM6 계획은 면출규칙에 의해 MTD를 결정하며, Storer(1989)의 Design D는 미리 정해진 표본수에서 시행을 종료한 후 MTD를 추정한다.

a. Storer(1989)의 Design A

실험 전에 복용량 수준이 정해지며 첫번째 환자군이 첫번째 용량수준을 복용하는 것은 DM 방법과 같다. 매 단계의 실험이 3명의 환자군으로 구성되는데 이 3명의 환자 중 아무도 독성반응을 보이지 않으면 다음 환자군은 한 단계 높은 복용 수준을 적용 받는다. 한명이라도 독성반응을 보이면 새로운 3명의 환자가 동일 복용수준을 적용 받으며 이를 총 6명중 1명이 독성반응을 나타내면 다음 환자군은 한 단계 높은 복용수준을 적용 받고, 나머지 경우에는 실험을 종료하고 (Ⓐ) 지점) 면출용량을 MTD로 결정한다.

Storer(1989)가 전통적인 계획이라고 일컬고 있는 이 계획은 그가 진행한 모의실험 연구에서 결과가 좋지 않은 것으로 밝혀졌는데 그 이유는 3명의 환자군에서 독성반응이 나타나 6명의 환자가 동일 용량수준에 배정되었을 때, 즉 위의 Ⓩ 지점에서, 6명중 1명을 제외한 2/6, 3/6, 4/6, 5/6, 6/6과 같은 독성반응율 결과에 대해 이를 구분함이 없이 실험을 종료하면서 면출용량을 MTD로 제시하므로 예민한 추정량이 되지 못하는 것이다.

b. Storer(1989)의 Design D

이 계획은 미리 정해진 표본수에서 실험을 종료한 후 모형에 의존하여 MTD를 추정하기 때문에 고정 표본수에 도달할 때까지 용량 증가, 감소 또는 동일 용량을 유지할 규칙을 제시하게 된다. 즉 3명의 환자 중 아무도 독성반응을 보이지 않는 경우에는 다음 환자군은 한 단계 높은 복용 수준을 적용 받으며, Design A와는 달리 2명 이상이 독성반응을 나타내면 한 단계 낮은 복용 수준을 적용 받는다. 나머지 경우, 즉 3명 중 1명의 환자가 독성반응을 나타낸 경우에는 추가로 3명이 동일 복용량을 받게 되며 이를 총 6명중 1명이 독성반응을 나타내면 다음 환자군은 한 단계 높은 복용수준을 적용 받고, 총 6명중 3명 이상이 독성반응을 나타내면 다음 환자군은 한 단계 낮은 복용수준을 적용 받고, 총 6명중 2명이 독성반응을 나타내면 다음 환자군은 동일 복용량을 적

용 받는다. 즉, 첫 3명은 $x(1) = x(2) = x(3) = x_1$ 을 복용한 후 다음 환자군($j=4, \dots, n-2$)의 복용수준은 아래와 같이 결정된다. O'Quigley와 Chevret(1991)는 이 Design D 방법을 CRM 방법과 비교하였다.

$$x(j+2) = x(j+1) = x(j) = \begin{cases} x(j-1) + dI(x(j-1) < x_k) & \text{if } \sum_{k=j-3}^{j-1} Y_k = 0 \\ x(j-1) - dI(x(j-1) > x_1) & \text{if } \sum_{k=j-3}^{j-1} Y_k > 1 \\ x(j-1) & \text{otherwise} \end{cases} \quad (2)$$

c. Korn et al.(1994)의 Standard Method 3(SM3)와 Standard Method 6(SM6)

표준방법에는 SM3와 SM6 계획이 있는데 이 표준방법들은 환자들의 독성반응이 면출규칙의 조건에 만족되면 시행을 종료하고 MTD를 제시하게 된다.

SM3 계획은 3명 환자 중 아무도 독성반응을 보이지 않는 경우에는 다음 환자군은 한 단계 높은 복용량을 적용 받으며 2명 이상이 독성반응을 나타내면 실험을 종료하고 한 단계 낮은 복용수준을 MTD로 결정하며, 3명 중 1명의 환자가 독성반응을 나타낸 경우에는 (독성반응율: 1/3) 추가로 3명이 동일 복용수준을 적용 받으며 이를 총 6명 중 1명이 독성반응을 나타내면 (독성반응율: 1/6) 다음 환자군은 한 단계 높은 복용수준을 적용 받고, 나머지 경우에는 (독성반응율: 2/6, 3/6, 4/6 - ④) 실험을 종결하고 한 단계 낮은 용량수준을 MTD로 결정한다. Goodman et al.(1995)은 이 SM3 방법을 CRM 방법과 비교했으며, SM3 방법이 현재 미국에서 주로 사용되고 있는 표준방법이라 설명하고 있다.

SM6 계획은 위의 SM3 방법과 동일하게 진행되지만 실험을 종결하고 한 단계 낮은 용량수준을 MTD로 결정할 때 이 낮은 용량수준에 만약 3명의 환자 자료만이 있다면 다시 3명을 이 낮은 용량에 추가한 후 앞에서와 같이 시행하다가 MTD를 제시하는 방법으로서 SM3 와 구분 지어 이를 SM6 계획이라 한다. 따라서 SM6 계획에서는 SM3 계획보다 더 많은 환자수가 요구된다.

일반적으로 실험을 종료한 후 결정된 용량수준을 MTD로 일컬고 있지만, Korn et al.(1994)은 MTD를 허용할 수 없는 (unacceptable) 독성 용량수준으로 정의하고 있으며, 이러한 정의에 의해 저자들은 실험을 종료한 후 한 단계 낮은 용량수준을 MTD로 제시하고 있다.

한편, SM3 또는 SM6 계획에서 33%이면 면출용량수준을 MTD로 정하고, 33%를 초과하면 한 단계 아래 낮은 용량수준을 MTD로 정하도록 세분한다면 이는 더욱 보완된 면출규칙이 된다. 즉, 본 연구에서 제시하는 첫번째 수정안은 위의 ④에서 독성반응율이 2/6이면 면출용량수준을 MTD로 정하고, 독성반응율이 3/6과 4/6이면 한 단계 아래 낮은 용량수준을 MTD로 정하는 좀 더 세분된 계획이며, 이를 ‘수정 SM3 계획’과 ‘수정 SM6 계획’이라 하고 위의 계획들과 함께 비교할 것이다.

2.1.3 두 단계 Up-and-Down 방법

제 1상 임상시험은 특징적으로 낮은 용량수준에서 출발하여 용량을 증가시키고 있다. 낮은 용량

에서 다수 환자의 소모를 되도록 줄이기 위해서 1명 환자에 의존하는 DM 방법으로 용량을 증가시키다가 1명 환자의 결과가 독성반응일 때에 비로서 3명을 배정하는 표준방법을 적용하는 두 단계 계획인 BC와 BD 계획을 Storer(1989)는 제시하였다. 여기서 Storer가 말하는 첫단계에 해당하는 B 방법은 바로 DM 방법이며, D 계획은 이미 앞에서 설명하였다. 아직 설명치 못한 C 계획은 DM 방법과 비슷하지만 2명의 연속적인 무반응을 확인한 후에 용량을 증가시킴이 DM 방법과 다른 점이다.

2.2 MTD 추정법

제 1상 임상시험의 MTD 추정법을 요약하면 다음과 같다. 첫째는 계획법 내에 MTD 추정이 포함되어 있는 형태, 즉 용량의 증가와 감소에 대한 멈춤규칙의 적용으로 실험의 종료와 동시에 MTD를 제시하는 방법인데 계획에 따라서 멈춘 후에 멈춤용량수준(stopping dose), 또는 한단계 아래 용량수준(lower dose), 또는 추천용량수준(recommended dose)을 MTD로 제시하고 있다. 이 멈춤규칙에 의해 실험을 종료하는 경우의 표본수는 매우 유동적이다. 둘째는 미리 정해진 표본수에서 실험을 종료하고서 MTD를 추정하는 방법인데, 위와 같이 (a) 멈춤용량수준, 한단계 아래 용량수준, 또는 추천용량수준을 MTD로 제시하거나, (b) 수집된 자료에 로짓모형 또는 다른 모형을 적합시켜 MTD를 추정하거나 또는 비모수적 방법의 적용으로 MTD를 추정하는 방법으로 다시 구분된다. 여기서 추천용량수준은 미리 정해진 표본수(n)에서 멈춘 후, n 번째 환자의 반응에 따라 증가 또는 감소된 용량수준 $x(n+1)$ 으로 MTD를 결정하는 방법이며 주로 DM 방법 또는 Storer(1989)가 제시한 BC 방법, BD 방법에서 사용한다. 멈춤규칙에 의해 멈춤용량수준 또는 한단계 아래 용량수준은 주로 SM 계획에서 사용한다.

필요한 정보가 충분히 수집된 후 멈춤규칙에 의해 실험을 종료하는 것이라면 첫째 방법이 고정 표본수 방법보다 훨씬 효율적일 수 있겠고, 또한 실험 종료와 더불어 즉시 MTD를 추정할 수 있다는 장점이 있다. 그러나 이 추정법에서 실험의 종료에 영향을 미치는 것은 마지막 환자군의 반응뿐이다. 따라서 매우 다양한 환자들이 실험에 참여하는 순서가 달라짐에 따라 MTD를 잘못되게 추정할 수 있는 가능성도 가지고 있는 것이다.

Storer(1989)에 의하면 여러 계획법은 단순히 자료를 효율적으로 수집하는 과정이며, 환자들의 독성반응에 대한 자료수집이 끝난 후 이 자료를 분석하여 MTD는 추정되어야 한다. 실제로 여러 계획법의 용량의 증가, 감소 규정은 목표로 정한 독성 반응율에 해당하는 용량을 추정토록 정해지는 것이므로 실험에서 환자들이 복용한 용량수준들은 목표로 정한 MTD를 중심으로 분포되어 있을 것이라는 사실에 기반을 두고서 다음과 같은 MTD 추정량을 제시할 수가 있다. 이들은 자료분석으로 MTD를 추정하는 방법으로 모형에 의존한 MLE 방법을 제외하고는 모두가 비모수적 방법들이다.

(1) Mean (평균)

여러 연구자들(Durham et al.(1997); Hsi(1965); Wetherill(1963); Brownlee et al.(1953))이 평균을 DM 방법에서 사용했으나 표준방법에서 사용한 경우는 없으며 또한 이 추정량을 멈춤규칙 MTD 추정량과 구체적으로 비교한 연구는 없다. 실험에서 다루어진 모든 복용량 $x(1), \dots, x(n)$

의 표본평균 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x(i)$ 를 MTD 추정값으로 제시한다.

(2) Median (중앙값)

실험에서 다루어진 모든 복용량 $x(1), \dots, x(n)$ 의 순위 통계량 $x_{(1)} \leq \dots \leq x_{(n)}$ 중에서 중앙값 \tilde{x} 를 MTD 추정값으로 제시한다. k 번째 분위수의 추정량은 $\tilde{x}_k = (1-g)x_{(j)} + g x_{(j-1)}$ 이며, 여기서 $(n+1)k = j + g$ 이고 j 는 $(n+1)k$ 의 정수부분에 해당한다. 중앙값 \tilde{x} 는 $k = \frac{1}{2}$ 일 때이다.

(3) Mode (최빈값)

실험에서 다루어진 모든 복용량 $x(1), \dots, x(n)$ 중에서 가장 빈번하게 출현한 값을 MTD 추정값으로 제시한다. 실제 계산에서 최빈값이 동점으로 여럿인 경우에는 이들의 평균값의 반올림을 MTD 추정값으로 정하였다.

(4) Maximum Likelihood Estimator (최대우도 추정량, MLE)

미리 정해진 표본수에서 임상시험을 종료하는 경우에 사용하는 추정법으로서, 로짓모형(logistic model)을 이용한 모수의 최대우도 추정량(MLE)으로부터 MTD를 정하는 모수적 방법이다. 이 MLE 추정량의 장점은 용량수준 사이의 값이나 외삽(extrapolation)이 가능하여 만약 참 MTD가 용량범위밖에 위치할 때에 이를 추정할 수가 있다는 것이다.

로짓모형으로 표현된 독성반응율 $P(Y=1 | X=x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$ 에서 최대우도추정량 $\hat{\alpha}$, $\hat{\beta}$ 로 표현된 p 분위수에 해당하는 추정 용량은 다음과 같다.

$$\hat{x} = \frac{\ln\left(\frac{p}{1-p}\right) - \hat{\alpha}}{\hat{\beta}} \quad (3)$$

여기서 50%를 목적으로 하는 MLE MTD는 $p = \frac{1}{2}$ 인 경우로서 $\hat{x}_{MLE} = -\hat{\alpha}/\hat{\beta}$ 이며,

33%를 목적으로 하는 MLE MTD는 $p = \frac{1}{3}$ 인 경우로서 $\hat{x}_{MLE} = \frac{\ln(\frac{1}{2}) - \hat{\alpha}}{\hat{\beta}}$ 이 된다.

네가지 추정량의 서로 다른 통계적 성질은 이미 알려져 있으며 따라서 제 1상 임상시험 자료에서 네가지 MTD 추정량의 효능도 짐작할 수 있으나, 실제로 환자들의 독성반응에 따라 다음 용량수준이 결정되는 제 1상 임상시험의 축차 자료에서 어떤 추정량이 과연 좋을지는 모른다. 모수적 추정법인 로짓모형을 이용한 MLE MTD는 대표본의 경우에 작은 편의와 분산 외에 여러 통계적 성질이 뛰어난 것으로 알려졌으나 제 1상 임상시험의 특징인 소표본의 경우 Newton-Raphson 방법에 의해 계산되는 MLE MTD는 비수렴하는 경우(Silvapulle, 1981)가 자주 발생하기 때문에, 다른 MTD 추정량이 제시되어야 함을 뜻하고 있다.

Durham et al.(1997)은, 모의실험에서 곧 설명하게 되는, 편의동전 계획(Biased Coin Design)을 제시하였는데 이는 DM 방법에 확률을 도입한 계획으로서 저자들은 본 연구의 곡선 1에 해당하는 이론적인 곡선에 대해 MLE, 평균, 최빈값 MTD 추정량의 모의실험 결과를 표본수 25와 50에서 비교하였다. DM 방법의 경우, 표본수 25에서 MLE의 편의가 가장 작았으나 MLE의 분산은 평균, 최빈값보다 커으며, 표본수 50에서도 마찬가지의 결과이었다. 또한 평균의 분산이 중앙값, 최빈값의 분산보다도 작은 것으로 알려져 있으며, 최빈값은 편의와 분산이 모두 크다.

여러 추정량을 비교해 보면 장, 단점이 서로 다르고, 특히 어떤 독성반응율을 곡선인가에 따라 바람직한 추정량이 서로 다르다. 연구자는 실제로 독성반응율 곡선에 대한 정확한 지식이 없기 때문에 실험을 진행하는 것이므로 본 연구에서 제시하는 두번째 수정안은 소표본의 경우에 MLE 추정량은 빈번하게 비수렴되므로 MLE를 제외하고, 면출규칙에 의한 MTD 추정량과 자료수집이 끝난 후에 쉽게 구해지는 평균, 중앙값, 최빈값을 중심축으로 요약한 종합추정량을 제시하고자 한다. 이 종합추정량은 여러 추정량의 장점을 함께 참고하여 제시하므로 로버스트(robust) 방법이다. 본 논문에서는 네가지 추정량의 평균에서 가장 가까운 용량수준을 궁극적인 추정량으로 정하는 종합추정량을 제시한다.

III. 모의실험

모의실험은 단순 Up-and-Down 계획법인 DM 방법, 그의 변형인 편의동전 계획(Biased Coin Design), 그룹 Up-and-Down 방법인 SM3, SM6, 그리고 두 단계 BC와 BD 계획법에 대해서 실시한다. DM 방법은 표본수가 달라짐에 따라 면출규칙의 효능이 어떻게 변해 가는지를 이해하는데 도움이 되며, Dixon과 Mood(1948)의 논문을 근거로 하여 실험의 표본수를 10, 25, 50으로 선정하였고, SM3와 SM6는 각기의 면출규칙에 의해 실험이 종료되며, 두 단계 계획법은 두번째 단계인 C와 D 계획의 표본수를 12(6)36으로 변화시켜 실시하였다.

모의실험의 난수발생은 IMSL(International Mathematics and Statistics Library)에서 제공하는 난수발생 서브루틴 DRNUNF를 사용하며, 각 방법의 1000번 반복에 대한 결과를 제시한다. 복용량 수준은 제 1상 임상시험의 모의실험에서 가장 많이 사용되고 있는 6 단계로 정하였으며, 용량수준 간격은 동일 간격 $d=1$ 로 정하였다. MTD 추정에서 MLE는 반복회수가 100회 이상인 경우에는 비수렴으로 간주하였다.

모의실험에서 연구한 각 수준의 독성반응율은 표 1과 같다. 곡선 1과 2는 목표로 하는 확률 $\theta = 0.33$ 을 중심으로 용량수준이 비교적 균일하게 분포되었고, 곡선 3과 4는 처음 용량수준부터 이미 높은 독성반응율을 보이고, 곡선 5와 6은 용량수준 4 또는 5까지 독성반응율이 매우 낮았으나 급격히 증가하는 경우이다.

Table 1. Family of true dose toxicity curves used in the simulations.
(All numbers represent percentage toxicity at that dose level)

Dose level	Curve 1	Curve 2	Curve 3	Curve 4	Curve 5	Curve 6
1	5	9	22	30	0	0
2	10	16	32	40	1	0
3	25	27	45	52	4	3
4	35	38	54	61	9	5
5	50	57	69	76	24	6
6	70	75	80	87	49	22

모의실험의 결과는 다음과 같다. 곡선 2, 4, 6의 결과는 각각 곡선 1, 3, 5와 비슷하여 홀수 곡선의 결과만을 논문에 제시한다.

3.1 DM 방법

3.1.1 DM 방법의 MED 추정

50% MED를 겨냥하는 DM 방법은 1명 환자의 독성반응에 의해 실험이 진행되므로 예외적으로 용량의 증가와 감소의 두가지 경우만 가능하며 따라서 구체적인 멈춤규칙이 있을 수 없고, 미리 정해진 표본수에서 종료하여 멈춤용량수준(stopping dose) 또는 추천용량수준(recommended dose)을 MTD로 결정하거나 또는 수집된 자료에 모형을 적합시켜 MTD를 추정한다. 추정이 잘 되었다는 기준은 실제로 MTD로 제시한 용량수준이 목적인 참(true) MTD를 제대로 지정하고 있는가가 중요하므로 멈춤용량수준 또는 추천용량수준의 빈도를 살펴보는 것에 중점을 둔다. 목적인 참 MTD를 지정하는 용량수준의 빈도가 높아야 하고 또한 이를 중심으로 되도록 좁게 분포되어 있어야 하며 이 빈도로부터 편의(bias)와 정도(precision)를 가늠할 수 있다. 실험에서 거쳐가는 용량수준(dosage level of experimentation)의 빈도와 관측 독성율(observed toxicity)은 실험의 안정성(stability)을 알려준다.

표 2-1, 2-2, 2-3은 곡선 1, 3, 5에 대한 DM 방법의 결과이다. 세 곡선에서 표본수가 증가함에 따라 낮은 용량수준에서 실험되어지는 빈도가 감소하고 높은 용량수준에서 실험되어지는 빈도는 증가하였다. 각 실험의 관측 독성율은 표본수가 증가하면서 높아지는데 그러나 모두가 50% 미만이다.

세 곡선에서 멈춤규칙에 의한 MTD를 살펴보면, $n = 10$ 과 같은 작은 표본수는 증가와 감소로서 멈추게 되는 지점이 자연스럽게 목적인 참 MTD 용량수준인 경우와 그렇지 못한 경우로 구분되어 결과가 크게 다르다. 다시 말하면 표 2-1에서 $n = 10$ 명으로서 멈춤규칙에 의해 멈춤용량수

준 4는 자연스럽게 도달되지만 참 MTD 수준인 용량수준 5는 용량수준 1 또는 6에서 더 이상 감소 또는 증가치 못하고 그대로 홀수회수를 더 머물러서 도달되는 경우이다. 한편, 추천용량수준은 $x(n+1)$ 으로 결정되는 방법이므로 면출용량수준과 한 용량씩 차이가 나서 용량수준 5가 자연스럽게 도달되는 용량이다. 이와 같이 최저와 최고 용량수준인 1 또는 6에 도달하여 더 이상 감소, 증가할 수 없었던 상태를 경계선 조건(boundary condition)이라 한다. 그러나 표본수 50으로 많아지면 소표본에서와 같이 참 MTD 수준, 또는 바로 아래, 또는 바로 위 용량수준에 집중되는 현상은 없어지지만 DM 방법은 계속 증가와 감소로 움직이게 되므로 MTD로 제시하는 용량이 넓게 분포되어 있다. 따라서 DM 방법에서 면출규칙은 소표본으로는 참 MTD 대신에 바로 아래, 또는 바로 위 용량을 지정하는 오류를 범할 확률이 매우 높으며 대표본의 경우에는 넓게 분포되어 있고 또한 높은 독성의 용량수준 빈도가 여전히 높다. 결론적으로 DM 방법에서 면출규칙만으로 MTD를 결정하는 것은 위험하다 하겠다.

이제 자료분석에 의한 MTD를 살펴보면 면출규칙에 의한 MTD 보다 안정되어 있고 표본수 50에서 편의와 정도가 크게 좋아진다. 곡선 1에서는 용량 5가 $\theta = 0.5$ 에 해당되는데 MLE MTD는 편의가 가장 작고 매우 커던 분산이 표본수가 증가할수록 작아지지만 무엇보다도 MLE MTD의 문제점은 수렴율이 낮아 수렴되는 경우가 50%를 넘지 못한다는 점이다. 또한 수렴된 MTD 중에서도 용량수준 6 이상인 추정값이 표본수 25에서 12% 가 된다. 이미 연구자들이 보고했듯이 수렴된 경우에서도 매우 큰 추정값이 간혹 제시되어 이러한 MLE MTD 추정값을 그대로 사용할 수가 없다(Storer, 1989). 평균은 표본수에 상관없이 결과가 낮게 추정되어 바람직하지 않지만 분산이 작으며, 반면에 최빈값의 분산은 큰데 예외적으로 곡선 5의 경우에는 최빈값의 편의가 작고 또한 분산도 작았다. 표본수 25 또는 50의 경우에 세 곡선에 걸쳐 중앙값의 효능이 가장 좋다.

3.1.2 DM 방법으로 여러 가지 θ 의 추정

DM 방법은 $\theta = 0.5$ 에 해당하는 용량수준을 추정하기 위한 방법이며 장점은 각 실험 단계에 1명 환자가 할당되므로 실험의 진행이 빠르다는 것인데 이러한 장점과 함께 여러 가지 θ 에 해당하는 용량수준의 추정에 사용할 수 있는 방법이 편의동전 계획(Biased Coin Design)이며, Durham et al.(1997; 1995)이 소개하였다. 편의동전 계획은 계획 내에 확률을 도입하여 독성반응이 나타나면 DM 방법과 동일하게 무조건 감소하지만 독성반응이 나타나지 않으면 확률 $b = \frac{\theta}{(1-\theta)}$ 에 의해 용량을 증가시키고 확률 $1-b$ 로 동일 용량수준에서 다시 실험하는 계획이다. 편의동전 계획에 대한 모의실험 결과는 표 3-1, 3-2, 3-3에 제시하였다.

표 3-1, 3-2, 3-3의 $\theta = 0.5$ 는 앞에서 설명한 단순 DM 계획이며 앞의 결과와 일치한다. $\theta = 0.10, 0.20, 0.25, 0.33$ 인 경우의 면출규칙에 의한 MTD 추정 결과를 살펴보면, 소표본에서 보였던 단순 DM 계획의 참 MTD 수준 바로 아래 또는 바로 위 용량수준에 집중되었던 현상은 확률의 도입으로 완화되었다. 그러나 표본수 증가에 따라 큰 변화를 볼 수 없는 것이 특징으로 면출규칙에 의해 MTD를 추정하고자 할 때 최저 용량수준에서 시작하여 참 MTD 용량수준에 도달하기에 충분한 표본수라면 그 이상의 큰 표본수는 별 도움이 되지 못하며 $n = 50$ 의 큰 표본수로도 참 MTD 용량수준을 실제로 MTD로 제시하는 확률이 41%를 넘지 못하고 있다. 어떠한 독성을 예상하는 용량수준이라도 추정할 수 있고, random walk rule의 이론적 배경의 장점을 가진

계획이지만(Durham et al., 1997), 실제 MTD 추정에는 유용한 계획이라 볼 수 없다.

3.2 SM 계획 및 수정 계획

독성반응율 33%에 해당하는 MTD를 추정하기 위한 표준방법의 모의실험 결과를 표 4와 5에서 살펴보면(SM3에서 표준계획과 수정표준계획의 결과가 비슷하여 표준계획의 것만을 표에 제시하였다), 우선 환자 자료에 근거한 면출규칙(data-dependent stopping rule)으로 계획을 종결하는 표준방법은 각 독성을 곡선에 따라 표본수가 다른데 대략 20명 이하에서 종결한다. 곡선 1의 경우 평균표본수는 SM3와 SM6의 각각에서 14.1명과 16.4명이며, 곡선 3의 경우 8.7명과 10.3명이며, 곡선 5의 경우 19.1명과 21.7명이다. 표 5의 면출규칙에 의한 MTD는 과거 논문(Goodman et al.(1995); Storer(1989))에서 밝혀진 대로 33%보다 낮은 용량을 추정하고 있으며, 일반적으로 제 1상 임상시험에서 20-33%를 겨냥하는 MTD를 바람직하다고 보는 기준에 의하면, 또한 O'Quigley 와 Chevret(1991)는 25%에 가까운 MTD가 적절하다는 기준에 의하면, SM3와 SM6 계획의 차이는 그리 크지 않다. 그러나 표준계획과 수정표준계획을 비교해 보면 우선 수정 SM6 계획에서 소요되는 평균환자수는 SM6 계획보다도 1명 정도 적은데도 수정표준계획은 표준계획보다 편의가 작아져 진보되었음을 본다. 표준방법에서 면출규칙의 단점은 실험의 종료 시점에서 무조건 바로 아래 용량을 MTD로 선택하므로 낮은 용량수준을 추정량으로 제시하는 빈도수가 크다는 점으로 따라서 모든 곡선에서 목적하는 MTD 용량수준보다 훨씬 낮은 용량수준을 MTD로 제시하는 빈도가 높은데, 수정표준계획은 이 낮은 용량수준의 빈도를 15% 이상 낮추고 있다. 구체적으로는 곡선 1에서 독성을 0.10 이하의 빈도가 SM3와 SM6 계획에서 48%와 51%인데 수정 SM3와 수정 SM6 계획에서 각각 30%와 32%이다. 곡선 3에서 독성을 0.22 미만(즉 용량수준 0)의 빈도가 SM3 와 SM6 계획에서 33%와 36%인데 수정 SM3와 수정 SM6 계획에서 각각 18%와 20%이다. 또한 곡선 5에서 독성을 0.09 이하의 빈도가 SM3와 SM6 계획에서 46%와 51%인데 수정 SM3와 수정 SM6 계획에서 각각 30%와 34%이다.

자료분석에 의한 MTD를 살펴보면, 특히 곡선 3에서 면출규칙에 의한 MTD보다 자료분석에 의한 MTD의 효능이 탁월하다. SM3에서 표준계획과 수정표준계획의 결과가 매우 비슷하여 표준계획의 것만을 표 5에 제시하였다. 여전히 표준계획에서도 MLE 추정량의 수렴문제가 존재한다. Storer(1989)가 보고한 Design D의 MLE 수렴율은 본 연구의 수렴율보다도 높은데 이는 본 연구에서는 임상시험의 현장에서 주로 사용하고 있는 6단계의 용량수준을 채택한 반면에, Storer(1989)의 모의실험에서는 본 논문에서보다 더욱 세밀한 0-50%의 독성을 사이에 9-10 단계의 용량수준을 채택하였기 때문이다. 다른 추정량들의 효능을 살펴보면 역시 DM 방법에서와 마찬가지로 여러 추정량, 즉 평균, 중앙값, 최빈값의 장, 단점이 각 곡선에 대해 달라서, 평균은 곡선 1과 3에서 좋으나 곡선 5에서는 다른 추정량보다 나쁘다. 세 곡선에 걸쳐 효능이 좋은 추정량은 중앙값이며, 또한 여러 추정량으로부터 구한 종합추정량 결과가 표 5의 마지막 열에 제시되었는데 이 종합추정량이 가장 바람직한 결과를 보이고 있다.

3.3 두 단계 BC와 BD 계획법

두 단계 계획인 BC와 BD 계획은 DM 방법 후에 C와 D 계획을 적용하는 방법으로서 이 C와 D 계획은 미리 정해진 표본수에서 멈추는 방법으로 Storer(1989)에 의해 제시되었다. 그러나 BC

와 BD 계획에 미리 정해진 표본수를 적용치 않고 만약 면출규칙을 적용시켜 MTD를 추정한다면 적은 환자수로 실험을 종료하면서도 효능을 유지할 수 있을 것으로 짐작된다. 이는 DM 방법으로 시작하여 면출규칙을 적용시킨 SM3 또는 SM6 계획을 두번째 단계로 진행하는 것과 동일하다.

두 단계 SM 계획의 모의실험 결과는 논문에 제시치 않았으나 간략히 설명하면 다음과 같다. 두 단계 SM 계획은 낮은 용량수준이 MTD로 채택되는 빈도가 줄어드는 장점이 있으나 참 MTD보다 더 높은 용량수준이 MTD로 채택되는 빈도가 현저히 높아 MTD 추정에 적절한 계획이 결코 아님을 시사하고 있으며, 이는 면출규칙에 의한 MTD나 자료분석에 의한 MTD 모두에서 동일하게 관측된 결과이었다.

IV. 결 론

MTD 추정은 오래 전부터 연구되어 왔으며 각 방법들에 대해서 많은 장단점이 제시되었으나 아직까지 어떤 방법이 가장 좋다는 결론을 내리지 못한 상태로 계속 연구되고 있다. 이와 같은 이유 중 하나는 주로 모의실험에 의존한 과거 여러 연구의 독성반응율 곡선, 시작용량, 용량단계가 서로 다르고, 여러 표준방법 중에서도 각 논문이 연구하기 위해 선택한 표준방법이 또한 다르기 때문에 즉각적인 비교가 어려워서였다. MTD 추정에 관한 연구에서 반드시 새로운 계획을 제시하는 것만이 최선책이 아니며 기존의 계획법에 대해서 엄밀한 검토로서 수정 방안을 찾는 것도 필요하다고 생각된다. 현재 가장 많이 사용되는 Up-and-Down 방법들을 개선할 수 있는 가능성을 본 연구의 모의실험으로 알아본 결과, 어떤 용량수준을 복용한 환자들의 독성반응율 관찰한 후 즉시 결정하는 면출규칙에 의한 MTD는 그 자체만으로는 불안정하며, 실험이 종결되기까지의 모든 환자 자료의 분석으로 뒷받침된다면 좀 더 유용한 MTD 추정량이 될 수 있음을 확인하였고, 또한 본 논문에서 제시한 두가지 수정방안으로 MTD 추정의 효능이 많이 진보되었음을 볼 수 있었다.

구체적으로 MED를 추정하는 DM 방법에서 면출규칙에 의해 MTD를 결정하는 것은 위험하며 적절한 표본수, 즉 25 또는 50에서 임상시험을 종결한 후 구하는 중앙값 MTD가 세 곡선에 걸쳐 효능이 가장 좋아 추천하는 MTD 추정량이다. Storer(1989)는 제 1상 임상시험의 계획에 랜덤화를 반영한다면 MTD 추정에 있어서 한층 진보된 계획이 될 것이라고 언급하였으나 본 연구에서 밝혀진 바로는 DM 방법에 랜덤화를 도입한 편의동전 계획이 좋은 결과를 보이지 않는다. 33%를 목표로 하는 표준계획에서도 MED와 마찬가지로 면출규칙에 의해 MTD를 결정하는 것은 위험하며 수정 표준계획으로 약간의 진보를 볼 수 있으나 역시 중앙값 MTD가 세 곡선에 걸쳐 효능이 좋으며, 면출규칙에 의한 MTD를 포함하여 평균, 중앙값, 최빈값 MTD로부터 구한 종합 MTD 추정량이 특히 곡선 5에서 중앙값 MTD 보다도 효능의 증가가 뚜렷하였다. 따라서 표준계획에서는 중앙값 MTD 또는 종합 MTD 추정량이 추천되는 MTD 추정량이다. MLE MTD는 비수렴하는 경우가 특히 표준계획에서 빈번하여 MTD 추정량으로서 적절치 않았다.

표준방법에서 사용하는 면출규칙의 효능을 알아보는 데에 있어 DM 방법의 모의실험 결과가 많은 도움이 되었는데, 그 이유는 표준방법은 면출규칙에 의해 실험을 종결할 때 매 임상시험의 시행 때마다 표본수가 변하므로 서로 다른 표본수, 다른 자료에 근거한 결과를 비교하는 데에는 한계가 있기 때문이다. 그러나 표준계획에서 여러 MTD 추정법은 면출규칙에 의해 종결된 자료를 분석하여 MTD 추정량을 계산하여 제시하므로 동일 표본수, 동일 자료에 근거한 결과의 비교이었다.

이제 본 연구에서 제시한 MTD의 수정 추정량을 CRM 방법의 추정량과 비교하는 연구가 장래

과제로 남아 있다.

참 고 문 헌

- [1] Brownlee, K. A., Hodges, J. L., Jr., and Rosenblatt, M. (1953). The up-and-down method with small samples. *Journal of the American Statistical Association*, Vol. 48, 262-277.
- [2] Chevret, S. (1993). The continual reassessment method in cancer phase I clinical trials: A simulation study. *Statistics in Medicine*, Vol. 12, 1093-1108.
- [3] Dixon, W. J. and Mood, A. M. (1948). A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, Vol. 43, 109-126.
- [4] Durham, S. D. and Flournoy, N. (1995). Up-and-down designs I: Stationary treatment distributions. In *Adaptive Designs*, Flournoy, N. and Rosenberger, W. F. (eds), 139-157. Hayward, California: Institute of Mathematical Statistics.
- [5] Durham, S. D., Flournoy, N. and Rosenberger, W. F. (1997). A random walk rule for phase I clinical trials. *Biometrics*, Vol. 53, 745-760.
- [6] Gatsonis, C. and Greenhouse, J. B. (1992). Bayesian method for phase I clinical trials. *Statistics in Medicine*, Vol. 11, 1377-1389.
- [7] Goodman, S. N., Zhurak, M. L. and Piantadosi, S. (1995). Some practical improvements in the continual reassessment method for phase I studies. *Statistics in Medicine*, Vol. 14, 1149-1161.
- [8] Hsi, J. L. (1968). The multiple-sample up-and-down method in bioassay. *Journal of the American Statistical Association*, Vol. 64, 147-162.
- [9] Korn, E. L., Midthune, D., Chen, T. T., Rubinstein, L. V., Christian, M. C. and Simon, R. M. (1994). A comparison of two phase I trial designs. *Statistics in Medicine* Vol. 13, 1799-1806.
- [10] Mick, R. and Ratain, M. J. (1993). Model-Guided determination of maximum tolerated dose in phase I clinical trials: Evidence for increased precision. *Journal of the National Cancer Institute*, Vol. 85, 217-223.
- [11] O'Quigley, J. (1992). Estimating the probability of toxicity at the recommended dose following a phase I clinical trial in cancer. *Biometrics*, Vol. 48, 853-862.
- [12] O'Quigley, J. and Chevret, S. (1991). Methods for dose finding studies in cancer clinical trials: A review and results of a monte carlo study. *Statistics in Medicine*, Vol. 10, 1647-1664.
- [13] O'Quigley, J. and Shen, L. Z. (1996). Continual reassessment method: A likelihood approach. *Biometrics*, Vol. 52, 673-684.
- [14] Piantadosi, S. and Liu, G. (1996). Improved designs for dose escalation studies using pharmacokinetic measurements. *Statistics in Medicine*, Vol. 15, 1605-1618.
- [15] Silvapulle, M. J. (1981). On the existence of maximum likelihood estimators for the

- binomial response models. *Journal of the Royal Statistical Society, Series B*, Vol. 43, 311–313.
- [16] Storer, B. E. (1989). Design and analysis of phase I clinical trials. *Biometrics*, Vol. 45, 925–937.
- [17] Wetherill, G. B. (1963). Sequential estimation of quantal response curves (with discussion). *Journal of the Royal Statistical Society, Series B*, Vol. 25, 1–48.

Table 2-1. Experimentation percentages and estimates of MTD for Dixon & Mood's Method to estimate the targeted probability $\theta = 0.5$

Table 2-2. Experimentation percentages and estimates of MTD for Dixon & Mood's Method to estimate the targeted probability $\theta = 0.5$

Table 2-3. Experimentation percentages and estimates of MTD for Dixon & Mood's Method to estimate the targeted probability $\theta = 0.5$

Table 3-1. Experimentation percentages and estimates of MTD for Biased Coin Design to estimate the targeted probability θ

θ	True probability of toxicity						Experimentation at $x_i(\%)$						Stopping at $x_i(\%)$						Recommendation at $x_i(\%)$						Observed toxicity (%)
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
Fixed Sample size n=10																									
0.10	71	25	4	0	0	0	54	37	8	1	0	0	52	38	9	1	0	0	7						
0.20	48	34	14	3	1	0	25	37	25	10	3	0	22	38	27	11	2	0	11						
0.25	40	33	19	7	1	0	17	34	28	16	4	1	16	33	29	15	6	1	13						
0.33	27	31	24	13	4	1	6	23	28	27	12	4	6	22	30	24	15	3	18						
0.50	12	17	21	24	16	10	0	7	0	52	9	32	1	0	26	5	55	13	31						
Fixed Sample size n=25																									
0.10	54	34	10	2	0	0	40	39	16	4	1	0	40	40	16	4	0	0	9						
0.20	29	34	24	10	3	0	17	32	30	16	4	1	16	34	27	17	5	1	16						
0.25	22	31	27	14	5	1	9	25	32	21	10	3	9	27	29	22	11	2	19						
0.33	13	22	26	22	13	4	4	15	24	32	17	8	4	15	27	27	21	6	26						
0.50	5	9	16	25	27	18	0	3	15	19	40	23	0	5	9	31	28	27	40						
Fixed Sample size n=50																									
0.10	47	37	13	3	0	0	39	41	17	3	0	0	41	41	16	2	0	0	10						
0.20	22	34	26	13	4	1	14	33	30	17	5	1	15	33	31	15	6	0	18						
0.25	15	29	29	18	7	2	9	28	32	20	9	2	8	29	30	22	8	3	22						
0.33	9	19	27	25	15	5	2	17	30	28	17	6	3	16	31	26	18	6	29						
0.50	3	6	14	26	30	21	0	3	10	28	35	24	0	2	13	26	36	23	43						

Table 3-2. Experimentation percentages and estimates of MTD for Biased Coin Design to estimate the targeted probability θ

θ	True probability of toxicity						Experimentation at $x_i(\%)$						Stopping at $x_i(\%)$						Recommendation at $x_i(\%)$						Observed toxicity (%)	
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6		
0.22 0.32 0.45 0.54 0.69 0.80																										
Fixed Sample size n=10																										
0.10	83	16	1	0	0	0	75	23	2	0	0	0	80	18	2	0	0	0	0	0	0	0	0	0	0	24
0.20	66	27	6	1	0	0	57	32	9	2	0	0	60	29	9	2	0	0	0	0	0	0	0	0	0	26
0.25	59	30	9	2	0	0	45	32	17	5	1	0	51	30	14	4	1	0	0	0	0	0	0	0	0	28
0.33	46	32	16	5	1	0	30	34	22	11	3	0	36	29	22	11	2	0	0	0	0	0	0	0	31	
0.50	22	30	22	17	7	2	0	40	0	50	2	8	17	0	53	2	26	2	0	0	0	0	0	0	40	
Fixed Sample size n=25																										
0.10	80	18	2	0	0	0	79	18	3	0	0	0	81	17	2	0	0	0	0	0	0	0	0	0	0	24
0.20	59	29	10	2	0	0	56	29	12	3	0	0	58	29	9	3	1	0	0	0	0	0	0	0	0	28
0.25	49	32	14	4	1	0	43	32	19	5	1	0	46	31	16	6	1	0	0	0	0	0	0	0	0	30
0.33	36	33	20	9	2	0	28	33	23	12	3	1	30	33	23	9	4	1	0	0	0	0	0	0	0	34
0.50	13	22	26	22	13	4	13	2	48	7	28	2	3	31	6	46	5	9	0	0	0	0	0	0	0	46
Fixed Sample size n=50																										
0.10	78	19	3	0	0	0	79	19	2	0	0	0	80	18	2	0	0	0	0	0	0	0	0	0	0	25
0.20	55	31	11	3	0	0	52	34	10	4	0	0	56	30	11	3	0	0	0	0	0	0	0	0	0	29
0.25	45	34	15	5	1	0	41	35	17	6	1	0	44	32	16	7	1	0	0	0	0	0	0	0	0	31
0.33	31	33	22	10	3	1	27	33	25	11	4	0	31	30	25	11	3	0	0	0	0	0	0	0	36	
0.50	10	20	25	24	15	6	3	28	12	40	9	8	10	7	40	13	25	5	0	0	0	0	0	0	48	

Table 3-3. Experimentation percentages and estimates of MTD for Biased Coin Design to estimate the targeted probability θ

θ	True probability of toxicity						Experimentation at $x_i(\%)$						Stopping at $x_i(\%)$						Recommendation at $x_i(\%)$						Observed toxicity (%)	
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6		
Fixed Sample size n=10																										
0.10	62	28	8	2	0	0	37	40	17	5	1	0	34	40	19	5	2	0	1	26	31	23	11	2	2	
0.20	38	33	19	8	2	0	9	28	31	22	9	1	7	26	31	23	11	2	0	29	33	20	7	4	4	
0.25	29	28	23	14	5	1	3	13	30	31	17	6	3	9	28	33	20	7	0	3	11	30	36	20	8	
0.33	20	22	22	19	12	5	0	4	15	31	33	17	0	3	11	30	36	20	0	2	6	41	51	51	21	
0.50	10	11	12	16	24	27	0	0	0	13	29	58	0	0	0	1	9	35	0	0	1	9	36	54	33	
Fixed Sample size n=25																										
0.10	37	31	19	10	3	0	10	27	32	23	7	1	8	27	33	23	8	1	3	4	17	34	33	12	9	
0.20	17	19	22	23	15	4	0	4	18	37	31	10	0	2	9	32	40	17	0	2	9	31	38	20	13	
0.25	13	15	18	24	21	9	0	2	9	32	40	17	0	0	5	22	44	29	0	0	5	22	40	33	19	
0.33	8	10	13	22	28	19	0	0	5	22	44	29	0	0	0	10	37	53	0	0	1	9	35	55	30	
0.50	4	4	6	12	31	43	0	0	0	10	37	53	0	0	0	1	10	34	55	0	0	1	9	36	54	
Fixed Sample size n=50																										
0.10	19	25	27	21	7	1	1	14	33	35	14	3	1	14	32	36	14	3	5	0	2	14	39	33	12	14
0.20	9	10	18	30	24	9	0	2	14	39	33	12	0	0	0	5	21	42	32	0	0	8	29	43	20	18
0.25	6	8	14	27	30	15	0	1	9	29	39	22	0	0	0	5	21	42	32	0	0	5	20	43	32	24
0.33	4	5	8	22	35	26	0	0	5	21	42	32	0	0	0	1	10	34	55	0	0	1	9	36	54	
0.50	2	2	3	11	33	49	0	0	1	10	34	55	0	0	0	1	9	36	54	0	0	1	9	36	54	

Table 4. Experimentation percentages of Standard and Modified Standard Methods to estimate the targeted probability $\theta = 0.33$

Dose level	True probability of toxicity	SM3	SM6	Modified SM6
Experimentation at x_i (%)				
1	0.05	24	23	23
2	0.10	26	27	26
3	0.25	27	27	27
4	0.35	16	16	17
5	0.50	6	6	6
6	0.70	1	1	1
Observed toxicity(%)		20(19.85)	20(19.9)	20(20.15)
Average number of cohorts (sample size)		4.7 (14.1)	5.5 (16.4)	5.2 (15.5)
Experimentation(%) <10, >40		24, 7	23, 7	23, 7
Experimentation at x_i (%)				
1	0.22	48	49	48
2	0.32	34	33	34
3	0.45	14	14	14
4	0.54	3	3	3
5	0.69	1	1	1
6	0.80			
Observed toxicity(%)		30(30.05)	30(39.95)	30(30.05)
Average number of cohorts (sample size)		2.9 (8.7)	3.4 (10.3)	3.3 (9.9)
Experimentation(%) <10, >40		0, 18	0, 18	0, 18
Experimentation at x_i (%)				
1	0.00	16	14	14
2	0.01	16	15	15
3	0.04	17	16	16
4	0.09	19	21	20
5	0.24	20	23	23
6	0.49	12	11	12
Observed toxicity(%)		13(13.23)	14(13.59)	14(13.99)
Average number of cohorts (sample size)		6.4 (19.1)	7.2 (21.7)	7.0 (20.9)
Experimentation(%) <10, >40		68, 12	66, 11	65, 12

Table 5. Simulated comparison of MTD estimates for Standard and Modified Standard Methods to estimate the targeted probability $\theta = 0.33$