비연속변수를 종속변수로 하는 비교 임상시험시 통계적 검정력 산출[†]

서울대학교 자연과학대학 통계학과, 서울대학교 의과대학 예방의학교실¹

곽민정, 이영조, 박병주¹

= Abstract =

Power Analysis for Comparative Controlled Clinical Trial with Non-continuous Outcome Variables

Minjung Kwak, Young-Jo Lee, Byung-Joo Park¹

*Department of Statistics, Seoul National University College of Natural Science,
Department of Preventive Medicine, Seoul National University College of Medicine¹*

Background : Calculating appropriate sample size is indispensable process in planning clinical trials. If the result of a comparative clinical trial cannot reject null hypothesis, four possibilities might be considered. Three of those can be the truly no difference of therapeutic effects between the two drugs, poor compliance of the patients to the prescribed drugs, and the effect of bias during the study. The last possibility can be the falsely accepting the null hypothesis due to the lack of statistical power. This paper reported power analysis with non-continuous outcome variables in comparative controlled clinical trials to point out latter problem.

Methods : A clinical trial of new anti-cancer agent using two groups was simulated with binary outcome variable. Another clinical trial was simulated to evaluate two treatment for lung cancer patients. After double-blind randomization to radiotherapy alone group and chemotherapy parallel group, follow-up observations were conducted. Outcome variable is survival time. By using PASS program, the statistical power and least significant number of sample size were calculated in the two different situation.

Results : The statistical power of anti-cancer agent data was 29%, which meant that sample size need to be increased. The least significant sample size for detecting the therapeutic difference between groups as statistically significant requires 200 persons. The statistical power of lung-cancer data was 39%, which also meant that sample size need to be increased. The least significant sample size for detecting the therapeutic difference between groups as statistically significant requires 84 persons.

Conclusion : Inadequate sample size in clinical trials can cause wrong decision due to decrease in statistical power, so proper power analysis is needed. The program introduced in this paper can be used in design, interim analysis, and final analysis of clinical trials to calculate statistical power very usefully.

KEY WORDS : Statistical power · Sample size · Clinical trial · Non-continuous outcome variable.

[†]본 연구는 1998년도 서울대학교병원 임상연구비(01-98-061)의 지원으로 이루어졌음.

교신저자 : 박병주

소 속 : 서울대학교 의과대학 예방의학교실

주 소 : 서울특별시 종로구 연건동 28 ☎ 110-799

전 화 : 02-740-8325 FAX : 02-747-4830 E-mail : bjpark@snu.ac.kr

서 론

두 약물간의 치료효능을 비교평가하기 위하여 수행한 임상시험의 결과가 두 약물간의 '치료효과에 유의한 차이가 없다'고 나타났을 경우에 연구자는 다음 네 가지 상황을 고려할 수 있다. 첫째는 실제로 치료효과 차이가 없어서 귀무가설을 참으로 받아들인 경우이며, 둘째는 연구대상자들이 처방된 약물에 대한 순응도가 낮아서 제대로 치료효과를 나타내지 못한 경우이며, 셋째는 연구수행상의 문제로 정확한 정보를 수집하지 못하여 바이어스가 개입된 경우를 생각할 수 있다. 마지막으로 통계적 검정력이 낮아서 실제로는 치료효능에 차이가 있는데도 불구하고 이를 입증하지 못하고 잘못된 결론을 내린 경우이다. 특히 연구대상수를 충분히 확보하기 어려운 현실적인 여건 때문에 국내에서 수행되어온 많은 임상시험연구들이 적은 수의 연구대상으로 종료되고 있어 제2종 오류를 범할 가능성이 높다고 판단된다. 이러한 경우에 연구결론을 내리기 전에 통계적 검정력 분석(power analysis)을 시행해봄으로써 타당한 결론을 얻을 수 있다.

그러므로 신약의 효능을 평가하기 위한 임상시험을 계획하는 단계에서 반드시 필요한 작업 중 하나는 연구가설의 검정에 적합한 연구대상수를 통계적으로 산출해내는 것이다. 이와 같은 통계적 근거에 의거하지 않고 임의로 연구대상수를 정하는 경우에 연구수행 및 결과해석에 중대한 문제가 생길 수 있다. 연구대상수를 불필요하게 많이 잡으면, 연구수행에 필요한 시간과 재정의 낭비를 초래할 뿐 아니라, 임상적으로 의미가 없는 사소한 치료효과 차이를 통계적으로 유의하게 만들므로써 효능이 낮아 임상적으로 사용할 가치가 없는 약물을 신약으로 승인받게 만든다. 이는 효능이 높고 안전한 약물을 신속하게 개발한다는 신약개발과정의 특성상 바람직하지 않으며 안전성이 완전히 입증되지 않은 신약을 불필요하게 많은 수의 대상자에게 투여하게 되므로 윤리적으로도 타당하지 않다. 반면에 연구대상수를 지나치게 적게 잡으면, 통계적 검정력이 떨어지게 되어 제2종 오류가 증가하므로써 실제로 약물의 효능이 임상적으로 유의함에도 불구하고 이를 입

증할 수 없는 결과를 초래하게 된다. 실제로 지금까지 시행되어 온 임상시험 중 많은 연구들이 불충분한 통계적 검정력으로 인하여 약물 혹은 치료의 효능을 올바르게 파악해내지 못하였다는 주장도 있다¹⁾.

이에 본 논문에서는 약물의 치료효능을 측정한 결과 변수가 비연속변수일 경우의 통계적 검정력 계산을 위하여 사용할 수 있는 프로그램과 실제 적용 결과를 제시할 것인데, 이는 결과변수가 연속변수일 경우의 통계적 검정력 산출²⁾에 대한 보완으로, 실제 임상시험의 계획에 있어서 표본 수 산출이나 검정력 계산 등의 문제를 살펴보면 결과변수가 이분적인 변수나 생존 시간을 나타내는 변수인 경우가 매우 많은데, 이런 비연속 변수의 경우에 대한 통계적 검정력을 고찰한 국내 논문이 부족하다고 생각하여 가상의 임상시험 자료를 이용하여 비연속 변수일 경우에 대한 검정력 계산을 하여 보았다.

연구대상 및 방법

1. 이론적 배경

임상시험의 수행에 있어서 알맞은 표본 수 계산, 최적의 추적조사 기간 산출 등과 함께 정확한 통계적 검정력의 계산은 통계학자들에게 자주 묻는 질문들 중의 하나이다. 통계적 검정력은 임상시험에서 관심의 대상이 되는 어떤 특성치 -예를 들면, 두 치료군 간의 혈압의 차이나 생존시간의 차이 등-의 정밀도를 바탕으로 계산되며 이는 표본 수와 제1종의 오류, 제2종의 오류의 함수로 나타내어진다. 임상시험에서는 연구 계획 수립 단계에서 관심이 되는 특정 가설을 세우고 연구 종료 후 그 가설의 검정에 관심이 있는데, 통계적 검정력은 그 가설의 검정에 쓰이는 검정통계량을 바탕으로 검정력 방정식을 세우고, 그 방정식의 해를 컴퓨터를 이용하여 찾게 된다.

통계적 검정력(statistical power)이란 귀무가설이 거짓인 상황에서 연구수행 결과로 귀무가설을 기각할 수 있는 확률, 즉 실제로 치료효과 차이가 있을 때 이를 차이가 있다고 판정할 수 있는 능력을 의미한다.

따라서 통계적 검정력이 낮은 경우에는 실제로는 비교하고자 하는 두 가지 약물간의 치료효과에 유의한 차이가 있음에도 불구하고 이를 차이가 없는 것으로 판단하는 오류를 범하게 된다. 참고로 제2종 오류는 실제 귀무가설이 거짓임에도 이를 채택할 확률이며, 따라서 $1-\beta$ 는 통계적 검정력과 같은 의미를 가진다. 반대로 제1종 오류는 실제 치료효과의 차이가 없는 상황에서 실제 연구를 수행한 결과로 귀무가설을 기각할 확률, 즉 효과의 차이가 있는 것으로 잘못 판단하게 될 확률을 의미한다. 간단히 말하면 통계적 검정력은 실제로 두 치료군간에 차이가 있을 때, 그 차이를 얼마나 잘 감지해내느냐의 정도라고 해석할 수 있다.

통계적 검정력의 계산은 임상시험의 연구 형태, 관심이 되는 가설, 특성치의 분포 등을 고려하여 여러 가지 방정식을 세우게 된다. 본 논문에서는 기존에 발표된 연속 변수일 경우에 대한 검정력 계산²⁾을 확장하여 비연속 변수에 대한 통계적 검정력을 계산한다. 첫 번째로 결과 변수가 범주형 변수인 경우 χ^2 분포를 이용한 검정력 방정식은 다음과 같다^{3,4)}. 비교 대조군

과 시험 처치군이 있어서 두 군간의 분율의 차이에 관심이 있다고 하자. 비교 대조군을 1로 했을 때, 시험 처치군의 배정비를 r 이라고 한다. n_2 는 비교 대조군의 필요 대상 수이고, rn_2 는 시험 처치군의 필요 대상 수이다. 비교 대조군의 분율을 π_1 , 시험 처치군의 분율을 π_2 라하고, 두 분율의 가중 평균을 $\bar{\pi}$, 알아내고자 하는 분율의 차이를 Δ 라하면 검정력 방정식은 다음과 같다. 단, Z_α 와 Z_β 는 각각 표준정규분포에서의 α 분위수, β 분위수를 나타낸다.

$$n_2 = \frac{(Z_\alpha \sqrt{(\gamma+1)\pi(1-\pi)/\gamma} + Z_\beta \sqrt{\pi_1(1+\pi_1) + \pi_2(1-\pi_2)})^2}{\Delta^2}$$

여기서 전체 필요대상자 수는 두 군의 대상자 수를 합친 $(1+r)n_2$ 가 된다.

두 번째로 결과 변수가 사람의 생존시간인 경우 생존시간이 지수 분포를 따른다고 가정하고 세운 검정력 방정식은 다음과 같다⁵⁾. 비교 대조군과 시험 처치군이 있어 두 군의 생존시간을 분석할 때, 여기서는 두

Fig. 1. Window frame of PASS program.

Table 1. Simulated data for analysing anti-cancer agent with binary outcome variable with N = 50

Control anti-cancer agent		New anti-cancer agent	
Patient No.	Remission Achievement	Patient No.	Remission Achievement
A-01	Fail	A-02	Success
A-03	Success	A-04	Success
A-05	Fail	A-07	Success
A-06	Fail	A-10	Fail
A-08	Success	A-11	Success
A-09	Fail	A-12	Fail
B-02	Success	A-13	Success
B-03	Success	B-01	Fail
B-07	Fail	B-04	Fail
B-08	Success	B-05	Success
B-09	Fail	B-06	Success
B-12	Success	B-10	Fail
B-13	Fail	B-11	Fail
C-04	Success	B-14	Success
C-05	Fail	C-01	Fail
C-06	Fail	C-02	Success
C-08	Fail	C-03	Success
C-10	Success	C-07	Fail
C-11	Fail	C-09	Success
C-12	Fail	C-13	Fail
C-14	Success	C-15	Success
C-16	Fail	D-01	Success
D-02	Fail	D-03	Success
D-04	Success	D-05	Fail
		D-06	Success
		D-07	Success

군의 위험율의 비 Δ 를 가지고 검정력 방정식을 세운다. d_1 은 비교 대조군에서 사건이 일어나는 대상자 수이고, d_2 는 시험 처치군에서 사건이 일어나는 대상자 수이다.

$$\frac{1}{d} + \frac{1}{d} = \frac{[\log(\Delta)]^2}{(Z_\alpha + Z_\beta)^2}$$

2. 통계적 검정력 계산을 위한 프로그램

본 논문에서는 가상적인 임상시험 자료를 이용하여,

분율의 분석과 생존시간의 분석에 대한 검정력을 계산하여 보았다. 이용된 검정력 계산 프로그램은 Power and Sample Size(PASS) 2000 Release로서, 여러 가지 경우의 검정력 계산과 표본 수 계산뿐만 아니라 검정력 곡선을 그림으로 나타내어 주므로 연구대상수나 치료 효과 크기의 차이가 변함에 따라 검정력이 어떻게 변하는 지를 가시적으로 이해할 수 있는 장점이 있다. 프로그램은 NCSS 홈페이지 <http://www.ncss.com>에서 얻을 수 있다. 이 프로그램은 윈도우 방식으로 되어 있어 사용자가 이용하기에 매우 편리하다. 그

Table 2. Simulated data for analysing lung-cancer patients with survival time variable with N=30

Radiotherapy Only			Radiotherapy with Chemotherapy		
Follow-up Start	Follow-up Finish	Result	Follow-up Start	Follow-up Finish	Result
99. 12. 28	00. 3. 28	Dead	98. 5. 9	99. 6. 29	Unknown
99. 2. 1	01. 1. 1	Alive	00. 3. 11	01. 1. 1	Alive
98. 6. 22	98. 12. 15	Dead	00. 2. 26	01. 1. 1	Alive
98. 7. 5	99. 1. 15	Unknown	00. 2. 6	01. 1. 1	Alive
98. 12. 22	99. 7. 14	Dead	98. 3. 28	98. 8. 17	Dead
99. 11. 21	00. 3. 18	Dead	00. 1. 29	01. 1. 1	Alive
99. 9. 18	00. 12. 17	Dead	00. 1. 15	01. 1. 1	Alive
99. 6. 15	01. 1. 1	Alive	99. 12. 31	01. 1. 1	Alive
99. 6. 13	00. 4. 26	Unknown	99. 5. 28	00. 3. 23	Dead
99. 1. 15	00. 6. 8	Dead	99. 7. 31	01. 1. 1	Alive
98. 6. 29	01. 1. 1	Alive	98. 8. 18	01. 1. 1	Alive
98. 4. 15	98. 12. 25	Dead	98. 1. 30	00. 3. 1	Dead
98. 7. 29	01. 1. 1	Alive	98. 10. 26	00. 4. 19	Dead
98. 1. 15	98. 11. 30	Unknown	99. 5. 16	00. 4. 17	Dead
98. 4. 26	99. 6. 25	Dead	98. 2. 8	99. 10. 26	Unknown

림 1은 카이 제곱 검정을 이용한 통계적 검정력 계산의 예에서 사용된 해당 메뉴의 초기 화면이다.

카이 제곱 검정을 위한 통계적 검정력 계산에 대한 화면에서 입력항목은 구하고자 하는 것이 통계적 검정력임을 지정해주고, 자료의 수를 입력해야 하며, 제1종의 오류와 제2종의 오류를 지정해주어야 한다. 또, 자료로 적절한 테이블을 만들었을 때의 자유도와 치료 효과의 차이도 입력해 주어야 한다. 프로그램은 메뉴 방식으로 여러 가지 값을 마우스로 클릭함으로써 지정할 수 있지만 직접 원하는 값을 키보드로 입력할 수도 있다.

3. 통계적 검정력 분석 방법

두 가지의 가상적인 자료를 만들어 통계적 검정력을 분석하였다. 첫 번째로, 범주형 변수에 대한 분석으로

항암제의 치료 효과를 평가하기 위한 임상시험을 가정하고 표본의 수가 50명 일 때와, 200명 일 때, 각각의 표를 카이제곱 검정을 이용하여 분석해보고 분석 결과를 해석함에 있어서 표본의 수가 적절한 지를 통계적 검정력 계산을 통해 살펴보았다. 보통 제 3상 임상시험의 표본의 수는 200-300명 정도이나, 본 논문에서는 적절하지 못한 표본의 수가 통계적 검정력을 떨어뜨림으로써 유의하지 못한 분석결과를 가져올 수도 있다는 점을 지적하기 위하여 가장 적은 경우로 50명의 표본 수를 상정하여 보았고, 200명으로 하였을 때의 결과도 살펴보았으며, 표본의 수가 200명인 경우, 표본의 수가 50명인 경우와 각 칸의 숫자가 비례하도록 가상적인 자료를 만들었다. 표본의 수가 50명인 경우, 계산에 있어서 입력되는 값은 치료효과의 차이 0.2와 제1종의 오류 0.05, 자유도 1이다. 덧붙여 치료 효과의 차이와 자유도를 변화시켜 가면서 검

Table 3. 2×2 table for data for analysing anti-cancer agent with binary outcome variable with $N=50$

		Remission Achievement	
		Fail	Success
Treatment	Control anti-cancer agent	14	10
	New anti-cancer agent	10	16

Table 4. Numerical result using PASS program with data in Table 3.

N	50	70	90	110	130	150	170	190	197	210
Power	29%	39%	48%	55%	63%	69%	74%	79%	80%	83%

정력 곡선이 변화하는 양상을 파악하였다. 사용된 자료는 표 1과 같다.

두 번째로, 생존 자료에 대한 분석으로 30명의 폐암환자를 방사선 치료군과 화학요법 병행 군으로 무작위 배정하여 추적 조사한 자료⁹⁾(표 2)를 대상으로 통계적 검정력 분석을 실시하였다. 이 자료에 대한 입력값은 일정 기간후의 생존율과 유의수준, 검정의 종류, 전체 추적조사 기간 등인데, 500일까지의 생존율을 화학요법 병행군의 경우 0.7754, 방사선 치료군의 경우 0.4675로 입력하였다. 유의수준은 0.05로 하였으며 전체 추적조사 기간은 3년이고, 양측검정의 경우를 상정하여 계산하여 보았다. 덧붙여 유의수준과 검정의 종류를 변화시켜 가면서 검정력 곡선을 그렸다.

결 과

1. 종속 변수가 이분적인 2-by-2 table 에서의 통계적 검정력 분석

표 3은 새로운 항암제의 치료효과를 평가하기 위한 가상적인 임상시험의 결과를 2×2 표로 정리한 것이다. 표 3에 대한 Pearson's chi-square 통계량 값은 1.97 이고 자유도는 1이며, 유의확률은 0.159로서 유의수준 0.05에서 새로운 항암제가 기존의 항암제에

Fig. 2. Power curve using numerical result in Table 4.

비해 치료 효과가 더 좋다는 대립가설을 채택할 수 없다. 여기서 전체 표본의 수가 50명으로, 이 임상시험에서 원하는 결과를 얻기 위한 표본의 수가 부족한 것은 아닌지 알아보기 위하여, 통계적 검정력 분석을 실시하여 보았다. 분석 결과 50명의 환자를 대상으로 치료 효과의 차이 20%를 감지해 내기 위한 카이제곱 분석에서 자유도가 1이고, 유의수준이 5% 일 때, 통계적 검정력은 29%로 나타났다. 또, 통계적 검정력을 80%로 고정시켜 놓고 구한 최소 유효 표본 수는 197명이었다.

Fig. 3. Power curve with data in Table 3. varying effect size 0.1, 0.2, 0.3

카이제곱 검정을 이용하여 수치적으로 구한 전체 자료의 수 N과 통계적 검정력과의 관계를 표 4와 그림 2로 제시하였다. 그림 2에서 볼 수 있듯이 통계적 검정력 곡선은 연구대상수가 증가할수록 높아진다. 예를 들어, 연구대상수가 200명일 경우 약 80%의 통계적 검정력을 가지게 된다. 그림 3은 치료 효과의 차이가 0.1, 0.2, 0.3일 경우의 통계적 검정력 곡선과, 자유도가 1, 2, 3일 때의 통계적 검정력 곡선을 그려 비교하여 본 것이다. 감지해내고자 하는 치료 효과의 차이를 클수록, 자유도가 작을수록 통계적 검정력이 높게 나타남을 볼 수 있다.

전체 자료의 수를 200명으로 늘려서 가상적으로 만든 자료가 표 5와 같을 때, 이에 대한 Pearson's chi-square 통계량 값은 8.35 이고 자유도는 1이며, 유의확률은 0.0038 로서 유의수준 0.05에서 새로운 항

Fig. 4. Power curve with data in Table 3. varying degree of freedom 1, 2, and 3

암제가 기존의 항암제에 비하여 치료 효과가 더 좋다고 판단할 수 있었다.

2. 두 치료군의 생존 자료에 대한 통계적 검정력 분석

폐암환자를 1998년 1월 1일부터 2001년 1월 1일까지 방사선 단독치료군과 화학요법 병행군으로 무작위 배정하여 치료한 뒤 추적조사한 가상적인 연구결과 자료는 표 2와 같다. 표 2의 자료에 대해 로그순위 검정을 이용하여 두 생존곡선을 비교한 결과, 유의확률이 0.3247로 두 군간의 생존곡선이 다르다고 말할 수 없다. 여기서 전체 표본의 수가 30명으로, 이 임상시험에서 원하는 결과를 얻기 위한 표본의 수가

Table 5. 2×2 table for data for analysing anti-cancer agent with binary outcome variable with N=200

		Remission Achievement	
		Fail	Success
Treatment	Control		
	anti-cancer agent	52	39
	New		
	anti-cancer agent	40	69

Table 6. Numerical result using PASS program with data in Table 2.

N	30	50	70	84	90	110	130	150	170	190
Power	39%	58%	73%	80%	83%	89%	94%	96%	98%	98%

Fig. 5. Power curve using numerical result in Table 6.

부족한 것은 아닌지 알아보기 위하여, 통계적 검정력 분석을 실시하였다. 입력 값으로는 일정 기간후의 생존율의 경우, 500일 생존율이 화학요법 병행군은 0.7754, 방사선 치료군은 0.4675이었고, 전체 추적조사 기간은 3년이었다. 분석 결과 30명의 환자를 대상으로 로그순위 검정을 이용하여 두 치료군의 생존함수가 같은지를 검정한 경우, 유의수준이 5% 일 때, 통계적 검정력은 39%로 나타났다. 또, 통계적 검정력을 80%로 고정시키고 구한 최소 유효 표본수는 84명이었다.

로그순위 검정을 이용하여 수치적으로 구한 전체 연구대상수 N과 통계적 검정력과의 관계를 표 6과 그림 5로 나타내었다. 그림 5에서 볼 수 있듯이 통계적 검정력 곡선은 연구대상수가 증가할수록 커진다. 예를 들어, 연구대상수가 80명일 경우 약 80%의 통계적 검정력을 가지게 된다.

다음은 제1종의 오류를 범할 확률인 유의수준 α 를 0.01, 0.05, 및 0.1일 경우의 통계적 검정력 곡선을 함께 그려 비교하여 본 것이다. 역시 유의수준이 높은 경우에 통계적 검정력이 크게 나타남을 볼 수 있다.

Fig. 6. Power curve with data in Table 2. varying $\alpha=0.01, 0.05, 0.1$ in case of two-sided test

Fig. 7. Power curve with data in Table 2. varying $\alpha=0.01, 0.05, 0.1$ in case of one-sided test.

양측 검정과 단측 검정의 비교에서는 양측 검정의 경우 통계적 검정력이 더 작음을 볼 수 있다(그림 6, 그림 7).

고 찰

본 논문은 연속 변수에 대한 통계적 검정력 계산의 확장으로 비연속 변수에 대한 통계적 검정력을 계산하는 방법을 가상의 연구자료를 이용하여 적용함으로써 향후 실제 임상시험에서 많이 활용될 통계적 검정력 분석에 대한 이해를 보다 심화하고자 하였다. 분석 결과에서 여러 그림(그림 2-그림 7)을 통하여 알 수 있듯이, 통계적 검정력 곡선은 전체 연구대상수에 대한 증가 함수로 나타내어지며, 치료 효과의 차이가 클수록, 자유도가 작을수록 통계적 검정력이 증가하며, 단측 검정의 경우 양측 검정일 때보다 통계적 검정력이 더욱 크게 증가한다. 또, 제1종의 오류를 범할 확률을 크게 허용할수록 역시 통계적 검정력이 증가한다.

본 논문에서는 여러 가지 비연속 변수 중에서 이분적인 변수와 생존자료를 나타내는 변수인 경우를 대상으로 통계적 검정력을 산출하는 과정을 소개하였는데, 이 두 가지 형태의 변수는 비연속변수 중 가장 기본적인 형태로서 실제 임상연구자들에게 통계적 검정력에 대한 이해와 중요성을 강조하기 위하여 가상의 연구자료를 가지고 통계적 검정력을 계산하여 보았다. 실제 임상시험에서 얻어지는 자료의 형태는 매우 다양하므로 이 밖에 여러 가지 다른 형태의 변수에 대해서도 적절한 통계적 검정력 방정식을 세워 정확한 통계적 검정력과 필요한 최소 연구대상수를 산출할 수 있을 것이다. 또한 향후 특정 임상시험을 시행할 때, 사전에 잘 수립된 연구계획에 따라 연구가 수행된 것이라 하더라도 연구대상수가 부족할 경우에는 통계적 검정력이 떨어지게 되어 결과를 올바르게 판정할 수 없다. 반드시 이의 적절성 여부를 확인하는 과정이 필요할 텐데, 사용된 프로그램을 임상시험연구를 계획하는 단계에 활용하는 것은 물론, 연구수행 도중의 중간 분석시, 결과분석이 완료된 후의 확인과정에 활용함으로써 손쉽게 통계적 검정력을 산출할 수 있을 것이고, 결과적으로 효과적인 임상시험의 수행에 도움이 될 것이다.

임상시험의 설계 종류와 가설의 형태 등에 따라 대단히 많은 통계적 검정력 방정식이 존재하며⁷⁾, 이들이 컴퓨터 프로그램화 되므로써 사용하기에 매우 편리해졌다⁸⁾. 기존의 방정식이 전제로 하는 가설을 모두 만족하지 않는 임상시험의 경우, 가상의 임상시험을 수행하여 자료를 구한 다음 그 시험에서의 통계적 검정력을 계산하는 모의 실험적 방법이 가능하다⁹⁾. 본 논문에서는 비연속 변수의 가장 기본적인 예라 할 수 있는 범주형 변수와 생존 시간자료에 대해서 통계적 검정력을 계산하고 구하여 보았다. 이 외에도 오즈비(odds ratio)나 상대위험도(relative risk)를 추정하고자 할 때¹⁰⁻¹²⁾, 축차적인(sequential) 시험 설계의 경우¹³⁾, 결과 변수의 특성 값이 유계(bounded)인 경우¹⁴⁾에 대한 여러 가지 검정력 방정식이 존재한다. 일반적으로 생존분석에서는 각 관측치가 독립적이라는 가정을 하나, 독립성의 가정이 위배되는 경우에 통계적 검정력을 산출하는 방정식도 소개되어 있다¹⁵⁾.

요 약

연구배경 :

임상시험의 시행에 있어 통계적 검정력이 떨어지는 경우에는 실제 약물의 효과에 대한 타당한 결론을 내리기 어려우므로 통계적 검정력을 분석할 필요가 있다. 본 논문에서는 결과변수가 비연속변수일 경우의 가상의 임상시험 자료에 PASS 라는 컴퓨터 프로그램을 이용하여 간편하게 통계적 검정력을 산출하는 과정 및 그 결과를 제시하였다.

대상 및 방법 :

새로운 항암제의 치료효과를 평가하기 위한 가상의 비교임상시험과, 폐암환자를 방사선 단독 치료군과 화학요법 병행군으로 무작위 배정하여 치료한 뒤 추적조사를 실시한 가상의 자료를 이용하여 통계적 검정력과 최소 연구대상수를 산출하였다.

결 과 :

새로운 항암제의 치료효과를 평가한 자료를 분석한 결과 처치와 예후간에 통계적으로 유의한 연관성이 없

다는 결과를 얻었으나, 통계적 검정력 분석 결과 검정력이 약 29%로 낮은 것으로 나타나 연구대상수가 충분하지 않아서 나타난 결과임을 알 수 있었다. 그리고 30명의 폐암환자를 대상으로 한 가상적인 자료의 경우도 통계적 검정력 분석 결과 검정력이 약 39%로서 매우 낮은 것으로 나타나, 이 역시 부족한 연구대상수가 문제임을 알 수 있었다.

결 론 :

이와 같은 결과를 통하여 임상시험을 시행할 때 연구대상수가 부족한 경우에는 통계적 검정력이 떨어지게 되어 결과를 올바르게 판정해 낼 수 없으면 반드시 이의 적절성 여부를 확인하는 과정이 필요함을 확인하였다. 소개된 프로그램은 연구 계획시, 연구 도중의 중간분석시 및 결과분석이 완료된 이후의 각 단계에서 손쉽게 통계적 검정력을 산출하는데 유용하게 사용될 것이다.

참 고 문 헌

1. Frieman JA, Chalmers TC, Smith HJr et al. *The importance of beta, the type II error and sample size in design and interpretation of the randomized control trial:survey of "negative trial"*. *New Eng J of Med* 1978;290-294
2. 구혜원, 박병주, 배종면, 강종원, 손혜숙. 연속변수를 종속변수로 하는 비교 임상시험시 SAS Macro를 이용한 통계적 검정력 산출. *임상약리학 회지* 1997;5:212-224
3. Pocock SJ. *Statistical aspects of clinical trial design*. *The Statistician* 1982;31:1-18
4. 신영수, 안윤옥. *의학연구 방법론*. 서울대학교 출판부. p.191-192, 1998
5. Piantadosi S. *Clinical trials; A methodologic perspective*. John Wiley & Sons, Inc, New York. p.148-181, 1997
6. 유근영, 박병주, 김 현, 이무송. 의, 약, 보건학을 위한 PC-SAS. p.133-153, 1995
7. Lerman J. *Study design in clinical research: sample size estimation and power analysis*. *Can J Anaesth* 1996;43(2):184-191
8. Brown BW, Lovato J, Russell K. *Asymptotic power calculations: description, examples, computer code*. *Stat Med* 1999;18(22):3137-3151
9. Liu G, Liang KY. *Sample size calculations for studies with correlated observations*. *Biometrics* 1997;53(3):937-947
10. Ben Armstrong. *A simple estimator of minimum detectable relative risk, sample size, or power in cohort studies*. *Am J Epidemiol* 1987;126(2):356-358
11. Glen AS, Lawrence LK. *Sample size requirements for interval estimation of the odds ratio*. *Am J Epidemiol* 1990;131(1):177-184
12. Jay HL, Mitchell HG. *On power and sample size for studying features of the relative odds of disease*. *Am J Epidemiol* 1990;131(3):552-566
13. Rebecca A. Betensky. *Conditional power calculations for early acceptance of H_0 embedded in sequential tests*. *Stat Med* 1997;16:465-477
14. Emmanuel L, Ilse S, Jurgen F, Erich B. *Calculation of power and sample size with bounded outcome scores*. *Stat Med* 1993;12:1063-1078
15. Lee SY. *Power calculation for a score test in the dependent censoring model*. *Stat Med* 1996;15:1049-1058