연구대상자 수 결정

천 병 렬*

= Abstract =

Sample Size Determination

Byoung Ryul Chun

Department of Preventive Medicine, Kyungpook University College of Medicine, Taegu, Korea

서 론

연구자가 연구를 시작하기로 결정한 후에는 먼저 연구대상자 수를 결정해야 한다. 임상시험의 경우 연구가 일단 시작되고 난 후에는 연구대상자 수를 조절하기 대단히 어렵다. 만약, 연구대상자 수(표본 수)가 너무 적다면 아무리 연구를 엄격히 잘 시행하여도 연구에서 제시된 가설에 대한 통계적으로 의미 있는 결과를 도출할 수 없다. 반면에 연구대상자 수가 너무 많다면 필요이상의 시간과 경비가 소모되어 바람직하지 못하다. 따라서 주어진 연구목적에 달성하기 위해 가장 적절한 연구대상자 수를 결정짓는 것은 대단히 중요한 일이며 동시에 가장 어려운 일이다.

연구대상자 수를 결정짓는 방법은 크게 나누어 연구시작 전에 미리 고정된 수를 할당시켜 연구를 시행하는 것(Fixed sample size design)과 연구를 시행해가면서 정해진 결과의 차이에 대한 해답(효과에 대한 차이가 있다는 결과에 도달하거나 아니면 차이가 없다는 결과에 도달함)이 나올 때까지 계속 연구를 진행하는 방법(Sequential sample size design)이 있다. 여기서는 임상시험의 Fixed sample size design의 연구대상자 수를 결정하는 방법을 알아본다.

임상시험 대상자 수의 계산을 위해 필요한 사전 정보

1. 연구의 목적이 구체적이고 간결하게 명시되어야 한다.

귀무가설(Null hypothesis)과 대립가설(Alternative hypothesis)이 설정되어야 하고 단측검정(One-tailed test)인지 양측검정(Two-tailed test)인지를 결정하여야 한다. 일반적으로 단측검정은 신약이 기존의 약제보다 반드시 효과가 좋을 것이라는 확신이 있을 때라야 적용이 가능하다. 단측검정은 양측검정에 비해서 연구대상자수가 적어도 된다. 그러나 동일한 연구대상자 수일 때는 검증력(Power of test)는 떨어진다.

2. 제 1 종 오류(Type I error= α error)와 제 2 종 오류(Type II error= β error)를 결정하여야 한다.

제 1 종 오류란 시험군(신약 투여군)과 비교대조군(기존의 약제 투여군)의 결과를 비교하여 치료효과의 차이가 있는지의 여부를 판정할 때, 실제로는 차이가 없음(귀무가설이 참임)에도 불구하고 연구자는 차이가 있다고 판정할 오류이다. 일반적으로 α error는 0.05로 정한다. 제 2 종 오류는 제 1 종의 오류와는 반대되는 개념으로 시험군과 비교대조군의 치료효과의 차이를 판정할 때, 실제로는 차이가 있음(귀무가설이 거짓임)에도 불구하고 연구자는 귀무가설을 받아들여 차이가 없다고 판정할 오류이다. 따라서 1- β 를 차

*경북대학교 의과대학 예방의학교실
교신저자: 천병렬

이가 있는 것을 있다고 판정하는 힘(power of test)이라고 한다. 일반적으로 β error는 α error의 4배 이내의 값(0.1~0.2)을 기준으로 한다.

연구가 끝나면 연구자는 귀무가설의 기각을 위한 통계적검증을 실시한다. 모집단에서 귀무가설의 참, 거짓여부에 따라 임상시험의 결과해석의 오류여부가 결정된다. 즉, 표본과 모집단의 결과가 일치하면 오류가 없지만, 일치하지 않을 때는 오류가 발생한다. 표 1을 참고로 보면 두 종류의 오류를 완전히 제거하고 싶을 것이지만 그렇게 할 수 없고 가능한 작게할 수 있다. 그런데 오류를 줄이려면 연구대상자의 수를 많게 하면 된다. 그러므로 연구대상자수를 결정할때 제 1종 및 제 2종의 오류를 허용할 수 있는 정도까지 낮추면서 불필요하게 어렵거나 비용이 많이 들지 않는 수준까지 연구대상자의 수를 조절하는 것이 중요하다.

3. 알아내고자 하는 차이의 정도

(effect size= Δ)

모집단에서 신약이 기존의 약제보다 치료효과가 어느 정도 나은지를 밝혀내려면 표본에서 실제 어느 정도의 차이가 있으면 이런 결론을 내릴 수 있는지에 대한 기준이 있어야 한다. 이 차이는 임상시험을 통해서 연구자가 알아내고자 하는 차이의 최소치로서 이를 effect size라고 한다. 이는 주로 연구자가 사전 연구(pilot study)나 다른 연구결과를 참고로 하여 정할 수 있다. 그러나 만약, 이런 자료가 없을 때는 임상적으로 의미있는(clinically significant) 가장 작은 effect size를 선택하는 것이 타당하다. 결국 effect size의 선택은 임의적이며 실현가능성(feasibility)이

중요한 결정요인이 된다. Effect size의 크기가 클수록 연구대상자의 크기는 작아진다.

4. 연구결과에 적용시킬 통계검정 공식(Type of statistical test)

연구대상자의 수를 결정하기 위한 공식은 수학적인 가정(assumption)에 근거를 둔다. 따라서 연구대상자 수를 결정짓기 전에 자료의 분석을 위한 통계적 기법의 적용에 대해 먼저 결정해 두어야 한다. 이것이 결정되면 이 공식을 이용하여 연구대상자의 수를 계산한다. 이때 통계적기법의 적용을 결정짓는 기준은 연구방법, 목적 그리고 독립변수와 종속변수의 측정 척도 형태에 따라 달라진다.

5. 그외에 고려할 사항들

1) 시험군과 비교대조군의 설정비율

$$N = rN_t + N_c$$

N : 전체 연구대상자수

r : 시험 치료군의 수

N_t : 시험군의 연구대상자 크기

N_c : 비교대조군의 연구대상자 크기

(1) 할당비가 동일할 때(Uniform allocation scheme) : 일반적으로 개개의 시험군과 비교대조군의 규모는 동일하게 하는 것이 원칙이다.

(2) 할당비가 다를 때(Nonuniform allocation scheme) : 비교를 더 정확하게 하거나 2차적인 목적을 만족시키기 위해 어떤 특정 치료군에 더 많은 연구대상자를 할당시키는 것

표 1. 제 1종의 오류와 제 2종의 오류의 정의

		모집단의 진실	
		신약과 기존의 약제는 치료효과가 다르다. (귀무가설이 거짓인 경우)	신약과 기존의 약제는 치료효과가 동일하다. (귀무가설이 참인 경우)
표 본 에 서 얻 어 진 결 과	귀무가설 기각시킴	참	Type I error 제 1종의 오류 (α error)
	귀무가설 기각에 실패함	Type II error 제 2종의 오류 (β error)	참

2) Variability of size(분산의 중요성)

연구대상자의 수를 결정하는 데 effect size 뿐만 아니라 그 variability도 상당히 중요하다. 변수가 dichotomous variable일 때는 variability가 이미 연구대상자수를 계산하는 공식 속에 반영되지만, 변수가 continuous variable 일 때는 variability를 고려해야 한다.

3) 추적연구설계에서의 탈락정도(drop out)

연구대상자들이 연구기간 도중에 탈락할 것을 고려하여 연구대상자 수를 계산해야 한다. 그 방법은 예상탈락율을 추정하여 산출된 연구대상자 수에 보정지수를 곱해준다.

보정지수 = $1/(1-d)$, d : 예상탈락율

예를 들면 산출된 연구대상자 수가 1,000명인데 예상탈락율이 10%라면 실제 필요한 연구대상자 수 = $1000 \times (1/(1-0.1)) = 1,110$ 명이 된다.

4) 치료거부(Treatment noncompliance)

할당된 치료법에 대한 연구대상 환자들이 치료받기를 거부하거나 연구자들이 약의 투여를 거부하는 경우 또는 환자나 연구자들이 치료약의 전 용량을 투여받거나 투여하기를 거부하는 경우에도 3)의 경우와 같은 공식에 의해 보정하면 된다.

5) 추적관찰기간

추적기간을 정해줄 것, 일반적으로 결과 사건의 누적발생률이 증가하고 추적기간이 길수록 필요한 연구대상자 수는 감소한다.

6) 혼란변수(confounding variable)

자료의 분석시 연구요인과 결과간의 관련성을 정확히 평가하기 위해 혼란변수의 영향을 조정하려면 연구대상자의 크기가 더 증가해야 한다. 더 요구되는 연구대상자의 수는 다음과 같은 세가지 요소에 의해 결정된다.

(1) 혼란변수의 수

(2) 연구요인과 혼란변수와의 관련성의 강도

(3) 결과변수와 혼란변수의 관련성의 강도

그러나, 이들 요소들의 복잡한 영향을 정확히 평가하기는 어려우므로 대개는 중요한 혼란변수 한가 지마다 연구대상자의 수를 약 10~20%씩 증가시키

도록 정하고 있다.

7) Multiple hypothesis

귀무가설의 수가 여러가지일 때는 개개의 귀무가설마다 유의수준을 0.05로 한다면 귀무가설의 수를 고려하여 유의수준을 재조정 해야한다. 예를 들면, 4 가지 가설을 한 연구에서 검증하려면, 유의수준은 $0.05/4 = 0.0125$ 로 하여야하므로 실제 연구대상자 수는 증가한다. 그러나, 이런 기준은 너무 엄격하므로 검증해야 할 가설들이 서로 완전히 독립적이지 않을 때만 적용하도록 한다.

8) Treatment lag time

치료가 시행되면 그 효과가 즉각 나타날 때를 no treatment lag이라고 한다. 따라서 연구대상자 수를 정할 때 치료효과가 나타날 시간적인 lag을 고려해야한다. 그러나, 실질적으로 이것을 고려한 연구대상자 수의 계산은 어렵다.

임상시험의 연구대상자 수를 계산하기 위한 공식

임상시험의 수행에 필요한 가장 적합한 연구대상자수의 계산을 하기 위해서는 표 2의 요약된 내용을 기준으로 하여 알맞은 공식을 찾아내어 일차적인 연구대상자수를 계산한다. 그 다음 여러가지 정보를 고려하여 최종적인 연구대상자 수를 결정짓는다.

공식 1 : Fisher's exact method

Fisher's exact test에 의한 연구대상자 수의 계산은 Gail and Gart, Haseman, Casagrand등이 computer program으로 계산한 표를 이용하면 된다.

공식 2 : Chi-square approximation

단순배정일 때

(Uniform allocation, $N_t = N_c$)

$$N_c = \frac{(Z\alpha\sqrt{2PQ} + Z\beta\sqrt{P_cQ_c + P_tQ_t})^2}{\Delta^2}$$

공식 3 : Chi-square approximation

이형배정일 때

(Nonuniform allocation, $N_t = RN_c$)

$$N_c = \frac{(Z\alpha\sqrt{PQ(R+1)/R} + Z\beta\sqrt{P_cQ_c + P_tQ_t/R})^2}{\Delta^2}$$

N_c : 비교대조군의 필요대상수

N_t : 시험군의 필요대상수

r : 시험군의 수

R : 배정비(비교대조군을 1로 한 시험군의 배정비)

P_c : 비교대조군에서 예상되는 결과(치료효과 : 치료율, 사망율)

P_t : 시험군에서 예상되는 결과(치료효과 : 치료율, 사망율)

Δ : $P_c - P_t$ Q_c : $1 - P_c$

Q_t : $1 - P_t$

P : P_c 와 P_t 의 가중평균치 $R=1$ 이면 $(P_c + P_t)/2$,

$R \neq 1$ 이면 $(P_c + RP_t)/(1 + R)$

Q : $1 - P$

Z_α : α -error에 해당되는 Z 값, 단측검증(One-tailed test) 적용

Z_β : β -error에 해당되는 Z 값

N : 전체 필요한 연구대상자 수

공식 4 : Inverse sine transform approximation

단순배정일 때

(Uniform allocation, $N_t = N_c$)

$$N_c = \frac{(Z\alpha + Z\beta)^2}{2(\sin^{-1}\sqrt{P_c} - \sin^{-1}\sqrt{P_t})^2}$$

공식 5 : Inverse sine transform approximation

이형배정일 때

(Nonuniform allocation, $N_t = RN_c$)

$$N_c = \frac{(Z\alpha + Z\beta)^2(R+1)/R}{4(\sin^{-1}\sqrt{P_c} - \sin^{-1}\sqrt{P_t})^2}$$

공식 6 : Poisson approximation

단순배정일 때

(Uniform allocation, $N_t = N_c$)

$$N_c = \frac{(Z\alpha + Z\beta)^2(P_c + P_t)}{\Delta^2}$$

공식 7 : Poisson approximation

이형배정일 때

(Nonuniform allocation, $N_t = RN_c$)

$$N_c = \frac{(Z\alpha + Z\beta)^2(P_c + P_t/R)}{\Delta^2}$$

공식 8 : Normal approximation for 2 independent means

단순배정일 때

(Uniform allocation, $N_t = N_c$)

$$N_c = \frac{\alpha(Z\alpha + Z\beta)^2V}{(M_c - M_t)^2}$$

M_c : 비교대조군에서의 평균

N_t : 시험군에서의 평균

V : 전체적인 결과 관찰치의 분산(δ^2)

공식 9 : Normal approximation for 2 independent means

이형배정일 때

(Nonuniform allocation, $N_t = RN_c$)

$$N_c = \frac{(Z\alpha + Z\beta)^2V(R+1)/R}{(M_c - M_t)^2}$$

M_c : 비교대조군에서의 평균

N_t : 시험군에서의 평균

V : 전체적인 결과 관찰치의 분산(δ^2)

공식 10 : Normal approximation for mean change

단순배정일 때

(Uniform allocation, $N_t = N_c$)

$$N_c = \frac{\alpha(Z\alpha + Z\beta)^2V_d}{(Md_t - Md_c)^2}$$

V_d : 각 개인에서의 처치 전 후간의 변동치들의 분산

공식 11 : Normal approximation for mean change

이형배정일 때

(Nonuniform allocation, $N_t = RN_c$)

$$N_c = \frac{(Z\alpha + Z\beta)^2V_d(R+1)/R}{(Md_t - Md_c)^2}$$

V_d : 각 개인에서의 처치 전 후간의 변동치들의 분산

연구대상자 수의 실제적인 계산

1. 독립적인 두 집단의 평균을 비교하기 위한 연구를 수행할 때, 두 집단의 평균의 차이(Δ)가 유의한 의미를 갖기 위해서 필요한 한 집단의 연구대상자 수는?

- 1) 연구대상자 수를 결정하는 공식

$$n = \frac{2s^2}{\Delta^2} \times (Z_\alpha + Z_\beta)^2$$

- 2) 사전에 필요한 정보

Δ : 연구자가 사전에 결정한 의미있는 차이

s : 표준편차

Z_α : 유의수준 α 일때의 값

Z_β : 검정력 $1 - \beta$ 일때의 값

〈문제〉 항고혈압 약물의 혈압강하 효과를 비교하기 위해 두 집단의 고혈압 환자들을 대상으로 한 집단에는 항고혈압 약물을 투여하고 나머지 집단에는 위약(Placebo)을 투여하려고 한다. 만약 두 집단의 연구대상자 수를 동일하게 한다면 각 집단에 필요한 연구대상자 수는? (두 집단의 감소된 혈압의 차이가 5mmHg, 표준편차 = 10mmHg, 양측검정이면서 유의수준 0.05, 검정력 90%)

〈해답〉 표 3을 참조하면

$$n = \frac{2 \times 10^2}{5^2} \times 10.51 = 84$$

따라서 필요한 연구대상자 수는 한 집단마다 84명의 고혈압 환자가 필요하다.

2. 독립적인 두 집단의 비율(%)를 비교하기 위한 연구를 수행할 때, 두 집단의 비율의 차이($P_1 - P_2$)가 유의한 의미를 갖기 위해서 필요한 한 집단의 연구대상자 수는?

표 2. 연구대상자 수의 결정을 위한 공식 요약

통계처리 방법	연구대상자 수 결정공식	가 정 (Assumption)	적 용 기 준
1. 결과변수가 상호배타적인 두가지 범주일 때(Binary outcome)			
Fisher's exact test	공식 1	독립적인 관찰	전체 사건 발생률이 0-1의 범위일 때
Chi-square approximation	공식 2, 3	독립적인 관찰	P_c and $P_t \geq 0.2$ but ≤ 0.8 : $N_c P_c$, $N_c Q_c$, $N_t P_t$, and $N_t Q_t$ all ≥ 15
Inverse sine transform approximation	공식 4, 5	독립적인 관찰	P_c and $P_t \geq 0.05$ but ≤ 0.95 : $N_c P_c$, $N_c Q_c$, $N_t P_t$, and $N_t Q_t$ all ≥ 15
Poisson approximation	공식 6, 7	독립적인 관찰	Low event rate(eg, P_c and $P_t \leq 0.05$: $N_c P_c$ and $N_t P_t$ ≥ 10
2. 결과변수가 연속적일 때(Continuous outcome)			
Normal approximation for 2 independent means	공식 8, 9	독립적인 관찰 분산이 공통적 정규분포할 때	N_c and $N_t \geq 30$
Normal approximation for mean changed	공식 10, 11	환자의 치료 전후 변화 관찰 분산이 공통적 정규분포할 때	N_c and $N_t \geq 30$

표 3. 유의수준과 검정력에 따른 $(Z_\alpha + Z_\beta)^2$ 값

Power, $1-\beta$	Significance level, α			
	One-tailed test		Two-tailed test	
	0.05	0.01	0.05	0.01
0.5	2.71	5.41	3.84	6.63
0.8	6.18	10.04	7.85	11.68
0.9	8.56	13.02	10.51	14.88

- 1) 연구대상자 수를 결정하는 Pocock(1984)의 공식

$$n = \frac{P_1(100 - P_1) + P_2(100 - P_2)}{(P_1 - P_2)^2} \times (Z_\alpha + Z_\beta)^2$$

- 2) 사전에 필요한 정보

$P_1 - P_2$: 연구자가 사전에 알고있는 각 집단의 비율(의미있는 비율의 차이)로 일반적으로 무작위 임상실험(randomised controlled trial)에서는 대조군의 비율(p_2)은 정해져 있기 때문에 중요한 임상적인 차이(Δ)만 사전에 결정되어 있으면 실험군의 비율(p_1)은 $p_1 = p_2 + \Delta$ 에 의해 구해진다.

Z_α : 유의수준 α 일때의 값

Z_β : 검정력 $1 - \beta$ 일때의 값

〈문제〉 금연을 위한 보건교육의 효과를 조사하기 위해서 금연교육을 실시하지 않은 흡연집단의 금연율이 40%이라면 각 집단의 필요한 연구대상자 수는? (최소한 10% 이상의 차이가 나면 금연교육의 효과를 인정하기로 하고, 양측검정이면서 유의수준 0.05, 검정력 80%)

〈해답〉 표 3을 참조하면

$$n = \frac{40 \times 60 + 35 \times 65}{10^2} \times 7.85 = 367$$

따라서 필요한 연구대상자 수는 한 집단마다 367명이다.

References

- 1) Curtis LM, Susan T. *Clinical trials, design, conduct and analysis*, NY, Oxford University Press, 1986, p71-89
- 2) Dobson AJ. *Calculating sample size. Transactions of the Menzies Foundation* 1984, p75-79
- 3) Freiman JA, Chalmers TC, Smith H, Kuebler RR. *The importance beta, the type II error and sample size in the design and interpretation of the randomized controlled trial. New England Journal of Medicine* 1978 ; 299 : 690
- 4) Lawrence MF, Curt DF, David LD. *Fundamentals of clinical trials, 2nd ed*, PSG Publishing Company, Inc, Littleton, 1985, p83-107
- 5) Pocock SJ. *Statistical aspects of clinical trial design. The Statistician* 1984 ; 31 : 1
- 6) Richard S. *Size of phase III cancer clinical trials. Cancer Treatment Report* 1985 ; 69(10) : 1087-1093
- 7) Stephen BH, Steven RC. *Designing clinical research, an epidemiologic approach. Williams & Wilkins, Baltimore, 1988, p128-150*