

## 임상 연구에서 연구 표본수의 산출

가톨릭대학교 의과대학 서울성모병원 정형외과

이호진 · 김양수 · 박 인

### Calculation of Sample Size in Clinical Trials

Hyo-Jin Lee, M.D., Yang-Soo Kim, M.D., In Park, M.D.

*Department of Orthopedic Surgery, Seoul St. Mary's Hospital Catholic University of Korea*

**Purpose:** This review aims to explain the definition and basic principle of statistical analysis and to clarify statistical issues related to the sample size calculation.

**Materials and Methods:** Many formulas are available that can be applied for different types of data and study design.

**Results:** The sample size is the number of patients or other experimental units that need to be calculated prior to the study. Determining the appropriate sample size is required to answer the research question.

**Conclusion:** Caution is needed when applying formula for the calculation of the sample size, as it is sensitive to error and even small differences in selected parameters can lead to large differences in the sample size.

**Key Words:** Sample size, Statistica analysis, Significance, Power, Effect size

## 서 론

연구 표본수(Sample size)란 연구에 참여하는 환자의 수 혹은 실험 단위를 뜻하는 것으로, 임상 연구에서 표본 수 산출의 목적은 의미있는 연구 결과 혹은 치료 효과를 산출해내기 위한 참여 단위의 수적 파악에 있다.<sup>1,2)</sup> 임상 연구를 시작하기 전에 연구자는 연구 표본수를 미리 결정하여야 한다.<sup>3)</sup> 엄격하게 잘 시행된 연구라 하여도 표

본수가 충분치 않으면 원하는 연구 결과를 얻지 못하거나, 얻더라도 그 결과의 신뢰도는 떨어진다. 반대로 연구 표본수가 너무 크다면 시간, 비용 측면에서 비효율적이며, 추적 손실이나 순응도 저하 등의 관리 문제가 발생할 수 있다. 원하는 연구의 결과가 유의한 의미를 갖기 위한 최소한의 연구 표본수를 정하는 것은 초보 연구자에게 있어서 큰 부담으로 다가오지만, 연구의 정당성을 뒷받침할 수 있는 표본수 산출의 정확한 기제는 현 시대의 학

※통신저자: 김 양 수

서울특별시 서초구 반포대로 222

가톨릭대학교 서울성모병원 정형외과

Tel: 02) 2258-6117, Fax: 02) 535-9834, E-mail: kysoos@catholic.ac.kr

접수일: 2013년 6월 4일, 1차 심사완료일: 2013년 6월 17일, 게재 확정일: 2013년 6월 20일

술지 혹은 연구 발표의 필수적인 요소가 되었다.

## 본 론

모든 임상 연구는 “다른 상황을 생각하게 하는 명확한 근거가 없는 한 현상적인 모든 차이는 없다(zero hypothesis, 영가설)”라는 과학 철학적 명제로부터 시작한다. 새로운 치료법의 효과를 입증하고자 한다면 “새 치료법은 효과가 없다”라는 영가설을 전제로 하여 연구를 시작한다.<sup>4)</sup> 연구자가 증명하고자 하는 실험 가설과 반대되는 입장, 증명되기 전까지는 효과도 없고 차이도 없다는 영가설을 귀무가설(null hypothesis,  $H_0$ )이라 하고, 반대로 연구자가 실험을 통해 규명하고자 하는 가설을 대립가설(alternative hypothesis,  $H_1$ )이라고 한다.

연구 표본수 산출의 첫 단계는 분석하고자 하는 변수의 성격을 이해하는 것이다. 변수는 크게 범주를 서로 구분하는 범주형(categorical)과 수량화가 가능한 연속형(numerical, continuous)으로 구분된다. 그리고 범주형 자료는 서열성 여부에 따라 명목(nominal) 척도와 순위(ordinal) 척도로 나누어지며, 연속형 자료는 실질적인 0(영)점 유무에 따라 비 척도와 구간 척도로 나뉜다. 가설을 검정에 사용할 통계적 모형은 결과 변수의 성격에 의해 거의 결정되는데, 예를 들면, 독립표본 t 검정이나 일원배치 분산분석은 연속형 변수의 비교에 쓰이고(예, BMI, ASES score), 카이제곱 검정은 범주형 변수의 비교(예, 당뇨와 비만과의 관계)에 이용된다. 이렇게 변수의 성격과 통계적 모형이 결정되면 이제 연구 표본수를 추정할 수 있다. 이 때 필요한 세 가지 요소가 유의수준, 검정력, 효과 크기이다. 이들 요소에 대해 하나씩 알아보자.

### 1. 유의 수준(Significance level) 및 제 1종 오류

유의 수준이란 통계 분석에서 제 1종 오류를 범할 가능성을 뜻하는 것으로서,  $p$  (probability)값으로 표기한다. 여기서 제 1종 오류( $\alpha$ )란 실제로 효과가 없는데 결과상 효과가 있다고 나오는 오류를 일컫는 것으로, 귀무가설을 거짓으로 기각하게 된다.<sup>5)</sup> 가장 문제가 되는 오류로 일반적으로 학계에서는 이를 5% 미만으로 통제하고 있다. 바꾸어 말하면 실제 효과가 없는데 효과가 있다는 결론이 도출될 가능성을 5% 미만으로 묶어두어야만 임상 시험 결과의 유효성을 인정하겠다는 것이다.

### 2. 검정력(Power) 및 제 2종 오류

검정력( $1-\beta$ )이란 실제로 효과가 있는 것을 통계 분석을 통해 효과가 있다고 증명할 수 있는 힘으로, 잘못된 귀무가설을 기각하는 능력을 말한다.<sup>6)</sup> 일반적으로 검정력은 0.80(80%)로 설정하고, 그 이상의 검정력을 요구할 경우 더 많은 수의 연구 표본수를 필요로 한다.<sup>7)</sup> 연구 표본수가 작아지면 실제로 효과가 있지만 결과상 효과가 없다는 결론을 도출하는 제 2종 오류( $\beta$ )의 가능성이 높아지며 검정력은 감소하게 된다. 연구자들은 제 1종 오류를 5%로 유지하면서 검정력을 최대화하는 통계 기법을 사용하고자 한다.<sup>8-11)</sup>

### 3. 효과 크기(Effect size)

효과 크기( $\delta$ )의 통계학적인 정의는 귀무가설( $H_0$ )과 대립가설( $H_1$ ) 간 차이의 크기로,<sup>12)</sup> 비교할 두 집단 간의 차이가 임상적으로 의미가 있는 정도로 설명할 수 있다.<sup>13)</sup> 효과 크기를 구하는 방법은 통계모형마다 다르다. 예를 들어 상관분석에서는 상관계수  $r$ 로, 독립표본 t검정에서는  $d$ 로, 분산분석에서는  $f$ 로 표현된다. 효과 크기가 ‘0’이라는 것은 실험 집단의 평균과 통제 집단(control group 혹은 대조군)의 평균이 같다는 뜻이고, +값이 나왔다는 것은 실험 집단의 평균이 통제 집단의 평균보다 크다는 것을 의미하고, -값은 그 반대의 경우이다. 절대값이 클수록 실험 집단과 통제 집단의 차이가 크다는 것을 의미하는데, 어느 정도의 효과 크기를 사용할 것인지는 임상적인 판단에 의한다.

### 4. 표본수 추정을 위한 공식

1) 이분형 변수(Dichotomous variables)에 대한 표본수 산정

$$\text{산정공식 } n_1=n_2=[(2p_m q_m)^{1/2} z_{1-\alpha/2} + (p_1 q_1 + p_2 q_2)^{1/2} z_{1-\beta}]^2 / \Delta^2$$

$n_1$ =비교군의 연구 표본수

$n_2$ =시험군의 연구 표본수

$p_1, p_2$ =각 군의 예상되는 결과 발생률

$$p_m=(p_1+p_2)/2$$

$$q_m=1-p_m$$

$$\Delta=p_1-p_2$$

$Z$ =평균이 0, 분산이 1인 표준정규분포 곡선에서 해당하는 지점의 값

시험 결과가 생존 혹은 사망, 완치 혹은 재발 등과 같

이 서로 상반되는 종속 변수로 표현될 경우 이분형 변수라 한다. 효과의 차이를 계산하기 위해서 시험군과 비교군의 차이를 미리 설정하여야 한다. 일반적으로  $p_1$ 값은 특정 연구와 유사한 임상 시험의 결과로부터 얻어져야 하나 기존 문헌에서 값을 얻지 못하는 경우에는 사전연구(pilot study)를 통해 그 값을 취할 수 있다. 일반적으로  $p_1$ 값을 과대하게 추정하면, 연구 표본수가 필요한 수보다 적게 계산되어 문제가 될 수 있다.

2) 연속형 변수 (Continuous variables)에 대한 표본수 산정

$$n_1=n_2=2(\sigma^2)(z_{1-\alpha/2}+z_{1-\beta})^2 / \Delta^2$$

$n_1$ =비교군의 연구 표본수

$n_2$ =시험군의 연구 표본수

$\Delta$ =각 군의 평균값의 차이

$\sigma^2$ =시험군, 비교군 공통의 분산값

시험의 변수가 연속형일 때는 효과의 차이는 평균치의 함수로 표현되며, 이분형 결과에서  $p_1$ 을 추정하는 것과 마찬가지로 연구 표본수를 계산하기 위해서 분산의 추정치를 필요로 한다. 임상 연구에서 연속형 변수의 표본수 산정의 가장 중요한 요소는 임상적인 의미를 가질 수 있는 두 군간의 차이를 파악하거나, 환자에게 영향을 끼칠 수 있는 최소한의 차이점을 파악하는 것이다.

5. 탈락자의 처리(Accounting for dropout and loss of follow-up)

연구자는 임상 연구 등록 후에 명단에서 제외되거나 추적 관찰 중 누락될 수 있는 탈락자에 대해 고려하여야 한다. 이런 이유로 연구 전 추정된 표본수보다 많은 표본수를 산정하여야 한다. 또한, 추적 기간이 증가할수록 탈락자는 증가하게 된다. 연구 설계와 성격에 따라 다르겠지만, 일반적으로 10%에서 15%정도의 표본수를 증원한다.

6. 웹사이트에서 계산하기

표본수 산정을 위해 직접 위의 수식을 적용하는 데는 적잖은 노력과 시간을 요하기 때문에, 많은 기관에서 간단하게 표본수를 구할 수 있는 서비스를 웹 상으로 제공하고 있다. 그 중에 하나인 University of British Columbia (UBC)의 Department of Statistics 에서 제공하는 서비스(<http://stat.ubc.ca/~rollin/stats/ssize/>)를 “Is Early Passive Motion Exercise Necessary After Arthroscopic Rotator Cuff Repair”<sup>14)</sup> 논문을 예로 들어 소개하고자 한다.

1) 이분형 변수(Dichotomous variables)에 대한 표본수 산정

회전근 개 파열 환자의 수술적 치료 후, 조기 관절 운동군과 지연 관절 운동군의 재파열율을 비교하고자 한다. 본 연구는 80%의 검정력을 갖고 유의 수준 5%을 유지하고자 한다. 웹사이트 상에서 네번째에 있는 “Comparing Proportions for Two Independent Samples”를 선택한다(Fig. 1). 문헌 고찰을 통해 회전근 개 파열 수술 후 일반적인(지연) 관절 운동을 했을 경우 30%의 재파열을 보였다( $p_1$ ). 10%이상의 재파열을 감소를 효과 크기로 설정할 경우, 조기 관절 운동 군의 재파열율을 20%로 설정하여  $p_2$ 값에 대입한다. 대입한 결과 본 연구에서는 각 군당 최소 294명의 환자가 필요하며(Fig. 2), 10%의 탈락률을 고려하면 최종적으로 각 군당 327명을 모집하면 된다.

2) 연속형 변수(Continuous variables)에 대한 표본수 산정

회전근 개 파열에 대한 수술적 치료 후, 조기 관절 운동군과 지연 관절 운동군 간의 치료 효과를 비교하고자 한다. 치료 효과를 판단할 수 있는 기능 평가는 ASES score를 사용하기로 하였다. 20명의 환자를 대상으로 시행한 예비 연구를 통하여, 조기 관절 운동군의 평균 ASES score는 75점이었으며( $\mu_1$ ), 의미 있는 평균의

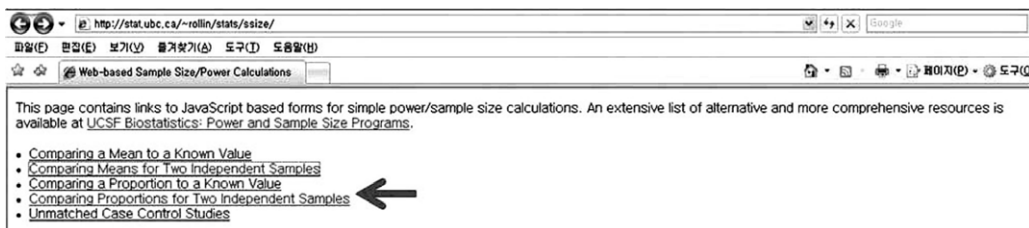


Fig. 1. Select “Comparing Proportions for Two Independent Samples” for sample size calculation on dichotomous variables.

차이(effect size)를 8점 그리고 표준편차를 12점으로 설정하였다. 이럴 경우, 지연 관절 운동군의 평균( $\mu_2$ ) 값으로는 조기 관절 운동군의 평균인 75점에서 효과 크기인 8점을 뺀 67점을 대입하면 된다. 90% 검정력을 갖

고 유의 수준 5%를 유지하고자 한다. 같은 방식으로 웹사이트에 접속하여 두번째에 있는 “Comparing Means for Two Independent Samples”를 선택한다(Fig. 3). 계산하면 각 군당 48명의 환자가 산출되며, 결과적으로

**Inference for Proportions: Comparing Two Independent Samples**

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values (as decimal fractions) for  $p_1$  (proportion in population 1) and  $p_2$  (proportion in population 2) and, if calculating power, a sample size (assumed the same for each sample). You may also modify  $\alpha$  (type I error rate) and the power, if relevant. After making your entries, hit the **calculate** button at the bottom.

- Calculate Sample Size (for specified Power)
- Calculate Power (for specified Sample Size)

Enter a value for  $p_1$ :

Enter a value for  $p_2$ :

- 1 Sided Test
- 2 Sided Test

Enter a value for  $\alpha$  (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

Reference: The calculations are the customary ones based on the normal approximation to the binomial distribution. See for example *Hypothesis Testing: Categorical Data - Estimation of Sample Size and Power for Comparing Two Binomial Proportions* in Bernard Rosner's *Fundamentals of Biostatistics*.

**Fig. 2.** Enter proportion (or expected result) for each group ( $p_1$  &  $p_2$ ). Enter values for required type I error level ( $\alpha$ ) and desired power ( $1-\beta$ ). After filling out the blanks with appropriate values, press “calculate”.

Web-based Sample Size/Power Calculations - Windows Internet Explorer

http://stat.ubc.ca/~rollin/stats/ssize/

This page contains links to JavaScript based forms for simple power/sample size calculations. An extensive list of alternative and more comprehensive resources is available at UCSF Biostatistics: Power and Sample Size Programs.

- [Comparing a Mean to a Known Value](#)
- [Comparing Means for Two Independent Samples](#) ←
- [Comparing a Proportion to a Known Value](#)
- [Comparing Proportions for Two Independent Samples](#)
- [Unmatched Case Control Studies](#)

**Fig. 3.** Select “Comparing Means for Two Independent Samples” for sample size calculation on continuous variables.

**Inference for Means: Comparing Two Independent Samples**

(To use this page, your browser must recognize JavaScript.)

Choose which calculation you desire, enter the relevant population values for  $\mu_1$  (mean of population 1),  $\mu_2$  (mean of population 2), and  $\sigma$  (common standard deviation) and, if calculating power, a sample size (assumed the same for each sample). You may also modify  $\alpha$  (type I error rate) and the power, if relevant. After making your entries, hit the **calculate** button at the bottom.

- Calculate Sample Size (for specified Power)
- Calculate Power (for specified Sample Size)

Enter a value for  $\mu_1$ :

Enter a value for  $\mu_2$ :

Enter a value for  $\sigma$ :

- 1 Sided Test
- 2 Sided Test

Enter a value for  $\alpha$  (default is .05):

Enter a value for desired power (default is .80):

The sample size (for each sample separately) is:

**Fig. 4.** Enter mean of population from each group ( $\mu_1$  &  $\mu_2$ ). Enter standard deviation for sigma value. Enter values for required type I error level ( $\alpha$ ) and desired power ( $1-\beta$ ). After filling out the blanks with appropriate values, press “calculate”.

10%의 탈락률을 적용시킨다면 본 연구의 의미 있는 결과 산출을 위해서는 각 군당 최소 54명의 환자가 필요하다(Fig. 4).

## 결 론

연구를 계획하는데 있어, 표본수 산출은 가장 먼저 선행되어야 함과 동시에 가장 중요한 절차라고 할 수 있다. 연구의 성격을 올바르게 파악하여 표본수 산출에 적합한 공식을 설정하여야 하며, 표본수 산출을 통해 충분한 검정력 여부 및 임상적으로 유의한 결과를 도출할 수 있는 최소한의 효과 크기를 파악하는 것이 중요하다. 또한 과도한 크기의 표본수 산출은 오히려 연구를 비효율적으로 만들기 때문에 적절한 표본수 산출은 매우 중요하다. 표본수 산출은 오류에 매우 민감하며 그릇된 결과를 도출할 수 있기 때문에, 세심한 주의를 요하며 필요시 전문가의 도움을 요청하는 것에 주저하지 말아야 하겠다.

## REFERENCES

- 1) Whitley E, Ball J. *Statistics review 4: Sample size calculations. Crit Care.* 2002;6:335-41.
- 2) Kirby A, GebSKI V, Keech AC. *Determining the sample size in a clinical trial. Med J Aust.* 2002;177: 256-7.
- 3) Altman DG. *Practical statistics for medical research. 1st ed. London: Chapman and Hall; 1991. 25-101.*
- 4) Bae JM. *An illustrated guide to medical statistics using SPSS. 1st ed. Seoul: Hannarae; 2012. 20-1.*
- 5) Machin D, Campbell MJ, Fayers PM, Pinol APY. *Sample size tables for clinical studies. 2nd ed. Oxford: Blackwell Science; 1997. 1-13.*
- 6) Pallant J. *SPSS survival manual: A step by step guide to data analysis using SPSS. 4th ed. New York: McGraw-Hill; 2010. 121-202.*
- 7) Siegel S, Castellan NJ. *Nonparametric statistics for the behavioral sciences. 2nd ed. New York: McGraw-Hill; 1988. 19-37.*
- 8) Hamilton LC. *Data analysis for social scientists: A first course in applied statistics. 1st ed. London: Duxbury Press; 1996. 30-42.*
- 9) Cohen J. *Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: L. Erlbaum Associates; 1988. 19-74.*
- 10) Bland M. *An introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2000. 14-32.*
- 11) Munro BH. *Statistical methods for health care research. 5th ed. Philadelphia: Lippincott Williams & Wilkins; 2005. 73-106.*
- 12) Thorne BM. *Statistics for the behavioral sciences. 1st ed. Houston: Mayfield; 1989. 16-44.*
- 13) Friedman LM, Furberg C, DeMets DL. *Fundamentals of clinical trials. 4th ed. New York: Springer; 2010. 133-162.*
- 14) Kim YS, Chung SW, Kim JY, Ok JH, Park I, Oh JH. *Is early passive motion exercise necessary after arthroscopic rotator cuff repair? Am J Sports Med.* 2012;40:815-21.

### 초 록

**목적:** 임상 연구의 통계 분석에 있어서 표본수 산출의 의미와 기본 방식에 대해서 알고자 한다.

**대상 및 방법:** 자료의 분류, 연구 디자인 및 도출하고자 하는 결과의 성격에 따라 각기 다른 식을 적용시켜 표본수를 산출 한다.

**결과:** 표본수 산출은 임상 연구를 시작하기 전에 선행되어야 하며, 적절한 표본수 산출은 유의한 결론 도출에 필수 불가결하다.

**결론:** 표본수 산출은 예러나 작은 변수에 민감하기 때문에 특정 식에 적용시킬 때는 주의를 요한다.

**색인 단어:** 표본수, 통계학적 분석, 유의성, 검정력, 효과 크기