

## 1

## 치의학 연구에서의 표본크기 산출

전남대학교 치의학전문대학원 교정학교실, 치의학 연구소

임 회 정

## ABSTRACT

## Sample size determination in dental research

Department of Orthodontics, Chonnam National University School of Dentistry, Dental Science Research Institute  
Hoi-Jeong Lim, Ph.D.

Sample size determination is critical, but not easy to do. Sample size defined as the number of observations in a sample to be studied should be big enough to have a high likelihood of detecting a true difference between groups. Practical procedure for determining sample size, using G\*power and previous dental articles, was shown in this study. Examples involving independent t-test, paired t-test, one-way analysis of variance(ANOVA), and one-way repeated-measures(RM) ANOVA were used. The purpose of this study is to enable researchers with non-statistical backgrounds to use in practice freely available statistical software G\*power to determine sample size and power.

Key words : sample size; power analysis; dental research

Corresponding Author

Hoi-Jeong Lim, Ph.D.

Department of Orthodontics, Chonnam National University School of Dentistry

Dental Science Research Institute

33 Yongbong-ro, Buk-gu, Gwangju 500-757, Korea

Tel : +82-62-530-5830 Fax : +82-62-530-5659, E-mail : hjlim@jnu.ac.kr

이 논문은 전남대학교 병원 학술연구비 지원(#CRI 11040-1) 에 의하여 연구되었음.

## I. 서론

표본크기(sample size)란 연구에 참여하는 환자, 사람, 혹은 동물의 수를 말하는 것으로, 임상연구에서

표본크기 산출의 목적은 유의한 연구 결과를 산출해내기 위한 표본크기를 파악하는 데에 있다. 너무 작은 표본으로 연구를 한다면 낮은 통계적 검정력(power)을 가지게 되어 증거가 부족하여 결론을 내릴 수 없거나

정확하지 않은 결론을 내릴 수 있다. 반대로 너무 많은 표본으로 연구를 한다면 시간, 자원 등을 낭비할 수 있고 추적 관찰 기간 중의 중도 탈락(follow-up loss)이 발생할 수 있으며 비윤리적인 문제가 생길 뿐 더러, 통계학적으로는 유의하다고 결론을 내릴 수 있지만 임상적으로 의미가 없는 경우가 생길 수 있다. 적절하지 않은 표본수로 연구를 디자인하는 것은 연구에 대한 잘못된 결론을 내릴 수 있고 부적절한 치료로 이어질 수 있다. 그러므로 연구 계획 단계에서 표본 수를 산출하는 것은 반드시 필요한 단계라 할 수 있다.

## II. 표본크기 산출에 필요한 고려할 사항들

검정력과 표본크기를 구하는 데는 여러 고려할 사항들이 많지만 그 중 중요하게 고려되어야 할 사항들은 제1종 오류와 제2종 오류, 효과크기(effect size) 등이다. 통계 방법마다 다른 공식을 가지고 표본크기를 계산해야 하므로 각자의 연구에서 어떤 변수를 주요 결과 변수(primary endpoint)로 하여 어떤 통계 방법을 사용하여 분석할 것인지 연구 계획단계에서 잘 계획되어야 한다. 아래에서 고려할 사항들에 대하여 더 자세히 알아보자.

### 1. 제1종 오류( $\alpha$ )와 제2종 오류( $\beta$ ), 그리고 검정력(power)

가설검정은 표본에 기초하여 의사 결정이 이루어지기 때문에 오류가 발생할 가능성이 항상 존재한다. 이러한 오류에는 제1종 오류와 제2종 오류가 있다. 제1종 오류는 효과가 없는데도 있다고 할 확률이고 제2종 오류는 효과가 있는데도 없다고 할 확률이다. 그러므로 검정력은 효과가 있을 때 있다고 할 확률이다. 대개 제1종 오류는 0.05, 검정력은 0.8의 고정된 값을 사용한다. 그러므로 오류의 개념은 어렵지만 사용하기

는 쉽다.

$$\text{제1종 오류}(\alpha) = P[\text{Reject } H_0 | H_0 : \text{True}]$$

$$\text{제2종 오류}(\beta) = P[\text{Do not reject } H_0 | H_0 : \text{False}]$$

$$\text{검정력}(\text{Power}) = 1 - \beta = P[\text{Reject } H_0 | H_0 : \text{False}]$$

where  $H_0$  : 효과가 없다는 귀무가설(Null hypothesis)

### 2. 효과크기(effect size)

효과크기란 임상적으로 의미 있는 최소한의 차이를 말한다. 이러한 효과크기를 알아내는 것이 임상시험의 목표이기도 하다. 효과크기가 클수록 검정력은 커지지만 표본크기는 작아진다. 반대로 효과크기가 작다면 그 작은 차이를 유의하게 발견하기 위해서 표본크기는 커져야 한다. 효과크기는 선행논문이나 예비연구(pilot study)에서 주어진 평균, 표준편차, 비율 등에 의해 얻어진다.

### 3. 가설검정의 종류

양측검정이란 귀무가설과 대립가설에서 두 그룹의 평균이 같은지 다른지를 검정하는 것이며 단측검정은 한 그룹의 평균이 다른 그룹의 평균보다 큰지 작은지를 검정하는 것이다.

### 4. Drop rate

보통 drop rate는 20%미만으로 설정한다. 대개의 경우는 10%의 drop rate를 사용하는데 연구기간이 길어지면 더 많은 추적손실(follow-up loss)이 발생하므로 drop rate를 좀 더 높게 잡을 수 있다. 이 경우  $G^*$ power로 구한 N값을 0.9로 나누어 Final N을 구한다(Final N : N = 1 : 0.9).

### 5. 주요 결과 변수(primary endpoint) 결정

연구 계획 단계에서 주요 결과 변수(primary endpoint)가 결정되어야 하고 그 변수에 따른 표본크기를 산출해야 한다. 주요 결과 변수는 주어진 치료가 효과가 있는지 보기 위해 연구 마지막에 측정되는 주요 결과이며, 예를 들어 연구 대상 질병의 사망자수, 치료군과 대조군 사이에 생존자수의 차이 등을 말한다.

### Ⅲ. G\*power를 이용한 표본크기 산출

G\*power는 독일의 듀셀도르프 대학에서 만든 것이며 <http://www.gpower.hhu.de/en.tml> 에서 G\*power 3.1.9.2를 무료로 다운로드를 받을 수 있다. 아래는 통계 방법별로 치의학 저널에 실린 논문들을 바탕으로 G\*power를 이용하여 검정력과 표본크기를 계산하는 방법에 대해 다루었다.

#### 1. Independent t-test를 이용한 표본크기 산출

2004년 미국 교정 학회지에 출판된 Miyawaki 등이 쓴 논문<sup>1)</sup>에서 사용된 Table 1을 근거로 검정력과 표본크기를 구하였다. 이 Table 1에서는 이같이 환자

와 정상집단 모두에서 occlusal splint의 사용 여부가 긴장을 풀고 있는 동안, 이를 악 물고 있는 동안, 씹는 것 같은 운동을 하고 있는 동안의 타액흐름속도를 증가시키는지에 대한 가설을 검정하였다. 그 결과 이같이 환자 8명과 정상인 8명에서의 타액 흐름속도를 비교하여 모든 변수에서 유의한 차이를 얻지 못했다. 이 경우 G\*power를 이용하여 8명의 표본크기를 가지고 검정력이 80%가 넘는지 알아보기 위하여 테이블에 나와 있는 표준오차(SE)를 이용하여 아래의 식으로 표준편차(SD)를 구하였다.

$$SE = \frac{SD}{\sqrt{n}}$$

G\*power를 이용하여 Figure 1 아래의 1번부터 10번의 순서로 검정력을 구한 결과 스프린트를 사용했을 때나 사용하지 않았을 때 모두 검정력은 모두 40% 미만으로 낮게 나왔다.

각 그룹 당 8명의 표본크기를 가지고 independent t-test를 한 결과 두 그룹의 차이가 유의하지 않다는 결론을 내린 연구지만, 80%의 검정력을 가지려면 표본크기가 얼마인지 알아보니 Without splint의 Relaxing 이라는 변수에서는 각 군 당 96명의 표본크기를 가져야 한다는 것을

Table 1. Power and sample size determination based on the Table1 of the study<sup>1)</sup> that used independent t-test

splint/test movements	Control group (n = 8)			Bruxism group (n = 8)			P	Power*	Min N <sup>‡</sup> ( per gp )	Final N <sup>§</sup> ( per gp )
	Mean	SE	SD*	Mean	SE	SD*				
Without splint										
Relaxing	0.49	0.05	<b>0.1414</b>	0.42	0.07	<b>0.1980</b>	ns	<b>0.1182</b>	<b>96</b>	<b>107</b>
Chewing-like movement	0.66	0.05	<b>0.1414</b>	0.49	0.11	<b>0.3111</b>	ns	<b>0.2589</b>	<b>33</b>	<b>37</b>
Clenching	0.52	0.05	<b>0.1414</b>	0.37	0.07	<b>0.1980</b>	ns	<b>0.3688</b>	<b>22</b>	<b>25</b>
With splint										
Relaxing	0.64	0.05	<b>0.1414</b>	0.52	0.14	<b>0.3960</b>	ns	<b>0.1171</b>	<b>98</b>	<b>109</b>
Chewing-like movement	0.89	0.08	<b>0.2263</b>	0.81	0.21	<b>0.5940</b>	ns	<b>0.0627</b>	<b>497</b>	<b>553</b>
Clenching	0.63	0.05	<b>0.1414</b>	0.64	0.15	<b>0.4243</b>	ns	<b>0.0504</b>	<b>15701</b>	<b>17446</b>

SE, Standard error; ns, not significant between groups (unpaired t test or Mann-Whitney U test between groups).

\*SD, standard deviation

+Power, probability that it rejects a false null hypothesis.

‡Min N, minimum sample size calculated by G\*Power

§Final N; Min N considering the drop rate 10%

G\*power를 통해 알아냈다. drop rate 10%를 고려해서 96명을 0.9로 나누어 107명의 표본크기를 얻었다. 이 연구에서 주요 결과변수가 Without splint의 Relaxing이라는 변수라면 적어도 그룹 당 107명의 연구대상을 모집해야 할 것이다(Table 1).

## 2. paired t-test를 이용한 표본크기 산출

2005년 Angle Orthodontist 저널에 출판된 Benson 등이 쓴 논문<sup>2)</sup>에서 Captured Slides와 Digital Camera Image의 demineralized lesion area의 차이를 측정된 결과 두 이미지 간에

유의한 차이가 있었고(p=0.029) 병소가 있는 부위의 평균 gray level과 건전한 부위의 평균 gray level에서도 두 이미지 간에 유의한 차이가 있었다(p=0.002, p=0.001). 그러나 luminance intensity(LI)%에서는 두 이미지 간에 유의한 차이를 발견할 수 없었다(p=0.148). 이 테이블에 대해 표본크기가 적절한지 검정력을 계산해 보았다.

27 sites의 표본크기를 가지고 paired t-test를 한 결과 Area에서 두 이미지 간의 차이가 유의하다는(p=0.029) 결론을 내렸지만 0.8보다 작은 검정력(power=0.62)을 얻었다. 즉, p-value가 0.05보다 작다고 해서 모든 검정력이 0.8보다 큰 것은 아니기

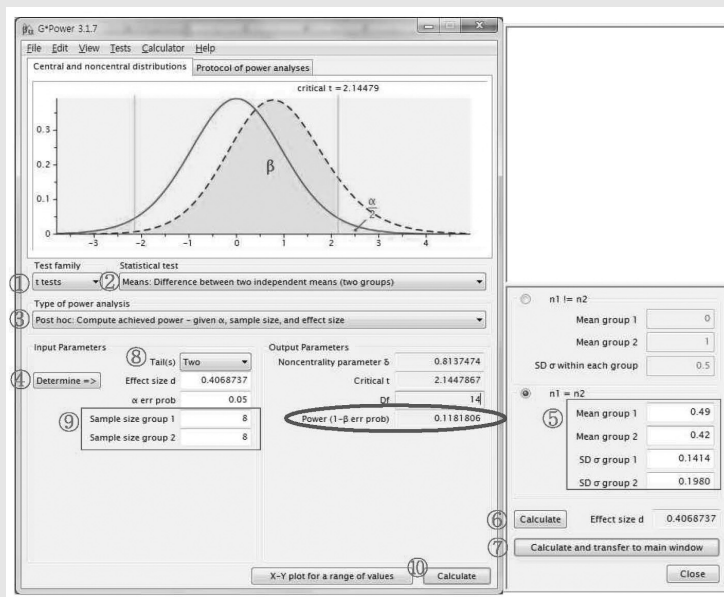


Fig. 1. Power analysis based on independent t-test

- ① Test family에서 t tests를 선택한다.
- ② Independent t-test를 선택하기 위해 Means: Difference between two independent means를 선택한다.
- ③ 검정력을 계산하기 위해 Post hoc을 선택한다. sample size를 계산하려면 A priori를 선택하면 된다.
- ④ Determine을 눌러 효과크기를 계산한다.
- ⑤ 옆의 윈도우로 옮겨 sample size가 같은 경우이므로 n1 = n2에서 두 그룹의 평균과 표준편차를 입력한다.
- ⑥ Calculate 버튼을 누르면 효과크기가 계산된다.
- ⑦ Calculate and transfer to main window를 눌러 계산된 효과크기를 옆의 윈도우에 있는 Effect size d로 옮긴다.
- ⑧ Tail(s)에서 Two를 선택함으로써 양측 검정을 선택한다.(One: 단측 검정을 말한다.)
- ⑨ 두 그룹의 sample size 8을 입력한다.
- ⑩ Calculate 버튼을 눌러 Power(검정력)를 계산한다.

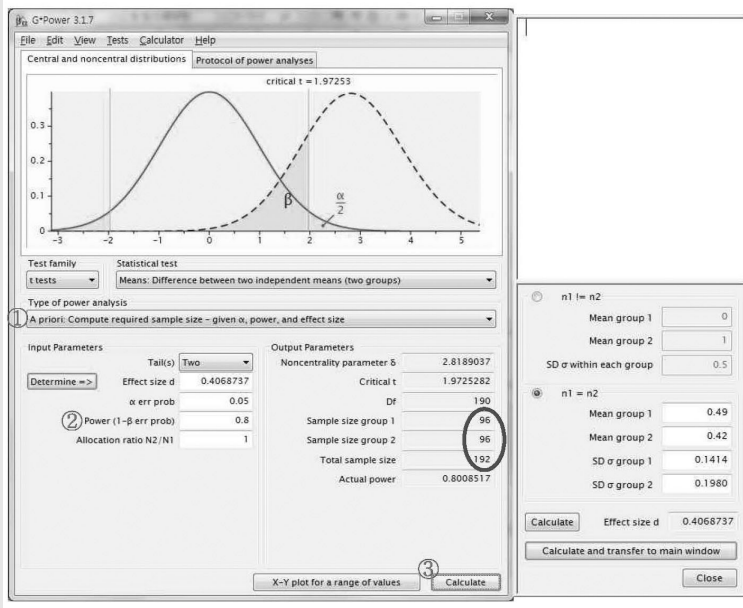


Fig. 2. Sample size determination based on independent t-test

나머지는 Post hoc인 위의 경우와 같고

- ① Figure 1에서 Type of power analysis에서 post hoc을 A priori로 바꾼다.
- ② Power를 0.8로 바꾼다.
- ③ Calculate 버튼을 눌러 그룹 별 sample size 96을 구했다.

때문에 이 경우 사후 검정력을 계산할 필요가 있다. 즉,  $p < 0.05$ 라고 해서 충분한 표본크기를 가졌다고 말할 수 없다<sup>3)</sup>. p-value가 0.05보다 작으면서 검정력이 0.8보다 크기 위해서는 적어도 41명의 연구대상자를 가지고 연구를 해야 한다. Figure 5를 참조하면,  $p > 0.05$ 일 때 표본크기는 22 미만이며, 22를 넘더라도 표본크기 40까지는 0.8미만의 검정력을 가진다. 아래의 테이블에 있는 Lesion과 Sound 변수에서는 검정력이 0.9를 넘었고 위 테이블에서 보여진 Lesion  $N=27$ 과 Sound  $N=17$ 은 80% 검정력을 가진 최소한의 표본크기( $N=24, 14$ )를 확보했다고 보여진다. 그러나 LI% 변수에서는 유의한 차이를 발견해내기 위한 표본크기는 10%의 drop rate를 고려한 109명의 연구대상을 확보해야 할 것이다.(Table 2, Table 3, Figure 5)

### 3. one-way ANOVA

2014년 Angle Orthodontist 저널에 출판된 Aslan으로 연구<sup>4)</sup>는 세 그룹 (1) miniscrew anchorage를 사용한 Forsus™ Fatigue Resistant Device(FRDMS) 그룹, (2) conventional FRD 그룹, (3) untreated class II control 그룹에서 Skeletal, Dental, and Soft Tissue Parameters들의 계측치들의 평균을 가지고 비교하여 dentofacial effect를 평가하고자 한 것이다. 그 결과 U1/HRL과 L1-VRL 계측치가 세 그룹간의 유의한 차이를 보였다.

위 테이블이 U1/HRL과 L1-VRL 변수의 검정력을 구해본 결과 0.8보다 컸으며, 필요한 표본크기도 15, 45명으로 최소한의 표본크기는 확보된 것으로 보

Table2. Power and sample size determination based on the Table4 of the study<sup>2)</sup> that used paired t-test

	Paired Differences		95% CI of the Difference		P	N	Power <sup>+</sup>	Min N <sup>≠</sup>	Final N <sup>§</sup>
	Mean	SD	Lower	Upper					
Area	-0.29	0.64	-0.54	-0.03	0.029	27	<b>0.6206</b>	<b>41</b>	<b>46</b>
Lesion	9.08	13.77	3.63	14.52	0.002	27	<b>0.9094</b>	<b>21</b>	<b>24</b>
Sound	12.11	12.93	5.46	18.76	0.001	17	<b>0.9517</b>	<b>12</b>	<b>14</b>
LI%	-2.86	9.96	-6.80	1.08	0.148	27	<b>0.3009</b>	<b>98</b>	<b>109</b>

+Power; probability that it rejects a false null hypothesis.

≠Min N; minimum sample size calculated by G\*Power

§Final N; Min N considering the drop rate 10%

Table3. The relationship among P-value, power, and sample size in demineralized lesion area

p>0.05 and power<0.8	p<0.05 and power<0.8	p<0.05 and power>0.8
<22	22 ≤ <41	≥41

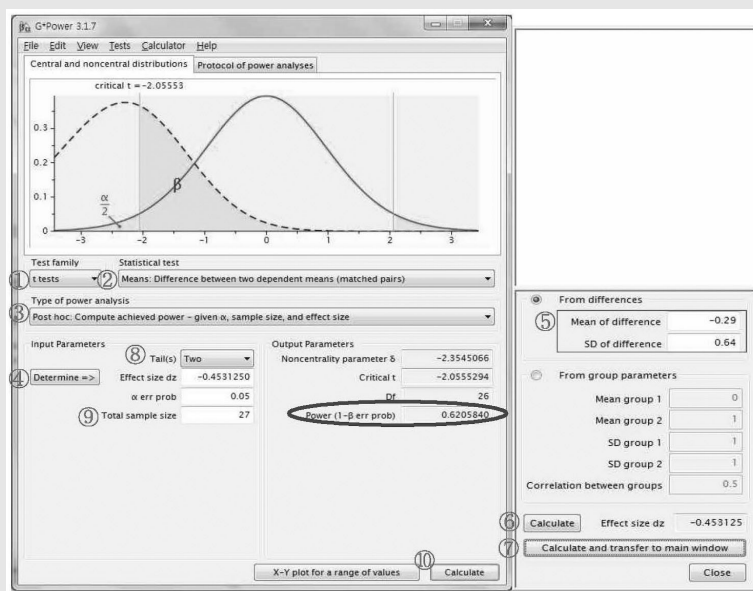


Fig. 3. Power analysis based on paired t-test

- ① Test family에서 t tests를 선택한다.
- ② Paired t-test를 선택하기 위해 Means: Difference between two dependent means를 선택한다.
- ③ 검정력을 계산하기 위해 Post hoc을 선택한다. sample size를 계산하려면 A priori를 선택하면 된다.
- ④ Determine을 눌러 효과크기를 계산한다.
- ⑤ 옆의 윈도우로 옮겨 From differences에서 차이의 평균과 표준편차를 입력한다.
- ⑥ Calculate 버튼을 누르면 효과크기가 계산된다.
- ⑦ Calculate and transfer to main window를 눌러 계산된 효과크기를 옆의 윈도우에 있는 Effect size d로 옮긴다.
- ⑧ Tail(s)에서 Two를 선택함으로써 양측 검정을 선택한다.(One: 단측 검정을 말한다.)
- ⑨ Total sample size 27을 입력한다.
- ⑩ Calculate 버튼을 눌러 Power(검정력)를 계산한다.

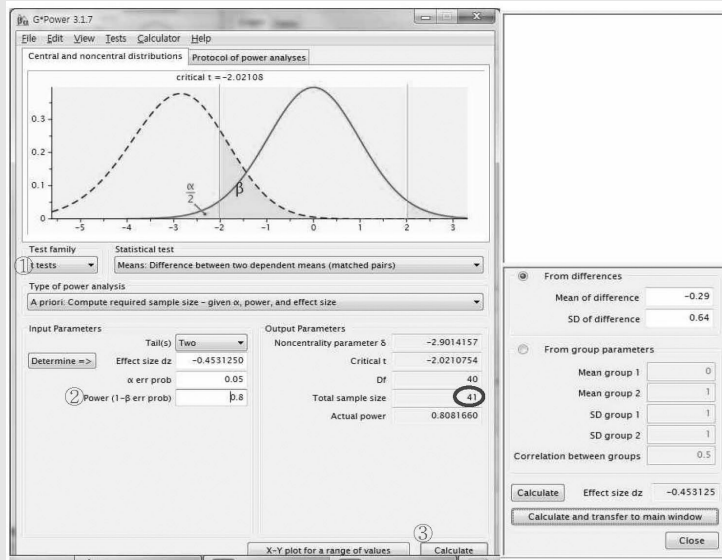


Fig. 4. Sample size determination based on paired t-test

- ① Figure3에서 Type of power analysis에서 Post hoc을 A priori로 바꾼다.
- ② Power를 0.8로 바꾼다.
- ③ Calculate 버튼을 눌러 그룹 별 sample size 41을 구했다.

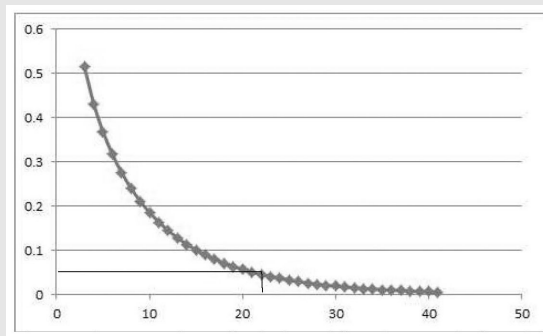


Fig. 5. Plot for total sample size(x-axis) against p-value(y-axis)

- ① 표본크기(N) 3부터 41까지의 paired t-test의 검정 통계량  $T = \frac{\bar{D}}{S_D/\sqrt{N}}$  과 자유도(N-1)를 구한다.
- ② 표본크기와 Figure3에서 구한 효과크기로 검정력을 구한다.
- ③ 표본크기에 따른 p값을 y축으로, 표본크기를 x축으로 그래프를 그린다.
- ④ p값이 0.05 보다 큰 표본크기는 22보다 클 때라는 것을 알았다.

여 진다. 그러나 다른 변수에서는 검정력이 0.8보다 작아 그룹 간의 유의성을 발견하기 위해서는 더 많은 표본크기가 필요하였다. 여기서도 주요 결과변수 (prime endpoint)가 U1/HRL이나 L1-VRL 변수 라면 선택된 표본크기에 문제가 없지만 그 외의 다른

변수라면 더 많은 표본크기를 선택하여 연구를 시작해야 할 것이다(Table 4).

#### 4. one-way repeated-measures(RM) ANOVA

Table 4. Power and sample size determination based on the Table2 of the study<sup>3)</sup> that used one-way ANOVA

	1 (n = 16)	2 (n = 17)	3 (n = 15)	P	SD <sup>‡</sup>	Power <sup>+</sup>	Min N <sup>‡</sup>	Final N <sup>§</sup>
SN, mm	74.70 ± 3.70	76.87 ± 3.13	74.66 ± 3.46	.082	3.43	0.4323	108	120
SNA, °	80.53 ± 3.76	78.27 ± 2.19	78.14 ± 3.02	.067	3.04	0.5709	78	87
A-VRL, mm	70.03 ± 3.50	72.32 ± 3.72	69.54 ± 3.29	.079	3.52	0.5395	84	94
U1/HRL, °	112.03 ± 5.11	110.17 ± 5.31	95.55 ± 9.62	<.001*	6.89	0.9999	15	17
L1-VRL, mm	68.00 ± 3.98	68.91 ± 4.66	64.10 ± 4.14	<.001*	4.28	0.8309	45	50
L1-MP, mm	43.26 ± 2.24	43.01 ± 2.47	42.14 ± 2.87	.057	2.53	0.1836	282	314

\* FRD indicates fatigue resistant device; FRDMS, FRD treatment used with miniscrew anchorage.

Values for groups 1, 2, and 3 are expressed as the mean ± the standard deviation.

\* P < .01; one-way analysis of variance.

‡SD; standard deviation

+power; probability that it rejects a false null hypothesis

‡Min N; minimum sample size calculated by G\*Power

§Final N; Min N considering the drop rate 10%

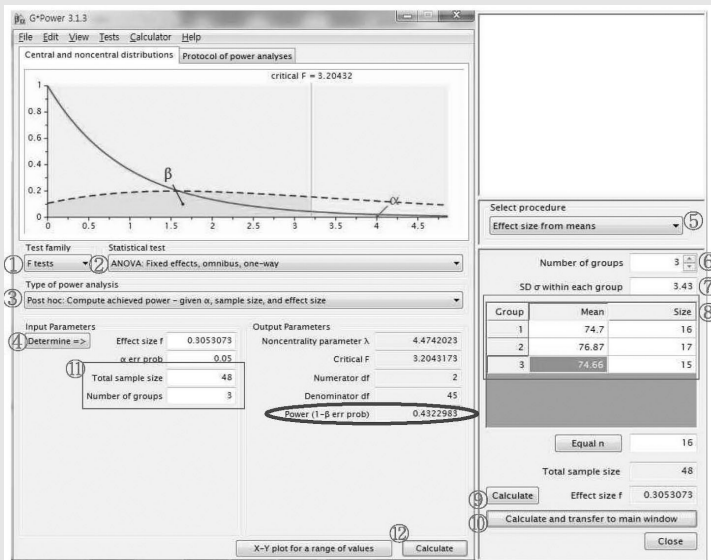


Fig. 6. Power analysis based on one-way ANOVA

- ① Test family에서 F tests를 선택한다.
- ② One ANOVA를 선택하기 위해 ANOVA: Fixed effects, omnibus, one-way를 선택한다.
- ③ 검정력을 계산하기 위해 Post hoc을 선택한다. sample size를 계산하려면 A priori를 선택하면 된다.
- ④ Determine을 눌러 효과크기를 계산한다.
- ⑤ 옆의 윈도우로 옮겨 Select procedure에서 Effect size from means를 선택한다.
- ⑥ 그룹의 수를 입력한다.
- ⑦ 그룹 내의 표준편차를 아래와 같이 계산하여 입력한다.

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_k - 1)s_k^2}{N - k}} \quad (\text{pooled estimate})$$

$n_1, n_2, \dots, n_k$  : 각 그룹의 표본 크기

$s_1, s_2, \dots, s_k$  : 각 그룹의 표준편차

$N = n_1 + n_2 + \dots + n_k$  : total sample size,  $k$  = 그룹 수

- ⑧ 각 그룹의 평균과 표본 크기를 입력한다.
- ⑨ Calculate 버튼을 누르면 total sample size와 효과크기가 계산된다.
- ⑩ Calculate and transfer to main window를 눌러 계산된 효과크기를 옆의 윈도우에 있는 Effect size d로 옮긴다.
- ⑪ Total sample size와 그룹의 수를 입력한다.
- ⑫ Calculate 버튼을 눌러 Power(검정력)를 계산한다.

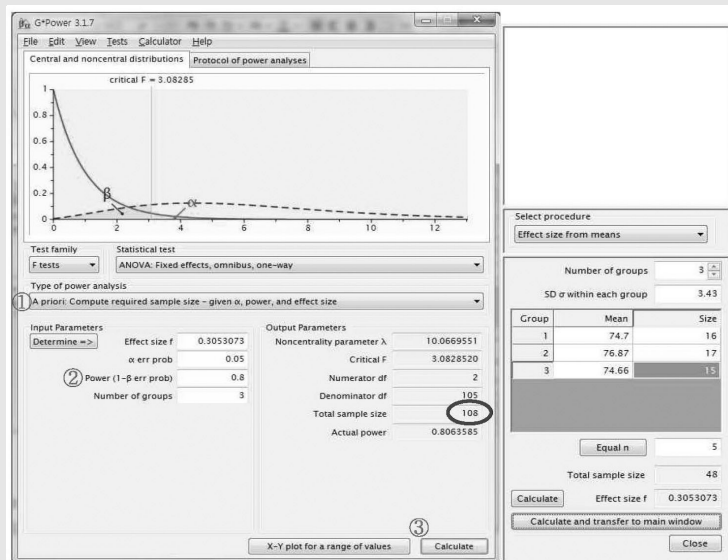


Fig. 7. Sample size determination based on one-way ANOVA

- ① Figure6에서 Type of power analysis에서 Post hoc을 A priori로 바꾼다.
- ② Power를 0.8로 바꾼다.
- ③ Calculate 버튼을 눌러 Total sample size를 구한다.(그룹 당 108/3=36명이 필요)

2011년 Oral Surg Oral Med Oral Pathol Oral Radiol Endod 저널에 출판된 Kim 등의 연구<sup>5)</sup>는 26명의 환자를 대상으로 CBCT를 이용하여 수술 전(T0), 수술 후(T1), 수술 후 6개월(T2)의 condylar axis와 anteroposterior condylar position을 비교하는 것이다.

반복측정 분산분석의 경우 선행 연구에 의해 데이터

가 존재하지 않는다면 표본크기나 검정력을 계산하기 어렵다. 아래 Table 5의 검정력과 표본크기를 계산하기 위해 테이블의 평균과 표준편차를 가지고 T0, T1, T2의 상관계수를 가상으로 설정하여 데이터를 제너레이션 해서 SPSS를 돌려 아래와 같은 결과를 얻었다.(Figure 8, Figure 9) Mauchly의 구형성 검정 결과 구형성을 만족하여 Nonsphericity

Table 5. Power and sample size determination based on the Table2 of the study<sup>5)</sup> that used one-way repeated-measures ANOVA

	Presurgery (T0)	Postsurgery (T1)	Follow-up (T2)	P-value	Power*	Min N*	Final N <sup>§</sup>
Axial condylar angle (right)	74.90 ± 6.88a	68.58 ± 7.46b	68.68 ± 7.16b	.048*	<b>0.9986</b>	<b>12</b>	<b>14</b>
Axial condylar angle (left)	73.24 ± 5.27a	67.44 ± 6.40b	67.13 ± 6.09b	.020*	<b>0.9995</b>	<b>11</b>	<b>13</b>
Anterior space (right)	1.81 ± 0.73a	2.70 ± 0.57	1.88 ± 0.42a	.043*	<b>0.9338</b>	<b>20</b>	<b>23</b>
Superior space (right)	2.67 ± 0.79	2.28 ± 0.64	2.42 ± 0.51	.464	<b>0.2076</b>	<b>99</b>	<b>110</b>
Posterior space (right)	2.43 ± 0.65	68.91 ± 4.66	2.25 ± 0.46	.703	<b>0.3181</b>	<b>66</b>	<b>74</b>

Distances are given in mm, angles in degrees.

FH, Frankfort horizontal.

\*Significant difference in the 3 groups by analysis of variance ( $P < .05$ ).

a,b,c The same superscripts indicate no statistically significant differences among the indicated groups ( $P > .05$ ).

+Power; probability that it rejects a false null hypothesis.

≠Min N; minimum sample size calculated by G\*Power

§Final N; Min N considering the drop rate 10%

correlation 엡실런은 1을 얻었다. 만약에 구형성을 만족하지 않는다면 Greenhouse-Geisser 엡실런 값을 선택한다. 또한 G\*power의 From variances의 Variance explained by effect와 Error variance는 SPSS 결과인 개체 내 효과 검정 테이블의 SS<sub>treat</sub>와 SSE의 값으로 얻었다. Number of Groups는 T0, T1, T2, 3개의 타임 포인트를 말하므로 3이며 Number of measurements는 각 타임 포인트 당 26명의 환자가 있으므로 26이며 Total sample size는 26\*3=78이다. Axial condylar angle과 Anterior space(right) 변수는 모두 p값이 0.05보다 작고 0.8이상의 검정력을 가져 충분한 표본크기가 확보되었음을 알 수 있었고 Superior

space(right)와 Posterior space(right)에서의 유의한 차이를 발견하려면 더 많은 표본크기를 확보해야 한다. 하지만 이 변수들이 주요 결과 변수가 아니라면 더 이상 표본크기를 확보할 필요가 없을 것이다.(Table 5)

#### IV. 결론

임상적, 과학적 근거 없이 표본크기를 30명으로 결정하여 임상시험을 수행하는 것은 윤리적으로 과학적으로 타당하지 않다. 표본 크기가 증가하면 항상 통계적으로 유의한 결과가 나온다<sup>7)</sup>. 통계적 유의성은 표본

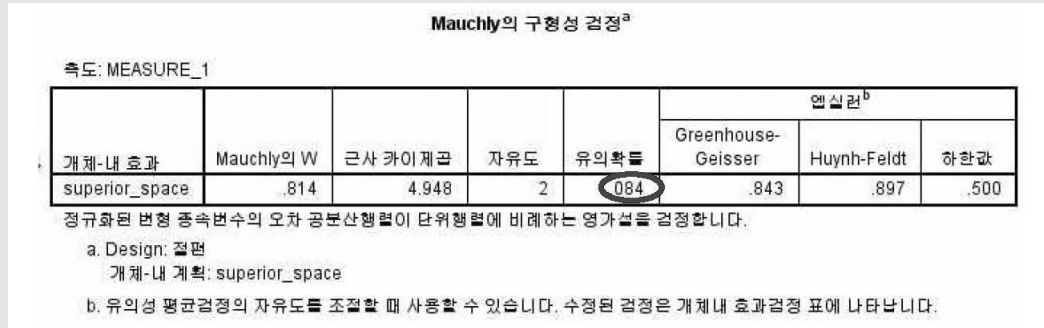


Fig. 8. Mauchly's sphericity test in SPSS output

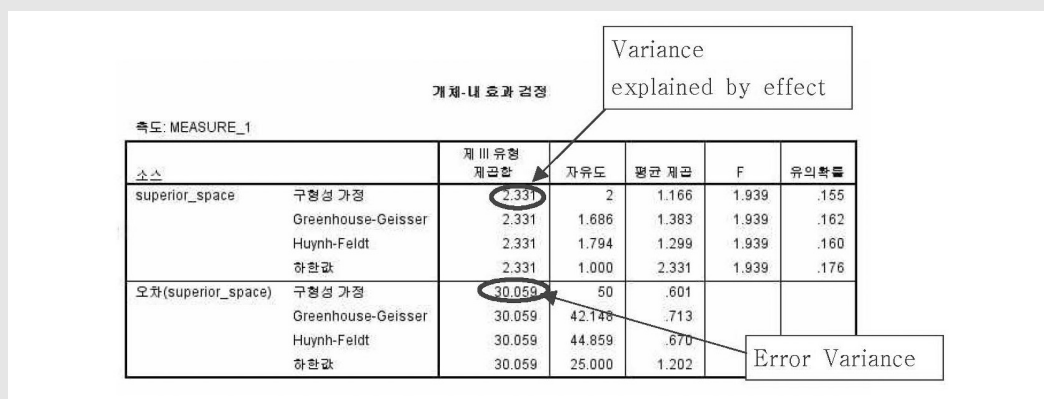


Fig. 9. Test of Within-Subjects Effects in SPSS output

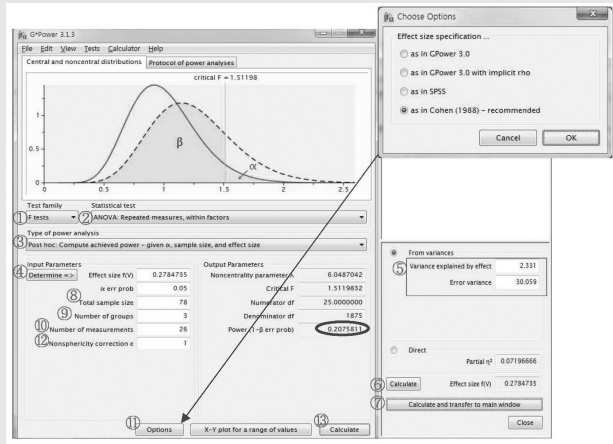


Fig. 10. Power analysis based on one-way repeated-measures ANOVA

- ① Test family에서 F tests를 선택한다.
- ② One RM ANOVA를 선택하기 위해 ANOVA: Repeated measures, within factors를 선택한다.
- ③ 검정력을 계산하기 위해 Post hoc을 선택한다. sample size를 계산하려면 A priori를 선택하면 된다.
- ④ 효과크기를 계산하기 위해 Determine을 누른다.
- ⑤ 옆의 윈도우로 옮겨 From variances에서 Variance explained by effect에는 SSTrt, Error variance에는 SSE를 SPSS output을 보고 입력한다.
- ⑥ Calculate를 눌러 Effect size를 계산한다.
- ⑦ Calculate and transfer to main window를 눌러 Effect size를 main window에 복사한다.
- ⑧ Total sample size를 입력한다.
- ⑨ 그룹의 수를 입력한다.
- ⑩ 각 그룹의 measurement의 수를 입력한다.
- ⑪ Options 버튼을 눌러 as in Cohen(1988)-recommend를 선택하고 OK 버튼을 누른다.
- ⑫ 구형성을 만족하므로 Nonsphericity correction  $\epsilon$ 값 1을 입력한다.
- ⑬ Calculate 버튼을 눌러 Power(검정력)를 계산한다.

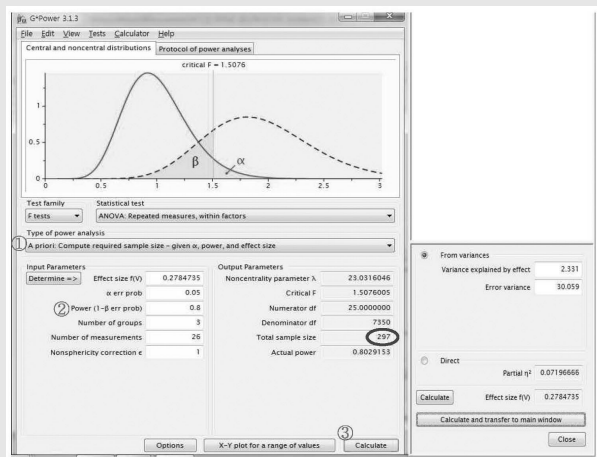


Fig. 11. Sample size determination based on one-way repeated-measures ANOVA

- ① Figure10에서 Type of power analysis에서 Post hoc을 A priori로 바꾼다.
- ② Power를 0.8로 바꾼다.
- ③ Calculate 버튼을 눌러 그룹 별 sample size를 구한다.(297/3=99 per time point)

크기에 매우 영향을 받기 때문이다. 예를 들어 매우 작은 효과크기의 차이를 통계적으로 유의하게 찾기 위해서는 10,000명이 필요하지만 비교적 큰 효과크기의 유의한 차이를 발견하기 위해서는 30명의 표본크기만으로도 충분하다. 그러나 통계적 유의차가 나오지 않는 경우가 두 가지가 있는데 첫째, 표본수가 부족하거나 둘째, 기대했던 효과크기보다 실제의 효과크기가 더 작은 경우이다. 기대했던 효과크기를 가지고 계산된 표본크기로는 유의한 차이를 발견해 내지 못한 것이다. 이 경우에도 표본크기를 늘리면 유의한 차이를

발견해 낼 수 있는데 유의한 차이가 발견되었다 하더라도 실제의 효과크기가 너무 작아 임상적으로 어떤 의미도 찾을 수 없다. 결국 가장 중요한 것은 통계적 유의성을 찾기 보다는 얼마의 효과크기가 임상적으로 의미 있는가를 생각하고 연구하는 것이 더 중요하다고 할 수 있다. 그러므로 높은 질의 연구를 수행하기 위해선 연구 계획 단계에서 최소한의 임상적으로 의미있는 효과크기를 알아내어 유의한 차이를 발견해 낼 수 있는 표본수를 결정하고 연구를 시작하는 것이다.

## 참 고 문 헌

- Miyawaki S, Katayama A, Tanimoto Y, Araki Y, Fujii A, Yashiro K, Takano-Yamamoto T. Salivary flow rates during relaxing, clenching, and chewing-like movement with maxillary occlusal splints. *Am J Orthod Dentofacial Orthop* 2004;126:367-70
- Benson PE, Shah AA, Willmot DR. Measurement of White Lesions Surrounding Orthodontic Brackets: Captured Slides Vs Digital Camera Images. *Angle Orthod* 2005;75:222-226
- 김지형. 샘플 수의 계산. 2012. 대한의학 서적
- Asian BI, Kucukkaraca E, Turkoz C, Dincer M. Treatment effects of the Forsus Fatigue Resistant Device used with miniscrew anchorage. *Angle Orthod* 2014;84:76-87
- Kim YI, Cho BH, Jung YH, Son WS, Park SB. Cone-beam computerized tomography evaluation of condylar changes and stability following two-jaw surgery: Le Fort I osteotomy and mandibular setback surgery with rigid fixation. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 2011;111(6):681-7.
- 임희정. SPSS와 G\*power를 활용한 통계학 개론 및 응용. 2014. 전남대 출판부
- Suresh KP, Chandrashekar S. Sample size estimation and power analysis for clinical research studies. *J Hum Reprod Sci* 2012;5(1):7-13.