

## Original Article

## Comparing Two Formulas of Sample Size Determination for Prevalence Studies

Hamed Tabesh<sup>1</sup> \*Azadeh Saki<sup>2</sup> Fatemeh Pourmotahari<sup>3</sup>

1-Assistant Professor of Biostatistics, Department of Biostatistics and Epidemiology, School of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

2-Assistant Professor of Biostatistics, Department of Biostatistics and Epidemiology, School of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

3- MSc Student of Biostatistics, Department of Biostatistics and Epidemiology, School of Health, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran

\* azadehsaki@yahoo.com

### Abstract

**Background and purpose:** Sample size and its determination is one of the most important problems in health researches. Calculating sample size for prevalence studies is one of the common questions of sample size topics. Minimum sample size with least complexity is desirable in order to achieve the basic goal of these studies. This study aims to compare two formulas of sample size calculation for prevalence researches and finally, to use the simplest formula to get the most appropriate sample size.

**Materials and Methods:** Sample size for proportions: 0.9, 0.95, 0.99, 0.999 candidates of  $p$  close to 1 proportions;  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$ , 0.05, 0.1 candidates of  $p$  close to 0, and proportions 0.3, 0.4, 0.5, 0.6, 0.7 candidates of  $p$  close to 0.5 were calculated. For comparing  $n_1$ ,  $n_2$ ;  $\varphi = n_1/n_2$ , it was computed by R package (2.10.1).

**Results:** Computed sample size by  $(f_2)$  is lightly greater than sample size computed by  $(f_1)$  and maximum value of  $\varphi$  index for comparing the two formulas equals 1.

**Conclusion:** Results show that the calculated sample size by  $(f_1)$  is similar to what was obtained by  $(f_2)$ , though, according to its interpretation and easy computation, it is suggested for all values of  $p$ . [Tabesh H. \* Saki A. Pourmotahari F. Comparing two Formulas of Sample Size Determination for prevalence studies. *IJHS* 2013; 1(2):56-60] <http://jhs.mazums.ac.ir>

**Key words:** Sample Size Calculation, Prevalence Study, Calculation Procedure.

## 1. Introduction

If observational and experimental studies are designed effectively, valuable results would be obtained. Good planning has many aspects such as exact definition of problem and method and enough sample size due to the goals. Enough sample size for a research is determined based on the type and object of a study, statistical methods for analyzing and interpreting, available data, validity and reliability for the generalized results by two general techniques: confidence interval and Bayesian methods (1). In simple term, sample size estimation means to estimate the minimum number of the sample for a study by using statistical methods based on specific situation, basic information and precision requirement, and under the premise of guaranteeing the reliability of the conclusion.(2) Calculated sample size should be large enough so that an effort of such magnitude to be of scientific significance & also be statistically significant. It is just as important, however, that the study not to be “too large”, where an effect of little scientific importance is nevertheless statistically detectable (3). Although, for such an important issue, there is not a large amount of published literature, there are several approaches for sample size. For example, one can specify the desired width of a confidence interval and determine the sample size that achieves that goal or Bayesian approach can be used where we optimize some utility function. For conducting sample size, statistical inference is based on estimating a parameter or testing a hypothesis which demand one type of the study. When estimating an infinite population parameter such as prevalence, incidence and chance of illness

recurrence is desirable. Several text books (4-7) recommended using (f<sub>1</sub>) and (f<sub>2</sub>).

$$n_1 = \frac{Z_{\alpha/2}^2 \hat{p} (1-\hat{p})}{d^2} \quad (f_1)$$

$$n_2 = \left( \frac{Z_{\alpha/2}^2}{\sin^{-1}\left(\frac{d}{\sqrt{\hat{p}(1-\hat{p})}}\right)} \right)^2 \quad (f_2)$$

Where n<sub>1</sub>, n<sub>2</sub> are the estimated minimum sample size d and d are the desired level of precision and estimated proportion of an attribute present in the population. It is clear that z<sup>2</sup> is the abscissa of the normal curve that cuts off an area α at the fails. (1- α equals the desired confidence level). Simplicity is one of the greatest properties of each statistical approach. Denominator of (f<sub>1</sub>) equals a half width of confidence interval which shows the estimation precision. So it is easy to understand and interpret. Also calculating sample size by (f<sub>1</sub>) is simple and does not have any complexity. But unlike the denominator of (f<sub>2</sub>), to the best of our knowledge, there is not any interpretation and computing sample size by (f<sub>2</sub>) which is not as convenient as computing by (f<sub>1</sub>). some literatures recommended using (f<sub>1</sub>) when p is close to 0.5 and (f<sub>2</sub>) when p is close to 0 or 1(6,8-9). In this study, comparing results of using (f<sub>1</sub>) and (f<sub>2</sub>) in the same situation was desirable so if possible, an alternative formula for (f<sub>2</sub>) could be suggested.

## 2. Materials and Methods

As mentioned before, several literatures recommended using  $(f_2)$  when crude estimation of prevalence, incidence, and success proportion or illness recurrence probability is close to 0 or 1. In this study, to compare the outputs of  $(f_1)$  and  $(f_2)$  sample size for proportions: 0.9, 0.95, 0.99, 0.999 as candidates of  $p$  close to 1 and,  $10^{-5}$ ,  $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$ , 0.05, 0.1 as candidates of  $p$  close to 0 and 0.3, 0.4, 0.5, 0.6, 0.7 as candidates of  $p$  close to 0.5 considered. Since  $d < K \cdot \min\{p, 1-p\}$  where  $K < \frac{1}{4}$ . So  $k$  equal to 10%, 15%, 20%, and 25% have been considered. Whereas softwares for sample size determination such as PASS, UnifyPow and Power and Precision determine sample size based on  $(f_1)$  so for comparing  $(f_2)$ , package R (2.10.1) was used and  $\varphi$  index ( $\varphi = n_1/n_2$ ) was

computed (10-12). As the most common confidence intervals in medical research cases are 95% and 99%, sample size was computed for these confidence intervals.

## 3. Results

Sample size for different estimated proportions in population,  $p$ , and distinct values of  $k$ , was computed by  $(f_1)$  and  $(f_2)$ .  $N_1$ ,  $n_2$  are estimated minimum sample size by  $(f_1)$  and  $(f_2)$  respectively. For comparing  $n_1$ ,  $n_2$ , index  $\varphi = n_1/n_2$  was calculated. Computed  $\varphi$  for  $p$  close to 0, 1 and 0.5 with 0.95 confidence limit are shown in tables (1), (2) and (3) respectively. The findings of this study show that  $n_1$  and  $n_2$  were very close to each other when estimated proportion of an attribute present in the population was very small.

**Table1.** Calculated  $\varphi$  index by  $(f_1)$  and  $(f_2)$  for very small values of  $\hat{p}$  when confidence limit is 95%

$\frac{\hat{p}}{K}$	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-2}$	0.05	0.10
10%	1.000000	1.000000	1.000003	1.000034	1.000175	1.000371
15%	1.000000	1.000001	1.000008	1.000076	1.000395	1.000834
20%	1.000000	1.000001	1.000013	1.000135	1.000703	1.001485
25%	1.000000	1.000002	1.000021	1.000211	1.001098	1.002323

**Table2.** Calculated  $\varphi$  index by  $(f_1)$  and  $(f_2)$  for very large values of  $\hat{p}$  when confidence limit is 95%

$\frac{\hat{p}}{K}$	0.9	0.95	0.99	0.995	0.999
10%	1.000371	1.000175	1.000034	1.000017	1.000003
15%	1.000834	1.000395	1.000076	1.000076	1.000008
20%	1.001485	1.000703	1.000135	1.000067	1.000013
25%	1.002323	1.001098	1.000211	1.000105	1.000021

**Table3.** Calculated  $\varphi$  index by (f<sub>1</sub>) and (f<sub>2</sub>) for mid- values of  $\hat{p}$  when confidence limit is 95%

$\hat{p}$ K	0.3	0.4	0.5	0.6	0.7
10%	1.001432	1.002230	1.003351	1.002230	1.001432
15%	1.003231	1.005040	1.007591	1.005040	1.003231
20%	1.005767	1.009018	1.013625	1.009018	1.005767
25%	1.009058	1.014206	1.021557	1.014206	1.009058

When  $p$  was close to 1,  $\varphi$  index was approximately equal to 1 which means that  $n_1$  and  $n_2$  were similar. For mid-values of  $p$ , similar conclusion could be drawn. The results revealed that both formulas, (f<sub>1</sub>) and (f<sub>2</sub>), would perform similarly when the estimated proportion of an attribute present in the population had very small, medium and very large values. The computations were done for 99% confidence limit, and the results were closely similar.

#### 4. Discussion

In medical research, it is important to determine size sufficiently enough to ensure reliable conclusions. On the other hand, prevalence studies are interesting research cases in medical sciences and consequently, adequate sample size for these research cases would be interesting, too. The most common formulas for prevalence or incidence studies are (f<sub>1</sub>) and (f<sub>2</sub>). (5-7). (f<sub>1</sub>) has simple structure made up of (f<sub>2</sub>) so it has more public

interest than (f<sub>2</sub>). But some literatures limited using (f<sub>1</sub>). They believe that (f<sub>1</sub>) could be useful when  $p$  is not to be close to 1 or 0. This study showed that the estimated sample size by (f<sub>1</sub>) is approximately similar to the estimated sample size by (f<sub>2</sub>). On the other hand, computing sample size by (f<sub>1</sub>) is easy. Therefore, using (f<sub>1</sub>) is recommended for estimating sample size of prevalence or incidence studies for infinite population

#### References

1. Lenth RV. Some practical Guidelines Effective sample size Determination. The American Statistical, 2001; 55(3): 187-193.
2. HU LP, Bao XL, Zhou SG, Guas X, Estimation of sample size and testing power (part1). J chin Integr Med. 2011;9(10): 1070-1074.
3. Sathian B, sreedharan y, Baboo NS, Sharan k, Abhilash ES, Rajesh E. Relevance of sample size Determination in Medical Research., Nepal Journal of Epidemiology, 2010;1(1): 4-10.

4. Daniel WW. Biostatistics: A Foundation for Analysis in the Health Sciences. 7th edition. New York: John Wiley & Sons.
5. Lwanga SK and Lemeshow S. Sample size determination in Health Studies: A practical Manual. Geneva: World Health Organization. 1999.
6. Cochran WG. Sampling Techniques, 3<sup>rd</sup> edition. New York: John Wiley & Sons. 1997.
7. Naing L, Winn T and Rusli BN. Sample size calculator for prevalence studies. Archives of Orofacial Sciences, 2006;1:9-14
8. Kreamer, H.C, and Thiemann, S. How many subjects & statistical power Analysis in Research, Newbury park, CA: Sage Publications. 1987.
9. Desu MM, Raghavao D. sample size methodology. Boston, MA: Academic press, Inc. 1990
10. Hintze, J. PASS 2000, Kaysville, UT: Number Cruncher Statistical System, Software for MS-DOS systems. 2001
11. O'Brien R.G. Unifypow.sas: version 98.08.25, Department of Biostatistics and Epidemiology, Cleveland Clinic Foundation, Cleveland, OH. 1998.
12. Borenstein, M., Rothstein, H., and Cohen, J. Power and Precision, Biostat, Teaneck, NY: software for MS-DOS systems. 1997.