Aceh International Journal of Science and Technology, 1 (2): 51-53
August 2012
ISSN: 2088-9860

SHORT COMMUNICATION

Some Practical Guidelines for Effective Sample-Size Determination in Observational Studies

[1]Wan Muhamad Amir W Ahmad, [2]Wan Abdul Aziz Wan Mohd Amin, [3]Nor Azlida Aleng, and [4]Norizan Mohamed

[1, 3, 4]Department of Mathematics, Faculty of Science and Technology, [2]Department of Language and Communication, Universiti Malaysia Terengganu (UMT)
[1]wmamir@umt.edu.my, [2]ziza@umt.edu.my, [4]norizan@umt.edu.my, [3]azlida_aleng@umt.edu.my

Abstract- The purpose of this present study was to determine the right sample size in observational study which is focusing on medical or health sciences field. Sample size calculation is actually depends on the type of how the study is designed. For example different formulas are used to calculate the sample size in different type of the study. In this article, the formula of single proportion and two proportions is discussed and an example from medical research, which may contribute to the understanding of this problem, is presented.

Keywords: Sample size determination, single proportion, and two proportions.

In a research study, most of researchers design experiments with dozens of treatment combinations. An experiment that they had design can involve several of measurement include of dichotomous variables. Sample size calculations for dichotomous variables do not require knowledge of any standard deviation. Statistical analysis is based on the key idea that observation on a sample of subjects is made and then draws inferences about the population from which the sample is drawn (Naing, 2003). If the study sample is not representative of the population, it will well mislead and statistical procedure cannot help. However, even a well-designed study can give only an idea of the answer sought because of random variation in the sample. Thus results from a single sample are subject to statistical uncertainty which is strongly related to the size of the sample.

Quality or strength of statistical inference depends largely on the size of the sample selected (Naing, 2003; Gardner and Altman, 1989). In this regard, there are some concerns about determining adequate sample size for studies where particular parameters of populations are estimated. In this study, we emphasize the parameter which is based on proportion of a population. For examples we want to estimate the prevalence of HIV-Infected Tuberculosis disease in a population.

According to Naing (2003) an appropriate inclusion and exclusion criteria are also important to obtain participants who have particular characteristics which are under research interest. The other area, which cannot be simply ignored, is testing validity and reliability of measurement tools before actual data collection. A pilot study or a pretest is required to test the validity and reliability of measurement tools. In addition, there are a number of potential biases which can be avoided while designing study and during planning of data collection. The following is the brief guide to approach on how to determine a sample size based on single proportion and two proportions. For example, we want to conduct a study on "Associated Factors of HIV-Infected Tuberculosis in Hospital Universiti Sains Malaysia, HUSM"**.** The aim of this study is to determine the prevalence of TB among HIV patients and associated factors of TB among HIV positive patients.

The following steps are the brief guide to determine a sample size based on parameter which is 'proportion' (Dupont and Plummer, 1990; Lwanga and Lemeshow, 1991). To illustrate this case, we fixed objectives as shown below. From the fixed objectives, we can determine the right sample size. Specific objectives were:

1. To determine the prevalence of TB among people living with the HIV from 2001 to 2010 in HUSM.
2. To identify associated factors of TB among HIV positive patients.

For objective 1, we used single proportion method. To determine the sample size required to estimate the proportion with the desired level of precision, some idea was required beforehand about the possible magnitude of the proportion. It can be obtained from previous studies in literature or from a pilot study if there was no similar study conducted before. This formula is usually used in prevalence studies.

$$\text{Sample size, } n = \left(\frac{z}{\Delta}\right)^2 p(1-p)$$

Calculation was done on proportion of TB in people living with HIV patients.

$$\text{Sample size, } n = \left(\frac{1.96}{0.05}\right)^2 0.072\ (1-0.072)$$

$$n = 103 \quad \text{Patients}$$

where

z = 1.96
Level of Significance, α = 0.05
Absolute precision (Δ) = 5%
p is expected proportion of individual in the sample with the characteristic of interest at the 100(1-α) % confidence interval. Anticipated population proportion (p) = 7.2% (Gao *et al*., 2010)

After adding 20% estimated missing data, we finally decided that total patients to be sampled is $n = 123.6 \approx 124$ patients. So, total sample size needed n = 124 of TB in people living with HIV patients.

For objective 2, we used two proportions method. The calculation of sample size can be done using Power and Sample Size Calculation (PS) software or manual calculation, with the significance level (α) 0.05 and the power of study (1-β) of 80%. The detectable hazard ratio of the presence of prognostic factor relative to absence of prognostic factors was decided by the researcher and expert opinion by clinicians (Dupont and Plummer, 1997, Mugusi *et al*., 2009). Below is the Two Proportions formula:

$$n = \frac{p_0(1 - p_0) + p_1(1 - p_1)}{(p_0 - p_1)^2}(z_\alpha + z_\beta)^2$$

where:

P_0 = Based on literature review
P_1 = Based on expert opinion
$z_\alpha = z_{0.05} = 1.9600$ (one tailed)
$z_\beta = z_{0.20} = 0.8416$ (one tailed)

Table 1. Sample size calculation

No	Objective	*P_1	P_0	Type 1 error	Power	Sample Size
1	CD4 Cell count (Nissapatorn *et al*., 2003)	0.52	0.36	5%	80%	142 patients
	Calculation $n = \frac{0.36(1-0.36)+0.52(1-0.52)}{(0.36-0.52)^2}(1.96+0.84)^2 = 142$ patients					
2	Heterosexual (Nissapatorn *et al*., 2003)	0.44	0.28	5%	80%	137 patients
	Calculation $n = \frac{0.28(1-0.28)+0.44(1-0.44)}{(0.28-0.44)^2}(1.96+0.84)^2 = 137$ patients					
3	Intravenous drug user (IVDU) (Nissapatorn *et al*., 2003)	0.56	0.4	5%	80%	149 patients
	Calculation $n = \frac{0.4(1-0.4)+0.56(1-0.56)}{(0.4-0.56)^2}(1.96+0.84)^2 = 149$ patients					
4	Blood transfusion (Nissapatorn *et al*., 2003)	0.41	0.25	5%	80%	132 patients
	Calculation $n = \frac{0.41(1-0.41)+0.25(1-0.25)}{(0.25-0.41)^2}(1.96+0.84)^2 = 132$ patients					
5	Age (Nissapatorn *et al*., 2003)	0.53	0.37	5%	80%	148 patients
	Calculation $n = \frac{0.37(1-0.37)+0.53(1-0.53)}{(0.37-0.53)^2}(1.96+0.84)^2 = 148$ patients					

**Determined by expert opinion*

From Table 1 we can see that, the largest sample size is taking into the account is n =149 patients. In this case, we consider 149 patients (because we choose the biggest sample size). After adding 20% estimated missing data, we get $n = 149 + (0.2 \times 149) = 178.8 \approx 179$ per group, which can be obtained as follows:

i. Patients with TB among HIV patients = 179 patients
ii. Patients with Non-TB among HIV patients = 179 patients

Therefore, a total patient to be sampled is $(179 \times 2) = 358$ patients.

The application of the proposed formula for the sample size determination has been discussed in this paper with approaching to the case of determination associated factors of HIV-infected Tuberculosis. Before we proceed the study, we

have to design the study which is including the sample size determination. This kind of study has been conducted in many parts of the world. In this paper, two different methods have been used: (i) Single proportion and (ii) Two proportions. Naing (2003) pointed out that the statistical importance of required sample size, to be earmarked as a standard value for a particular population, was rarely emphasized though values have been recognized as references for ages. Health personnel need to be aware of the fact that central to the planning of any such study is the decision on how large a sample to select from the population under study. This article is discussed a certain extension of important issue of sample size determination for observational studies. Ahmad *et al.* (2011) pointed out that the larger your sample, the more sure you can be that their answers truly reflect the population. However, there are a few guidelines that can have to be addressed in this particular area of health sciences.

References

Dupont, W. D., and Plummer, W.D. (1990). Power and sample size calculation, A review and computer program. Controlled Clinical Trails, 11: 116-128.

Mugusi, F. M., Mehta, S., Villamor, E., Urassa, W., Saathoff, E., Bosch, R. J., and Fawzi, W. W. (2009). Factors associated with mortality in HIV-infected and uninfected patients with pulmonary tuberculosis. BMC Public Health, 9: 409.

Gardner, M. J., and Altman D.G. (1989). Statistics with confidence-confidence intervals and statistical guidelines. British Medical Journal: 1-13.

Nissapatorn, V., Lee, C., Ithoi, I., Fong, M.Y., Abdullah, K. A. (2003). Tuberculosis in aids patients. Malaysian Journal of Medical Sciences, 10: 60-64.

Gao, L., Zhou, F., Li, X. W., and Jin, Q. (2010). HIV/TB co-infection in mainland China: A Meta-Analysis. Plos One Journal, 5(5): 1-6.

Lwanga, S. K., and Lemeshow, S. (1991). Sample size determination in health studies: A practical manual. World Health Organization, Geneva.

Ahmad, W.M.A.W., Aleng, N. A, Ali, Z., and Nawi, A.A. (2011). Modeling associated factors of HIV-infected tuberculosis (TB) patients using path model analysis. World Applied Sciences Journal, 12(9): 1584-1584.

Naing, N. N. (2003). Determination of sample size. Malaysian Journal of Medical Science, 10(2): 84-86.