Determination of Minimum Sample Size Requirement for Multiple Linear Regression and Analysis of Covariance Based on Experimental and Non-experimental Studies

Mohamad Adam Bujang [1], Nadiah Sa'at [2], Tg Mohd Ikhwan Tg Abu Bakar Sidik [2]

(1) Clinical Research Centre, Sarawak General Hospital,Malaysia
(2) Biostatistics Unit, National Clinical Research Centre, Malaysia

CORRESPONDING AUTHOR: *Mohamad Adam Bujang - Clinical Research Centre, Sarawak General Hospital, Jalan Tun Ahmad Zaidi Adruce, 93586 Kuching, Sarawak, Malaysia. Office: 6082 – 276823 - Fax: 6082 – 276823 - E-mail address: adam@crc.gov.my*

DOI: 10.2427/12117
Accepted on June 13, 2017

ABSTRACT

Background: MLR and ANCOVA are common statistical techniques and are used for both experimental and non-experimental studies. However, both types of study designs may require different basis of sample size requirement. Therefore, this study aims to proposed sample size guidelines for MLR and ANCOVA for both experimental and non-experimental studies.
Methods: We estimated the minimum sample sizes required for MLR and ANCOVA by using Power and Sample Size software (PASS) based on the pre-specified values of alpha, power and effect size (R^2). In addition, we also performed validation of the estimates using a real clinical data to evaluate how close the approximations of selected statistics which were derived from the samples were to the actual parameters in the targeted populations. All the coefficients, effect sizes and r-squared obtained from the sample were then compared with their respective parameters in the population.
Results: Smaller minimum sample sizes required for performing both MLR and ANCOVA when r-squared is used as the effect size. However, the validation results based on an evaluation from a real-life dataset suggest that a minimum sample size of 300 or more is necessary to generate a close approximation of estimates with the parameters in the population.
Conclusion: We proposed sample size calculation when r-squared is used as an effect size is more suitable for experimental studies. However, taking a larger sample size such as 300 or more is necessary for clinical survey that is conducted in a non-experimental manner.

Key words: Analysis of Covariance; coefficients; Multiple Linear Regression; r-squared; sample size

INTRODUCTION

Multiple Linear Regression (MLR) and Analysis of Covariance (ANCOVA) are the two common statistical analyses in multivariate models. These two statistical tools share several common assumptions, however they are usually applied in different scenarios. MLR is usually been applied as a statistical tool to predict the event of dependent variable based on a set of predictors [1-2]. On the other hand, ANCOVA, is more commonly used when the research question aims to determine the effect of an independent variable on an outcome after certain variable(s) is/are being adjusted in the analysis [3]. This technique is usually applied when data was collected in a non-experimental manner, such as for cross-sectional or cohort study designs. Besides that, ANCOVA is also applied for determining the presence of an association between a set of risk factors and an outcome [4].

A major pre-requisite for using MLR and ANCOVA is to determine the minimum required sample size. Conventionally, the minimum required sample size for almost all types of multivariate analysis is determined using a rule-of-thumb which is mostly derived from MLR. Some sample size guidelines proposed a minimum required sample size based on ratio between number of independent variables and number of case such as 30 to 1 [5] and 10 to 1 [6]. Tabachnick and Fidell (2013) proposed by using formula of "50 + 8m" where "m" is the number of factor [7]. Gregory and Daniel (2008) proposed a guideline for determining the minimum required sample size for using MLR for prediction [8]. They proposed that the sample should vary from small to large, according to the effect sizes. With regards to ANCOVA, George et. al. (2007) recommended a simple formula to estimate the minimum required sample size for a randomized controlled trial study [9].

Nowadays, it is feasible to use software for determining the minimum sample size required for performing multivariate analysis. Such a software will be very useful for those who are familiar with selecting the correct statistical technique for a particular research design, and also in the application of a particular statistical software for the same purpose. These are usually the statisticians, who will be able to determine the minimum required sample sizes by themselves. However, researchers with a different background may require a simple guide to estimate a minimum required sample size for their research. This is why one of the aims of this study is to tabulate the minimum sample sizes required for using both MLR and ANCOVA using sample size software.

In addition, a validation study using real patient data was conducted to evaluate to what extent all the different sample sizes used can affect the discrepancy between the sample statistics and the actual parameters in the target population. The purpose of this validation is to estimate a minimum sample size required for a research study which is able to derive a closest estimate for the coefficients and also r-squared. This is to indicate that the results inferred from the sample are able to be generalized to those from the target population.

The combination of findings from the minimum required sample sizes which were estimated by using the Power and Sample Size software (PASS) and the results obtained from the validation (to evaluate to what extent all the different sample sizes used for prediction by either MLR and ANCOVA can affect the discrepancy between the sample statistics and the actual parameters in the target population) were used as the basis for sample size recommendation for both MLR and ANCOVA.

METHODS

The sample size calculation was performed using Power and Sample Size (PASS) Software (PASS 11 citation: Hintze, J. (2011). PASS 11. NCSS, LLC. Kaysville, Utah, USA). When using this Power and Sample Size (PASS) Software, it is necessary to select "Regression" as the module, the values for both alpha and desired power were set at 0.05 and 0.8 respectively. Another parameters required for this calculation are the value of r-squared of "number of variables tested" define as R^2_T and r-squared of "number of variables controlled" define as R^2_C. To tabulate a range of estimated sample sizes required for MLR, the "number of variables controlled" was set at zero and the "number of variables tested" was set at two, or three, or four, or five, or six, or seven, or eight, or nine, or ten. The effect size was determined by the value of r-squared (R^2_T) and was set at 0.1, or 0.2, or 0.3, or 0.4, or 0.5, or 0.6, or 0.7.

Sample size calculation for ANCOVA was also performed by using the same values for both alpha and desired power. The "number of variables tested" was set at one to indicate that the effect of only one independent variable is being tested on an outcome. The r-squared (R^2_T) was set at 0.1, or 0.2, or 0.3, or 0.4, or 0.5. The effect size for a particular independent variable should have a minimum value of 0.1 and it shall ensure that the multivariate analysis (either MLR or ANCOVA) will be able to detect a significant effect size of up to 0.5. The "number of variables controlled" was set at one, or two, or three, or four or five, or six, or seven, or eight, or nine, or ten. The or effect size for the controlled variable(s) was also set at 0.1, or 0.2, or 0.3, or 0.4, or 0.5.

Besides that, a validation was conducted to verify the accuracy of these estimations. The validation was performed using a real patient data from "An Audit of Diabetes Control and Management (ADCM) 2009", which included all data collection (at a national-level) of patients with diabetes mellitus from all the government health clinics in Malaysia during the year 2009. The methodology of this data collection process was explained in a previous paper and published elsewhere [10]. We selected one government health clinic which have a relatively high number of patients with total population of 1595, therefore re-analysis can be done by using different sub-samples (n=30, 50, 100, 150, 300, 500, and 1000). All these different sub-samples were representing all the different sample sizes that were estimated using the Power and Sample Size software (PASS) which would subsequently be tested using real patient data obtained from an audit.

We tested a multivariate model by using eight explanatory (or independent) variables and one outcome (or dependent variable). The dependent variable was glycemic control (HbA1c) in a numerical form while a set of independent variables include gender, age, body mass index, diabetes treatment, duration of diabetes mellitus, systolic blood pressure, status of co-morbidity and low-density lipoprotein. Since all data were not collected in a prospective fashion, the model developed can only be used to test for an association between these independent variables and the outcome; rather than to identify and determine the risk factors or determinants for HbA1c [11].

Next, the findings obtained from the validation were then analyzed. The statistics such as r-squared, partial-eta-squared and coefficients that derived from the samples were then compared with the respective true values (parameters) in the targeted population. The analyses for the validation were carried out using IBM SPSS version 21.0 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.)

RESULTS

Findings from sample size calculation using PASS

When the values of both alpha and desired power were fixed, the sample size would vary from small to large depending on the number of independent variables and also on their r-squared. The calculated sample sizes required for MLR is presented in Table 1. A larger sample size is required for a smaller value of r-squared and when there are many explanatory (or independent) variables are to be tested in the regression model. Sample size for a value of r-squared which is more than 0.7 is not calculated since it will yield a very small sample size. In general, the minimum sample size required will usually be less than 200 for a maximum of 10 independent variables.

On the other hand, the calculated sample sizes required for ANCOVA are presented in Table 2, Table 3 and Table 4. Similarly, a larger sample size is required for a smaller value of r-squared and when there are many explanatory (or independent) variables are to be controlled in the model. By using the software, it was calculated that the minimum sample size required is less than 100 to test for one independent variable with a maximum of ten controlled variables.

For the same effect size and also the same number of independent variables, it was calculated using the software that the minimum sample size required for ANCOVA was generally larger when compared with that for MLR. For instance, a minimum sample size of 57 (R^2_T = 0.1 for test variable + R^2_C = 0.2 for controlled variables) or a minimum sample size of 30 (R^2_T = 0.2 for test variable + R^2_C=0.1 for controlled variables) will be needed to be able to detect a value of r-squared of 0.3 between two variables, one is a tested variable and the other is a controlled variable. However, a minimum sample size of only 26 is required for MLR to detect a value of r-squared of 0.3 between two independent variables.

TABLE 1. Sample size for Multiple Linear Regressions (MLR) based on number of tested variables with selected R^2_T for 0 controlled variable (Alpha = 0.05 and Power = 0.8)

Number of Tested Variable	R^2_T for Tested Variable						
	0.1	0.2	0.3	0.4	0.5	0.6	0.7
2	90	42	26	18	14	11	9
3	103	48	30	21	16	12	10
4	113	53	33	24	18	14	12
5	122	58	36	26	20	16	13
6	130	62	39	28	21	17	14
7	137	65	42	30	23	18	15
8	144	69	44	32	24	20	17
9	150	72	46	33	26	21	18
10	156	75	48	35	27	22	19

Findings obtained from the validation

The detail of the variable was presented in Table 5 and results obtained from the validation are presented in Figure 1 until Figure 5. Results had shown that having a minimum sample size of 500 and above, it is possible to ensure that the differences between the sample estimates and the population parameters of partial eta-squared, regression coefficients and r-squared to be sufficiently small (i.e. differences less than 0.05). This indicates that a minimum sample size of 500 or more will yield reliable and valid sample estimates for the intended population. A minimum sample size of 300 shows almost a similar result (Figure 1).

DISCUSSIONS

This study provides an estimation of the minimum required sample sizes for performing MLR and ANCOVA for a range of differing effect sizes. All these minimum required sample sizes are tabulated which can serve as a quick guide for researchers especially who are non-statisticians to estimate a minimum required sample size for their studies. Previous studies had already introduced several simple rules of thumb

TABLE 2. Sample size for Analysis of Covariance (ANCOVA) based on number of controlled variables with selected R^2_C for 1 tested variable with $R^2_T = 0.1$, 0.2, and 0.3 (Alpha = 0.05 and Power = 0.8)

R^2_T for Tested Variable	Number of Controlled Variable	R^2_C for Controlled Variable				
		0.1	0.2	0.3	0.4	0.5
0.1	1	65	57	50	42	34
	2	65	58	50	42	34
	3	65	58	50	42	34
	4	65	58	50	42	34
	5	65	58	50	42	34
	6	65	58	50	42	34
	7	66	58	50	42	34
	8	66	58	50	42	35
	9	66	58	50	42	35
	10	66	58	50	42	35
0.2	1	30	26	22	18	15
	2	30	26	22	19	15
	3	30	26	23	19	15
	4	30	27	23	19	15
	5	30	27	23	19	16
	6	31	27	23	19	16
	7	31	27	23	20	17
	8	31	27	23	20	17
	9	31	27	24	20	18
	10	31	27	24	21	18
0.3	1	18	16	13	11	9
	2	19	16	14	11	9
	3	19	16	14	12	10
	4	19	17	14	12	10
	5	19	17	15	13	11
	6	19	17	15	13	12
	7	20	18	16	14	13
	8	20	18	16	15	13
	9	20	18	17	15	14
	10	21	19	17	16	15

[5-7]. Although these are relatively easy to apply, however researchers may require more specific guidelines which are tailored to different situations. It is well-known that a sufficient sample size should not rely on the number of independent variables only. For instance, the findings from the sample size calculation showed that every additional independent variable in the model will not be needed an additional of 10 to 20 cases. It is also necessary to take into consideration the impact of the both effect size and number of independent variables on the minimum required sample size [8].

TABLE 3. Sample size for Analysis of Covariance (ANCOVA) based on number of controlled variables with selected R^2_C for 1 tested variable with $R^2_T = 0.4$ and 0.5 (Alpha = 0.05 and Power = 0.8)

R^2_T for Tested Variable	Number of Controlled Variable	R^2_C for Controlled Variable				
		0.1	0.2	0.3	0.4	0.5
0.4	1	13	11	9	8	6
	2	13	11	10	8	7
	3	13	12	10	9	8
	4	14	12	11	10	8
	5	14	13	11	10	9
	6	15	13	12	11	10
	7	15	14	13	12	11
	8	16	15	14	13	12
	9	16	15	14	14	13
	10	17	16	15	15	14
0.5	1	10	8	7	6	
	2	10	9	8	7	
	3	10	9	8	7	
	4	11	10	9	8	
	5	12	11	10	9	
	6	12	11	11	10	
	7	13	12	12	11	
	8	14	13	13	12	
	9	15	14	13	13	
	10	15	15	14	14	

Based on the findings from sample size calculation, we are concerned with the relatively small minimum sample sizes required for performing both MLR and ANCOVA when r-squared is regarded as an indicator of effect size. Therefore, besides tabulating all the minimum required sample sizes for both MLR and ANCOVA, a validation was also conducted to determine the minimum required sample sizes for yielding very close approximations of the sample estimates for partial eta-squared, regression coefficients and r-squared to their respective population parameters. This is because the ultimate aim for conducting an inferential study is to infer the true value of a target population parameter from a particular sample [12]. Hence, the validation study can serve as a viable research design to address this issue [13-14].

The term validation is used to compare results derived from the samples with the true values in the population (parameters). Since the parameters such as r-squared, effect size and coefficients are already known from the population data, therefore it is not necessary to do inferential study by reporting p-value and respective 95% confidence interval. Instead, the differences between the selected statistics and the respective parameters were

TABLE 4. Sample size for Analysis of Covariance (ANCOVA) based on number of controlled variables with selected R^2_C for 1 tested variable with R^2_T = 0.6 and 0.7 (Alpha = 0.05 and Power = 0.8).

R^2_T for Tested Variable	Number of Controlled Variable	R^2_C for Controlled Variable				
		0.1	0.2	0.3	0.4	0.5
0.6	1	8	7	6		
	2	8	7	6		
	3	9	8	7		
	4	10	9	8		
	5	10	10	9		
	6	11	11	10		
	7	12	11	11		
	8	13	12	12		
	9	14	13	13		
	10	15	14	14		
0.7	1	6	5			
	2	7	6			
	3	8	7			
	4	9	8			
	5	10	9			
	6	10	10			
	7	11	11			
	8	12	12			
	9	13	13			
	10	14	14			

TABLE 5. Information for an audit data, variables name and the code

ASSOCIATED FACTORS	CODE FOR VARIABLE
Categorical form	
Gender	
Male	1
Female	reference group
BMI category	
Normal	2
Underweight	3
Overweight	4
Obese	reference group
Duration of diabetes	
<5 years	5
5-10 years	6
>10 years	reference group
Treatment	
Diet only	7
Oral ADA only	8
Insulin only	9
Both oral and insulin	reference group
Co-morbidity	
No	10
Hypertension only	11
Dyslipidemia only	12
Hypertension and dyslipidemia	reference group
Numerical form	
Age	13
Low-Density lipoprotien	14
Blood pressure (systolic)	15

The code of the variables is the reference for Figure 4 - Figure 6

reported as been presented in Figure 1 until Figure 5.

The findings from this validation had shown that a minimum sample size of 500 was required to provide an almost accurate sample estimate for partial eta-squared, regression coefficients and r-squared of a target population. A minimum sample size of 300 also showed about similar results except that its value of r-squared was just slightly high. However, referring to Table 1, to detect a r-squared of 0.2 (since the parameter for the r-squared in our population is 0.261) only require a minimum sample of 69 for eight independent variables (since our modelling is based on eight independent variables). Results from the validation show that the samples of 100 or even 200 are not sufficient to detect a close approximation estimate for the r-squared and also other parameters such as regression coefficients.

A major concern for performing a statistical analysis is the validity of the inference drawn from the results obtained from a sample, and whether or not such an inference can be a close approximation of the true value obtained from the target population. Hence, a larger sample size will usually be required to obtain sample estimates which closely mimic those actual parameters from the target population. In any research study, there is always a possibility for its research findings to be false [15].

Most scholars agreed that this was due to a lack of consensus among different research studies especially when the investigator relies on the findings obtained from a single study only, and also when he/she depends only on statistical significance (i.e. p – value less than 0.05) [16-17]. This shall further support the notion that, although the results are found to be statistically significant (i.e. typically p - value less than 0.05); however it does not mean that the sample estimate can represent the true value of the population parameter from the intended population. Therefore, it is still very important to conduct a research study with a sufficient sample size to obtain reliable and valid estimates [13-14, 18-21].

FIGURE 1. The relation of the difference of r-squared between parameters and statistics and sample sizes

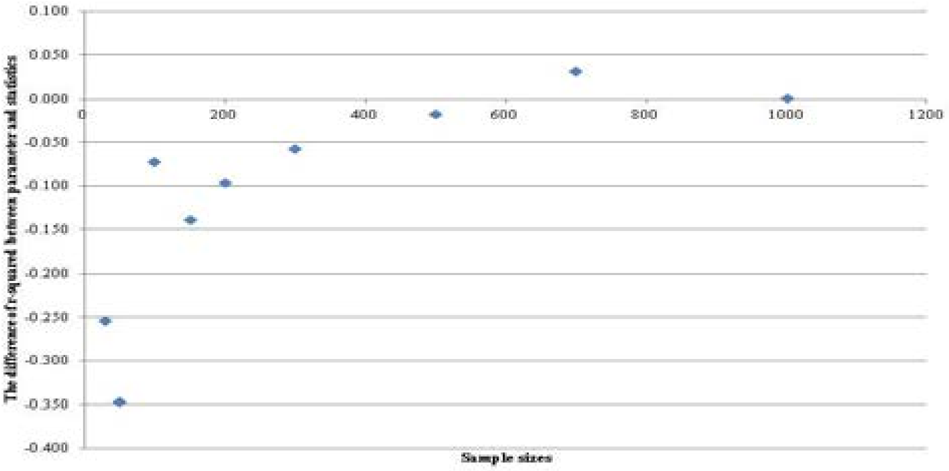


FIGURE 2. The relation of the difference of effect size (partial eat-squared) between parameters and statistics and sample sizes

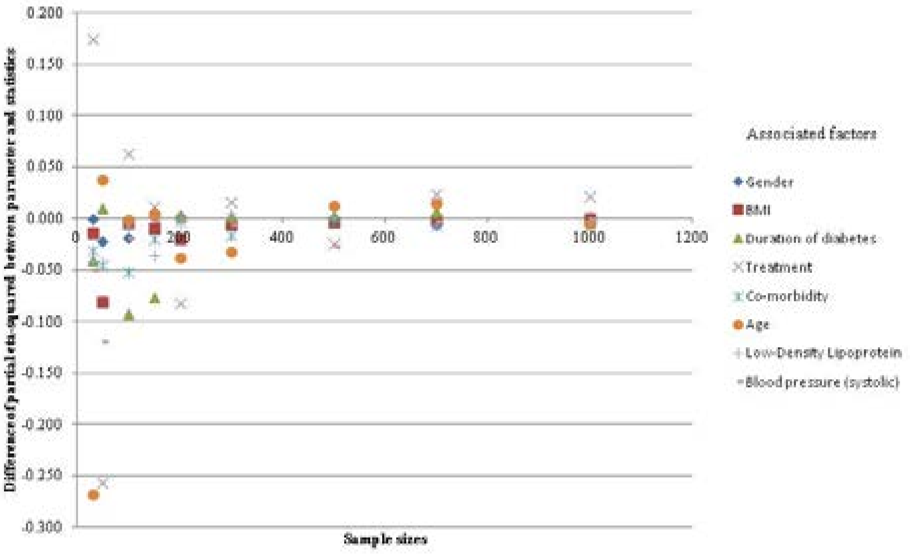


FIGURE 3. The relation of the difference of coefficient between parameters and statistics and sample sizes for sample size of 30, 50 and 100*

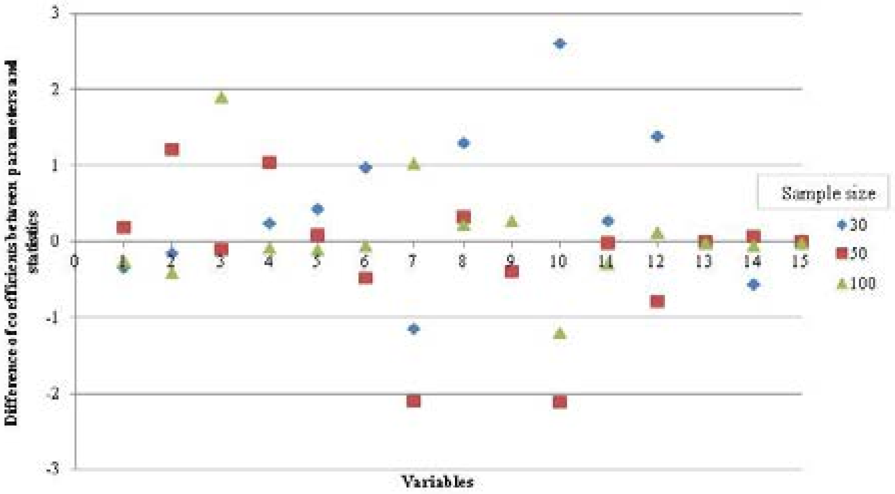


FIGURE 4. The relation of the difference of coefficient between parameters and statistics and sample sizes for sample size of 150, 200 and 300

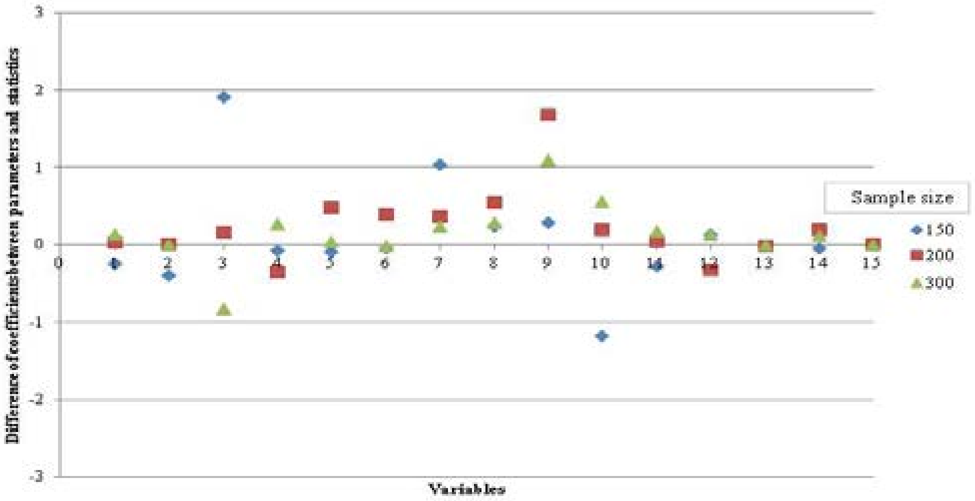
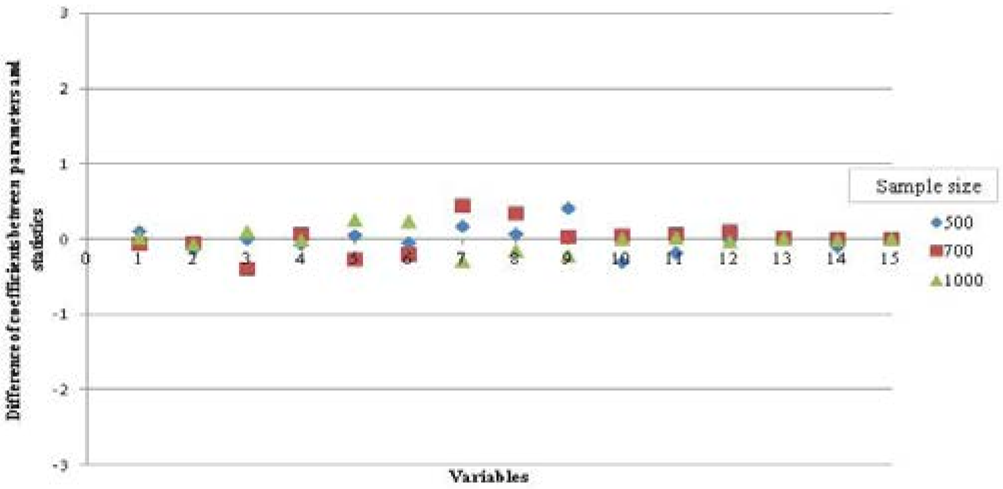


FIGURE 5. The relation of the difference of coefficient between parameters and statistics and sample sizes for sample size of 500, 700 and 1000



Based on the findings obtained from this validation, a sample size of at least 300 can yield the sample estimates which are almost accurate as the estimates for the population parameters. This finding is consistent with those from the previous studies [13-14]. This again had shown that the ideal sample size shall preferably be at least 300. During the stage of sample size planning, researchers are required to estimate a desired effect size. It is very likely for them to face some major problems when selecting a desired or clinically-relevant effect size for the purpose of sample size calculation.

This is because very often, the researchers shall only know the desired effect size after the entire analysis has been completed. Although information gathered from a pilot study or from a review of the literature can be helpful for deriving a desired or clinically-relevant effect size, however these estimates may not reflect the actual desired or clinically-relevant effect size. Therefore, a minimum sample size of at least 300 can serve as a simple rule of thumb for providing a sufficient sample size for both MLR and ANCOVA particularly for data that is collected in observational manner such as based on cohort, cross-sectional and case control studies.

The guides for estimating the minimum sample sizes for both MLR and ANCOVA which are presented in Tables 1, 2, 3 and 4 tend to have derived a smaller range of sample sizes required. We thereby proposed these guides to be used in studies that are conducted based on experimental manner whereby the samples are collected based on randomization. For any experimental studies, a smaller required sample size usually will be sufficient because an experimental study is usually a very well-planned study which will usually ensure that all potential confounding factors have been adequately controlled (or adjusted for) during the design stage.

However, if data is collected in a non-experimental manner, then collecting a larger sample size is necessary. Collecting large sample is necessary so that the statistical tests such as MLR or ANCOVA will have sufficient power to detect a sizeable effect size from each independent variable. To determine the risk factors or associated factors, confounding variable usually is a major problem in the multivariate analysis. Therefore, taking large sample such as minimum sample of 300 is necessary to control the effect from confounders' variables.

So, this study proposed sample size statement for experimental study that will apply MLR or ANCOVA and also for study that to determine risk factors or associated factors using MLR and ANCOVA. We first presented sample size statement based on an assumption that the samples based on experimental studies. Let's illustrate this point by using a simple case as an example. Say a researcher aims to predict an outcome (Y) based on five independent variables (Xs). The statement for a suggested minimum required sample size will be as follows: "The aim of this study is to predict an outcome (Y) based on five independent variables (Xs). Therefore, the minimum sample size of 58 is required to test whether a set of five predictor variables can predict the outcome with a minimum target value r-squared of 0.2. This calculation is based on a two-sided test with the values of alpha and desired power to be set at 0.05 and 0.8 respectively."

Next, let's illustrate another simple case by using the ANCOVA. Say a researcher aims to determine whether a particular independent variable "X1" is associated with an outcome "Y" after four other variables have been controlled (or adjusted for) in the analysis. The statement for a suggested minimum required sample size will be as follows: "The aim of this study is to determine whether an independent variable "X1" is associated with an outcome "Y" after four other variables have been controlled (or adjusted for) in the analysis. The value of r-squared (R^2_T) for "X1" is assumed to be estimated as 0.2 while the value of r-squared (R^2_C) for the other four controlled variables is assumed to be estimated as 0.1. Hence, a minimum sample size of 30 is required to test whether (or not) an independent variable "X1" is associated with an outcome "Y" while holding the assumption that its r-squared shall have a target value of 0.2 for its and the r-squared for a set of the other four controlled variables shall have a target value of 0.1. This calculation is based on a two-sided test with the values of alpha and desired power to be set at 0.05 and 0.8 respectively."

For data collection that is recruited based on observational studies, we recommended the minimum required sample size is 300 for both MLR and ANCOVA. So, the example of a scenario and sample size statement will be as follow. "The aim of this study is to determine to what extent a set of independent variables are associated with an outcome. Concerning on the data collection that was collected in a non-random manner and also to eliminate the influence from the confounders' variables, therefore, the recommended minimum required sample size is 300 to be able to estimate close approximation of the statistics towards the parameters in the targeted population.

A major limitation of this research study is its scope is limited to regression models which include the main effects only. So, there is no consideration of any interaction effect or of the effect of incorporating other higher order (polynomial) terms. There are also many other possible nuances which might be associated with the multiplicative terms that have been scaled in multiple regression models, which had not been considered in this research study [22]. Apart from that, this validation was performed using only a single dataset. It is beyond the scope of this study to perform audit using the other real datasets from the various non-clinical fields. Therefore, this shall provide opportunities for investigators of future research studies to consider exploring these areas.

Besides that, simulation work has not been conducted for this study. However, study with regards to estimate sample size for MLR using simulation analysis was published by Beaujean, 2014 [23]. He showed few necessary steps to estimate sample size for MLR using simulation approach. The process is highly depends on the regression model where the regression models can be varies depending on the scope of research including the number of variables and setting up the interactions between variables. Based on his findings, sample size between 200 and 400 is fairly sufficient to detect sufficient high power and also high accuracy.

In addition, the minimum sample size of 300 is larger than sample size estimated by Kelley and Maxwell (2003) which was 237. Kelley and Maxwell (2003) recommended the sample size for MLR is 237 and this rule of thumb was derived from simulation analysis [24].

CONCLUSIONS

With regards to statistical analyses that involve MLR and ANCOVA, this study proposes sample size requirement is differ for experimental and non-experimental studies. For experimental studies, estimation from Table 1 until Table 4 can become a guide to determine a minimum sample size for MLR and ANCOVA depending on the r-squared and the number of independent variables to be studied. Meanwhile, for non-experimental (observational) studies, this study recommended a rule of thumb with 300 subjects to be collected to derive statistics that are sufficiently accurate to represent the parameters in the targeted population.

Acknowledgement

The authors wish to thank the Director General of Health, Ministry of Health Malaysia for permission to publish this paper. We would like to extent our appreciation to Mr Hon Yoon Khee for his proofread of this paper. In addition, we would also thank the registry of An Audit Control Management for data sharing.

Declarations

Competing interest

All authors declare no conflicts of interest.

Ethical approval and consent to participate

This study used secondary data analysis from Patients Registry Data. There is no clinical interpretation is made since it is a methodology based study. Thus, ethics was not required.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. James S. The use of linear regression to predict digestible protein and available amino acid contents of feed ingredients and diets for fish. Aquaculture 2008;278(1–4):128–42.
2. Yamanaka N, Okamoto E, Kuwata K, Tanaka N. A multiple regression equation for prediction of posthepatectomy liver failure. Ann Surg. 1984;200(5):658–63.
3. Jennifer SG, Supriya GM, Charles EH, et al. A Phase III Randomized, Placebo-Controlled Study of Topical Amitriptyline and Ketamine for Chemotherapy-Induced Peripheral Neuropathy (CIPN): A University of Rochester CCOP Study of 462 Cancer Survivors. Support Care Cancer 2014;22(7):1807–14.
4. Wen JL, Ramli M, Thian FC, Christopher TSL, Zaki M, Adam B. Quality of life in dialysis: A Malaysian perspective. Hemodialysis International 2014;18(2):495–506.
5. Pedhazur EJ, Schmelkin LP. Measurement, design, and analysis: An integrated approach. Hillside, NJ: Lawrence Erlbaum. 1991.
6. Miller DE, Kunce JT. Prediction and statistical overkill revisited. Measurement and Evaluation in Guidance 1973;6:157-63.
7. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 6th ed. Boston: Pearson Education. 2013.
8. Gregory TK, Daniel M. Sample Sizes When Using Multiple Linear Regression for Prediction. Educational and Psychological Measurement 2008;68:431.
9. George FBJF, Wim AJGL. A simple sample size formula for analysis of covariance in randomized clinical trials. Journal of Clinical Epidemiology 2007;60(12):1234–38.
10. Ismail M, Chew B, Lee P, et al. Control and Treatment Profiles of 70,889 Adult Type 2 Diabetes Mellitus Patients in Malaysia - A Cross Sectional Survey in 2009. International Journal of Collaborative Research on Internal Medicine & Public Health 2011;3:98-113.
11. Ai TC, Ping YL, Shariff-Ghazali S, et al. Poor glycemic control in younger women attending Malaysian public primary care clinics: findings from adults diabetes control and management registry. BMC Family Practice 2013;14:188. doi: 10.1186/1471-2296-14-188
12. Fraenkel JR, Wallen NE. How to design and evaluate research in education. New York: McGraw-Hill. 2006.
13. Bujang MA, Ghani PA, Zolkepali NA, Selvarajah S, Haniff J. A comparison between convenience sampling versus systematic sampling in getting the true parameter in a population: explore from a clinical database: the Audit Diabetes Control Management (ADCM) registry in 2009. Int Conf Stat Sci Bus Eng 2009. 2012;1–5.
14. Bujang MA, Sa'at N, Joys AR, Ali MM. An audit of the statistics and the comparison with the parameter in the population. AIP Conference Proceedings, 1682, 050019. 2015. doi: 10.1063/1.4932510
15. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med 2005;2(8):e124. doi:10.1371/journal.pmed.0020124
16. Sterne JA, Davey SG. Sifting the evidence—What's wrong with significance tests. BMJ 2001;322:226–31.
17. Wacholder S, Chanock S, Garcia-Closas M, Elghormli L, Rothman N. Assessing the probability that a positive report is false: An approach for molecular epidemiology studies. J Natl Cancer Inst. 2004;96:434–42.
18. Sedlmeier P, Gigerenzer G. Do studies of statistical power have an effect on the power of studies?. Psychological Bulletin 1989;105:309–16.
19. Rossi JC. Statistical power of psychological research: What have we gained in 20 years?. Journal of Consulting and Clinical Psychology 1990;58:646–56.
20. Muller KE, Benignus VA. Increasing scientific power with statistical power. Neurotoxicology and Teratology 1992;14:211–9.
21. Cohen J. The earth is round (p < .05). American Psychologist 1994;49:997–1003.
22. Ken K, Scott EM. Sample Size for Multiple Regression: Obtaining Regression Coefficients That Are Accurate, Not Simply Significant. Psychological Methods 2003;8(3):305–21.
23. Beaujean, A. A. Sample size determination for regression models using Monte Carlo methods in R. Practical Assessment, Research & Evaluation 2014;19(12). Retrieved from http://pareonline.net/getvn.asp/v=19&n=12
24. Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. Psychological Methods 2003;8:305–21