DOI: 10.7860/JCDR/2016/18129.8744

Review Article

Statistics Section

Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis

MOHAMAD ADAM BUJANG[1], TASSHA HILDA ADNAN[2]

ABSTRACT

Sensitivity and specificity analysis is commonly used for screening and diagnostic tests. The main issue researchers face is to determine the sufficient sample sizes that are related with screening and diagnostic studies. Although the formula for sample size calculation is available but concerning majority of the researchers are not mathematicians or statisticians, hence, sample size calculation might not be easy for them. This review paper provides sample size tables with regards to sensitivity and specificity analysis. These tables were derived from formulation of sensitivity and specificity test using Power Analysis and Sample Size (PASS) software based on desired type I error, power and effect size. The approaches on how to use the tables were also discussed.

Keywords: Diagnostic, Screening, Type I error

INTRODUCTION

Sensitivity and specificity analysis is commonly used for the evaluation of screening or diagnostic studies. The most important aim of a screening or diagnostic study is, usually to determine how sensitive a screening or diagnostic test is in predicting an outcome when both the test and variable for clinical diagnosis are presented as dichotomous data. An important consideration to be made before conducting any screening or diagnostic studies is to plan and justify a sufficient sample size. This is to ensure that the results obtained from the subsequent analysis will provide the screening or diagnostic test with a desired minimum value for both its sensitivity and specificity, together with a sufficient level of power and a sufficiently-low level of type I error (i.e., its corresponding p-value).

There were studies conducted on sample size estimation for sensitivity and specificity analysis. A study by David et al., (1991) emphasized on the estimation of a minimum sample size required for a positive likelihood ratio with its respective confidence interval [1]. Meanwhile, another study by Nancy et al., (1996) emphasized on how to incorporate the value of the prevalence of a disease into the sample size calculation [2]. Besides that, a study by Claes et al., (2000) introduced an approach for estimating the minimum sample size required when the true state of disease is unknown [3]. Despite the provision of all these current guidelines developed by the scholars, it is still desirable for us to further improve the prospective estimation of a minimum sample size required for determining both the sensitivity and specificity especially for a screening and diagnostic tests.

Since the majority of researchers are not statisticians, it is likely that most researchers will require a guide to determine the minimum sample size for evaluating both the sensitivity and specificity of a screening or diagnostic test. In most instances, the minimum sample size required will depend on the objectives of the research study. For example, if an objective of the research study is to determine whether (or not) a specific tool or instrument can be used as a screening tool; then researchers will have to ensure that it has a sufficiently-high degree of sensitivity, but a lower degree of specificity can be tolerated [4,5]. On the other hand, if the researcher plans to develop a specific tool or instrument to be used as a diagnostic tool, then the researcher will usually have to target for a high degree of both sensitivity and specificity [6,7]. Due to the above, some research studies emphasize more on specificity than sensitivity [8]. Thus, different guides for estimation of a minimum sample size may be applicable for different objectives.

This review paper discusses on how to estimate sample size for sensitivity and specificity test. First of all, we presented the minimum sample sizes required for obtaining the desired sensitivity, specificity, power and type I error (i.e. p-value) for a range of low to high prevalence of the disease. Then, we provide convenient guide for researchers to follow when determining the minimum sample size required especially for two different types of studies, i.e., screening and diagnostic studies.

SAMPLE SIZE CALCULATION USING PASS SOFTWARE

The minimum sample size required for sensitivity and specificity test was calculated by using PASS software (PASS 11 citation: Hintze J (2011). PASS 11. NCSS, LLC. Kaysville, Utah, USA). PASS software is one of the commercial software that provides sample size tools for various statistical test and confidence interval scenarios [9]. We estimate the minimum sample size required, based on the different values of the prevalence of a disease and both sensitivity and specificity of a screening or diagnostic test (while in the meantime, the power is set to be at least 80% and the p-value, is set to be less than 0.05).

The values of the prevalence of a disease were set to be from 5%, and then subsequently increased to 10% and finally increased to 90% (i.e., with a stepwise increment of 10%). The values of both sensitivity and specificity to be adopted within the null hypothesis were set to range from 50% to 90% (i.e., with a stepwise increment of 10%) while those to be adopted within the alternative hypothesis were set to range from 60% to 95% {i.e., with a stepwise increment of 10%, except for the last category which consists of a stepwise increment of 5% (i.e., from 90% to 95%)}. The two different guides to be derived from this research study are namely:

(i) A guide to estimate the minimum sample size required for a screening study and,

(ii) A guide to estimate the minimum sample size required for a diagnostic study.

The overall rationale of determining the minimum sample size required for a screening study is to detect as many as true-positives as possible, hence it shall necessitate a sufficiently-high degree of sensitivity but it may not require a similarly high degree

of specificity. On the other hand, since the overall rationale of determining the minimum sample size required for a diagnostic study is to detect as many true-positives and also true-negatives at the same time, hence, it shall necessitate a sufficiently-high degree of both sensitivity and specificity.

REVIEW OF THE RESULTS

It is already well-understood that the minimum sample size required will be affected by the pre-specified values of the power of a screening or diagnostic test, its corresponding type I error and the effect size. The value of the effect size to be adopted within this research study is determined by the values of the prevalence of a disease and also the values of both sensitivity or specificity of the screening or diagnostic test {for both null (H_o) and alternative (H_a) hypotheses}.

By fixing the values of the power of a screening or diagnostic study and also the type I error, the minimum sample size required for determining both the sensitivity and specificity of a screening or diagnostic test will increase when there is a smaller clinically-important difference (in both sensitivity and specificity of a diagnostic test) between those proposed in null hypothesis and those proposed in alternative hypothesis, as illustrated by [Table/Fig-1–3]. A larger sample is also required for obtaining a higher sensitivity with a lower prevalence and vice versa (higher specificity with a higher prevalence).

The proposed estimation of the minimum sample size required for a screening study will range from 22 (Prevalence=90%, H_o=0.5 and H_a=0.8) to 980 (Prevalence=5%, H_o=0.5 and H_a=0.7), while the proposed estimation of the minimum sample size for a diagnostic study will range from 34 (Prevalence=90%, H_o=0.7 and H_a=0.9) to 4860 (Prevalence=5%, H_o=0.9 and H_a=0.95); depending on the prevalence of a disease and also on the change in the percentage values of both the sensitivity and specificity of a diagnostic test between those stated within the null hypothesis and those stated within the alternative hypothesis.

From the above, a rough guide has been prepared for estimating the minimum sample size required for both screening and diagnostic studies, which are provided in [Table/Fig-1-3]. The light grey areas are meant for proposing a minimum sample size required for a screening study, while those dark grey areas are meant for proposing a minimum sample size required for a diagnostic study (Refer to [Table/Fig-1-3]).

DISCUSSION

The concept of null hypothesis is to estimate the values of sensitivity and specificity before the study is conducted. The estimate can

n (Sensitivity)							n (Specificity)						
Prev	H_o	H_a	Power	p-value	N1	N	Prev	H_o	H_a	Power	p-value	N1	N
5%	0.50	0.60	0.804	0.047	199	3980	5%	0.50	0.60	0.804	0.047	10	209
5%	0.50	0.70	0.810	0.044	49	980	5%	0.50	0.70	0.810	0.044	3	52
5%	0.50	0.80	0.804	0.041	20	400	5%	0.50	0.80	0.804	0.041	1	21
5%	0.50	0.90	0.889	0.039	12	240	5%	0.50	0.90	0.889	0.039	1	13
5%	0.60	0.70	0.801	0.048	181	3620	5%	0.60	0.70	0.801	0.048	10	191
5%	0.60	0.80	0.826	0.034	45	900	5%	0.60	0.80	0.826	0.034	2	47
5%	0.60	0.90	0.885	0.035	19	380	5%	0.60	0.90	0.885	0.035	1	20
5%	0.70	0.80	0.818	0.044	155	3100	5%	0.70	0.80	0.818	0.044	8	163
5%	0.70	0.90	0.807	0.048	31	620	5%	0.70	0.90	0.807	0.048	2	33
5%	0.80	0.90	0.819	0.040	107	2140	5%	0.80	0.90	0.819	0.040	6	113
5%	0.90	0.95	0.839	0.043	243	4860	5%	0.90	0.95	0.816	0.048	12	243
10%	0.50	0.60	0.804	0.047	199	1990	10%	0.50	0.60	0.804	0.047	22	221
10%	0.50	0.70	0.810	0.044	49	490	10%	0.50	0.70	0.810	0.044	5	54
10%	0.50	0.80	0.804	0.041	20	200	10%	0.50	0.80	0.804	0.041	2	22
10%	0.50	0.90	0.889	0.039	12	120	10%	0.50	0.90	0.889	0.039	1	13
10%	0.60	0.70	0.801	0.048	181	1810	10%	0.60	0.70	0.801	0.048	20	201
10%	0.60	0.80	0.826	0.034	45	450	10%	0.60	0.80	0.826	0.034	5	50
10%	0.60	0.90	0.885	0.035	19	190	10%	0.60	0.90	0.885	0.035	2	21
10%	0.70	0.80	0.818	0.044	155	1550	10%	0.70	0.80	0.818	0.044	17	172
10%	0.70	0.90	0.807	0.048	31	310	10%	0.70	0.90	0.807	0.048	3	34
10%	0.80	0.90	0.819	0.040	107	1070	10%	0.80	0.90	0.819	0.040	12	119
10%	0.90	0.95	0.816	0.048	231	2310	10%	0.90	0.95	0.816	0.048	26	257
20%	0.50	0.60	0.804	0.047	199	995	20%	0.50	0.60	0.804	0.047	50	249
20%	0.50	0.70	0.810	0.044	49	245	20%	0.50	0.70	0.810	0.044	12	61
20%	0.50	0.80	0.804	0.041	20	100	20%	0.50	0.80	0.804	0.041	5	25
20%	0.50	0.90	0.889	0.039	12	60	20%	0.50	0.90	0.889	0.039	3	15
20%	0.60	0.70	0.801	0.048	181	905	20%	0.60	0.70	0.801	0.048	45	226
20%	0.60	0.80	0.826	0.034	45	225	20%	0.60	0.80	0.826	0.034	11	56
20%	0.60	0.90	0.885	0.035	19	95	20%	0.60	0.90	0.885	0.035	5	24
20%	0.70	0.80	0.818	0.044	155	775	20%	0.70	0.80	0.818	0.044	39	194
20%	0.70	0.90	0.807	0.048	31	155	20%	0.70	0.90	0.807	0.048	8	39
20%	0.80	0.90	0.819	0.040	107	535	20%	0.80	0.90	0.819	0.040	27	134
20%	0.90	0.95	0.816	0.048	231	1155	20%	0.90	0.95	0.816	0.048	58	289

[Table/Fig-1]: Sample size calculation for sensitivity and specificity analysis for prevalence of disease from 5% to 20%.
Prev = prevalence of disease H_o = Hypothesis null H_a = Hypothesis alternative
N1 = The minimum number of sample size for positive disease N = The minimum number of sample size requirement for total

n (Sensitivity)								n (Specificity)						
Prev	H_o	H_a	Power	p-value	N1	N		Perv	H_o	H_a	Power	p-value	N1	N
30%	0.50	0.60	0.804	0.047	199	663		30%	0.50	0.60	0.804	0.047	85	284
30%	0.50	0.70	0.810	0.044	49	163		30%	0.50	0.70	0.810	0.044	21	70
30%	0.50	0.80	0.804	0.041	20	67		30%	0.50	0.80	0.804	0.041	9	29
30%	0.50	0.90	0.889	0.039	12	40		30%	0.50	0.90	0.889	0.039	5	17
30%	0.60	0.70	0.801	0.048	181	603		30%	0.60	0.70	0.801	0.048	78	259
30%	0.60	0.80	0.826	0.034	45	150		30%	0.60	0.80	0.826	0.034	19	64
30%	0.60	0.90	0.885	0.035	19	63		30%	0.60	0.90	0.885	0.035	8	27
30%	0.70	0.80	0.818	0.044	155	517		30%	0.70	0.80	0.818	0.044	66	221
30%	0.70	0.90	0.807	0.048	31	103		30%	0.70	0.90	0.807	0.048	13	44
30%	0.80	0.90	0.819	0.040	107	357		30%	0.80	0.90	0.819	0.040	46	153
30%	0.90	0.95	0.816	0.048	231	770		30%	0.90	0.95	0.816	0.048	99	330
40%	0.50	0.60	0.804	0.047	199	498		40%	0.50	0.60	0.804	0.047	133	332
40%	0.50	0.70	0.810	0.044	49	122		40%	0.50	0.70	0.810	0.044	33	82
40%	0.50	0.80	0.804	0.041	20	50		40%	0.50	0.80	0.804	0.041	13	33
40%	0.50	0.90	0.889	0.039	12	30		40%	0.50	0.90	0.889	0.039	8	20
40%	0.60	0.70	0.801	0.048	181	452		40%	0.60	0.70	0.801	0.048	121	302
40%	0.60	0.80	0.826	0.034	45	112		40%	0.60	0.80	0.826	0.034	30	75
40%	0.60	0.90	0.885	0.035	19	48		40%	0.60	0.90	0.885	0.035	13	32
40%	0.70	0.80	0.818	0.044	155	388		40%	0.70	0.80	0.818	0.044	103	258
40%	0.70	0.90	0.807	0.048	31	78		40%	0.70	0.90	0.807	0.048	21	52
40%	0.80	0.90	0.819	0.040	107	268		40%	0.80	0.90	0.819	0.040	71	178
40%	0.90	0.95	0.816	0.048	231	578		40%	0.90	0.95	0.816	0.048	154	385
50%	0.50	0.60	0.804	0.047	199	398		50%	0.50	0.60	0.804	0.047	199	398
50%	0.50	0.70	0.810	0.044	49	98		50%	0.50	0.70	0.810	0.044	49	98
50%	0.50	0.80	0.804	0.041	20	40		50%	0.50	0.80	0.804	0.041	20	40
50%	0.50	0.90	0.889	0.039	12	24		50%	0.50	0.90	0.889	0.039	12	24
50%	0.60	0.70	0.801	0.048	181	362		50%	0.60	0.70	0.801	0.048	181	362
50%	0.60	0.80	0.826	0.034	45	90		50%	0.60	0.80	0.826	0.034	45	90
50%	0.60	0.90	0.885	0.035	19	38		50%	0.60	0.90	0.885	0.035	19	38
50%	0.70	0.80	0.818	0.044	155	310		50%	0.70	0.80	0.818	0.044	155	310
50%	0.70	0.90	0.807	0.048	31	62		50%	0.70	0.90	0.807	0.048	31	62
50%	0.80	0.90	0.819	0.040	107	214		50%	0.80	0.90	0.819	0.040	107	214
50%	0.90	0.95	0.816	0.048	231	462		50%	0.90	0.95	0.816	0.048	231	462
60%	0.50	0.60	0.804	0.047	199	332		60%	0.50	0.60	0.804	0.047	299	498
60%	0.50	0.70	0.810	0.044	49	82		60%	0.50	0.70	0.810	0.044	73	122
60%	0.50	0.80	0.804	0.041	20	33		60%	0.50	0.80	0.804	0.041	30	50
60%	0.50	0.90	0.889	0.039	12	20		60%	0.50	0.90	0.889	0.039	18	30
60%	0.60	0.70	0.801	0.048	181	302		60%	0.60	0.70	0.801	0.048	271	452
60%	0.60	0.80	0.826	0.034	45	75		60%	0.60	0.80	0.826	0.034	67	112
60%	0.60	0.90	0.885	0.035	19	32		60%	0.60	0.90	0.885	0.035	29	48
60%	0.70	0.80	0.818	0.044	155	258		60%	0.70	0.80	0.818	0.044	233	388
60%	0.70	0.90	0.807	0.048	31	52		60%	0.70	0.90	0.807	0.048	47	78
60%	0.80	0.90	0.819	0.040	107	178		60%	0.80	0.90	0.819	0.040	161	268
60%	0.90	0.95	0.816	0.048	231	385		60%	0.90	0.95	0.816	0.048	347	578

[Table/Fig-2]: Sample size calculation for sensitivity and specificity analysis for prevalence of disease from 30% to 60%.
Prev = prevalence of disease
H_o = Hypothesis null
H_a = Hypothesis alternative
N1 = The minimum number of sample size for positive disease
N = The minimum number of sample size requirement for total

be referred from either literatures, pilot study and sometimes by rough guidelines or target. Using some rough guidelines or target is necessary especially when there are no benchmark studies to be referred with and when pilot study could not be done for some reasons. For instance, the values of sensitivity in the null hypothesis for screening studies could be set at 50% as for rough guideline with the aim that the values should increase to indicate that the screening tool is sensitive in predicting the disease. Therefore, the role of alternative hypothesis is to estimate the values of sensitivity and specificity after the study is conducted. Basically, it is a targeted value that researchers are expecting from the performance of the screening or diagnostic tools.

The results showed that either a lower value of both sensitivity and specificity of a screening or diagnostic test to be adopted within the null hypothesis, or a smaller difference (in the values of both sensitivity or specificity of a screening or diagnostic test) between those adopted within the null hypothesis and those adopted within the alternative hypothesis, will increase the minimum sample size required. It is a similar concept in sample size calculation where larger sample is required to detect a lower effect size [10].

n (Sensitivity)								n (Specificity)						
Prev	H_o	H_a	Power	p-value	N1	N		Perv	H_o	H_a	Power	p-value	N1	N
70%	0.50	0.60	0.804	0.047	199	284		70%	0.50	0.60	0.804	0.047	464	663
70%	0.50	0.70	0.810	0.044	49	70		70%	0.50	0.70	0.810	0.044	114	163
70%	0.50	0.80	0.804	0.041	20	29		70%	0.50	0.80	0.804	0.041	47	67
70%	0.50	0.90	0.889	0.039	12	17		70%	0.50	0.90	0.889	0.039	28	40
70%	0.60	0.70	0.801	0.048	181	259		70%	0.60	0.70	0.801	0.048	422	603
70%	0.60	0.80	0.826	0.034	45	64		70%	0.60	0.80	0.826	0.034	105	150
70%	0.60	0.90	0.885	0.035	19	27		70%	0.60	0.90	0.885	0.035	44	63
70%	0.70	0.80	0.818	0.044	155	221		70%	0.70	0.80	0.818	0.044	362	517
70%	0.70	0.90	0.807	0.048	31	44		70%	0.70	0.90	0.807	0.048	72	103
70%	0.80	0.90	0.819	0.040	107	153		70%	0.80	0.90	0.819	0.040	250	357
70%	0.90	0.95	0.816	0.048	231	330		70%	0.90	0.95	0.816	0.048	539	770
80%	0.50	0.60	0.804	0.047	199	249		80%	0.50	0.60	0.804	0.047	796	995
80%	0.50	0.70	0.810	0.044	49	61		80%	0.50	0.70	0.810	0.044	196	245
80%	0.50	0.80	0.804	0.041	20	25		80%	0.50	0.80	0.804	0.041	80	100
80%	0.50	0.90	0.889	0.039	12	15		80%	0.50	0.90	0.889	0.039	48	60
80%	0.60	0.70	0.801	0.048	181	226		80%	0.60	0.70	0.801	0.048	724	905
80%	0.60	0.80	0.826	0.034	45	56		80%	0.60	0.80	0.826	0.034	180	225
80%	0.60	0.90	0.885	0.035	19	24		80%	0.60	0.90	0.885	0.035	76	95
80%	0.70	0.80	0.818	0.044	155	194		80%	0.70	0.80	0.818	0.044	620	775
80%	0.70	0.90	0.807	0.048	31	39		80%	0.70	0.90	0.807	0.048	124	155
80%	0.80	0.90	0.819	0.040	107	134		80%	0.80	0.90	0.819	0.040	428	535
80%	0.90	0.95	0.816	0.048	231	289		80%	0.90	0.95	0.816	0.048	924	1155
90%	0.50	0.60	0.804	0.047	199	221		90%	0.50	0.60	0.804	0.047	1791	1990
90%	0.50	0.70	0.810	0.044	49	54		90%	0.50	0.70	0.810	0.044	441	490
90%	0.50	0.80	0.804	0.041	20	22		90%	0.50	0.80	0.804	0.041	180	200
90%	0.50	0.90	0.889	0.039	12	13		90%	0.50	0.90	0.889	0.039	108	120
90%	0.60	0.70	0.801	0.048	181	201		90%	0.60	0.70	0.801	0.048	1629	1810
90%	0.60	0.80	0.826	0.034	45	50		90%	0.60	0.80	0.826	0.034	405	450
90%	0.60	0.90	0.885	0.035	19	21		90%	0.60	0.90	0.885	0.035	171	190
90%	0.70	0.80	0.818	0.044	155	172		90%	0.70	0.80	0.818	0.044	1395	1550
90%	0.70	0.90	0.807	0.048	31	34		90%	0.70	0.90	0.807	0.048	279	310
90%	0.80	0.90	0.819	0.040	107	119		90%	0.80	0.90	0.819	0.040	963	1070
90%	0.90	0.95	0.816	0.048	231	257		90%	0.90	0.95	0.816	0.048	2079	2310

[Table/Fig-3]: Sample size calculation for sensitivity and specificity analysis for prevalence of disease from 70% to 90%.
Prev = prevalence of disease H_o = Hypothesis null H_a = Hypothesis alternative
N1 = The minimum number of sample size for positive disease N = The minimum number of sample size requirement for total

From the above, it is clear that the minimum sample size required will depend on the pre-specified values of the power of the screening or diagnostic test, its corresponding level of type I error (i.e., its p-value) and the effect size. In this research study, we postulate that the values to be pre-specified for estimating a minimum sample size will depend on the research objectives of the study. Both screening and diagnostic studies are commonly evaluated by their sensitivity and specificity. We proposed that the basis for estimation of a screening study is that its sensitivity must be pre-determined to be at least 50.0% within the null hypothesis to indicate that the probability or chance for an instrument to detect a true-positive is in balance with at least 50.0%.

On the other hand, the minimum value of sensitivity to be adopted within the alternative hypothesis will be expected to be higher, of at least 70.0%, to indicate that the screening or diagnostic tool is fairly sensitive [11-13]. Meanwhile, the basis for estimation of a diagnostic study is that both its sensitivity and specificity will have to be pre-determined to be at least 70.0% within the null hypothesis to indicate that the probability or chance for an instrument to detect a true-positive or a true-negative is at least 70%. On the other hand, the values of both sensitivity and specificity to be adopted within the alternative hypothesis is expected to be at least 80.0% [14-16], in order to indicate that the instrument is fairly good as a diagnostic tool. However, these estimates could be arbitrary. These pre-determined values of both sensitivity and specificity of a screening or diagnostic test were adopted to ensure a valid estimation of the minimum sample size required.

It is always possible for the researchers to select different target estimates for the evaluation of both sensitivity and specificity of a screening or diagnostic study, such as aiming for higher or lower values of both their sensitivity and specificity. So, we now have illustrated two scenarios for the estimation of a minimum sample size required, along with their guiding statements for these estimations, which are based on the tabulated results.

Determination of a Minimum Sample Size Required for a Screening Study

Consider a study which aims to determine how sensitive a newly-developed instrument is in screening for Obstructive Sleep Apnea (OSA) in those patients who attended a respiratory clinic. The prevalence of OSA patients from a respiratory clinic is estimated to be approximately 80% [5]. Currently, these OSA patients will require their diagnosis to be confirmed by using Polysomnography (PSG) and such a diagnosis is costly and time-consuming. The

researcher will expect that the newly-developed instrument to be as sensitive as a screening tool in screening OSA patients, even though it may not be as accurate as a diagnostic tool. The sample size statement is as follow; "This study aims to determine to what extent a specific newly-developed instrument is as sensitive as a screening tool to screen patients for OSA."

By making reference to [Table/Fig-3], we can see that when prevalence of the disease is estimated to be 80% [5], a minimum sample size of 61 subjects (including 49 subjects having the disease) will be required to achieve a minimum power of 80% (actual power=81.0%) for detecting a change in the percentage value of sensitivity of a screening test from 0.50 to 0.70, based on a target significance level of 0.05 (actual p=0.044)."

It is important to bear in mind that the minimum sample size required for screening studies will depend on whether sensitivity or specificity of a screening test is being measured. A bigger minimum sample size will be required for measuring sensitivity of a screening test when the prevalence of a disease is lower, while a bigger minimum sample size will be required for measuring specificity of a screening test when the prevalence are higher. This is because sensitivity of a screening test aims to detect as many true-positives as possible, while specificity of a screening test aims to detect as many true-negatives as possible.

Determination of a Minimum Sample Size Required for a Diagnostic Study

Determination of a minimum sample size required for a diagnostic study will usually aim for a high value of both its sensitivity and specificity. Consider a study which aims to determine how sensitive a newly-developed instrument is in diagnosing those pre-mature babies with Retinopathy Of Prematurity (ROP). In this case, both the sensitivity and specificity of a diagnostic test are expected to be high. The prevalence of ROP among pre-mature babies is estimated to be approximately 20% [7].

So, the researcher will expect that the instrument to be both a sensitive and a specific tool to diagnose pre-mature babies with ROP. The sample size statement will be as follows; "This study aims to determine how sensitive this newly-developed instrument is in diagnosing pre-mature babies with ROP." By making reference to [Table/Fig-1], we can see that when the prevalence of the disease is estimated to be 20% [7], a minimum sample size of 535 subjects (including 107 subjects having the disease) will be required to achieved a minimum power of 80% (actual power=81.9%) in order to detect a change in the percentage value of sensitivity from 0.80 to 0.90, based on a target significance level of 0.05 (actual p=0.040). This minimum sample size is also sufficient to detect a change in the value of specificity from 80.0% to 90.0% which will only require a minimum sample of 134 subjects (including 27 subjects having the disease).

Other Considerations

Tables of minimum sample sizes required which are produced by this research study will only include discrete values of pre-specified parameters; such as a value of 5% or 20% for the prevalence of a disease and a value of 50% or 70% for the sensitivity of a test. However, estimates obtained from literature may report a more precise value of pre-specified parameters; such as given the prevalence until one or two decimal point. Thus, researchers are advised to adopt the discrete values which are nearest to these estimates obtained from literature, as illustrated and described within the two scenarios previously. This can usually be acceptable because sample size planning will only provide an estimate because it is sometime difficult to know the exact prevalence of a disease in the population and also the true performance of a specific screening or diagnostic tool until the research study has been completed.

Hence, if the researcher intends to know the minimum sample size required for obtaining an estimate of both sensitivity and specificity of a diagnostic or screening test, based on pre-specified values that beyond the estimates that we provided, then researcher may have to calculate it manually or by using a statistical software. The tables developed by this research study will therefore serve only as a rough guide in order to assist researchers in planning their sample size calculation for a screening or diagnostic study that requires the evaluation of both its sensitivity and specificity.

The prevalence of a disease is one of the pre-specified parameters which will affect the determination of a minimum sample size required for a screening or diagnostic study. As showed in the results, a larger sample will be required to detect a higher degree of sensitivity for a disease with a lower prevalence and vice versa (while a larger sample is also required to detect a higher degree of specificity for a disease with a higher prevalence). The prevalence of a disease varies from one population to another. For example, prevalence of OSA can be very low in a general patient population but it will be higher in a population with a higher risk of OSA, such as those patients attending a respiratory clinic. However, both screening and diagnostic studies will usually be conducted within the population with a higher risk of disease, because these tools (for either screening or diagnosing) are usually meant to be used in a specific patient population having the disease rather in a general patient population [4-7].

All results for the determination of minimum sample size required which were presented in this study have adopted a minimum value of 5% prevalence of a disease, which is sufficient for conducting both screening or diagnostic studies in a specific patient population having the disease. The estimated minimum sample size required will range from between 22 until 4860 depending on the pre-specified values of the power of both screening and diagnostic test, their corresponding type I error (i.e., their p-value), and the effect size. Researchers are advised not to obtain a very small sample size, such as 22 subjects (Prevalence=90%, H_o=0.5 and H_a=0.8) although its sample size calculation is still valid. At the same time, researchers may often be quite reluctant to recruit a large sample of patients because this will be costly and time-consuming.

Determination of a minimum sample size will provide only an estimate to ensure that the statistically-significant results can be obtained based on the desired effect size and a sufficient power of the screening or diagnostic test. Usually it is difficult to know the true values of these pre-specified parameters until the entire research has been completed and all analyses have been completed. Occasionally, it is possible that the true estimates for these pre-specified parameters; such as the effect size, the prevalence of a disease, the values of sensitivity and specificity of both the screening and diagnostic tests, are not yet known.

The rule-of-thumb is to obtain a large sample, which is reasonable since it will always increase the accuracy of the estimation process. Some studies had suggested that by obtaining a sample of more than 300 subjects, the estimated statistics that are derived from the sample will be likely to be the same as the true values within the intended population [17,18]. These findings were derived from an audit from several populations and tested with various statistical analyses (univariate and multivariate) and eight sub-samples were obtained for each statistical analysis. Therefore, it is possible to derive a rule-of-thumb in obtaining a sample of minimum 300 subjects, if researchers have difficulty in estimating a reliable estimate for the effect size. Based on the results that we have presented, a sample of minimum 300 subjects is often sufficiently large to evaluate both sensitivity and specificity of most screening or diagnostic tests.

CONCLUSION

Determination of a minimum sample size required for the evaluation of both sensitivity and specificity of a screening or diagnostic test will have to be based on various pre-specified parameters. Hence, a table which tabulates the estimated minimum sample sizes required for determining both sensitivity and specificity of a screening or diagnostic test (based on a set of pre-specified parameters such as prevalence of disease, etc.,) will be very helpful in providing researchers a rough guide for obtaining a minimum sample size required for their studies to be conducted on both screening and diagnostic tests.

ACKNOWLEDGEMENTS

We would like to thank the Director General of Health Malaysia for his permission to publish this article. We would like to extend the appreciation to Mr John Hon Yoon Khee for his effort in proofreading this manuscript.

REFERENCES

[1] David L Simel, Gregory P Samsa, David B Matchar. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidmiol*. 1991;44(8):763-70.
[2] Buderer NM. Statistical Methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med*. 1996;3(9):895-900.
[3] Enøe C1, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev Vet Med*. 2000;45(1-2):61-81.
[4] Netzer NC, Stoohs SA, Netzer CM, Clark K, Strohl KP. Using the Berlin questionnaire to identify patients at risk for the sleep apnea syndrome. *Ann Intern Med*. 1999;131:485-91.
[5] Yunus A, Seet W, Mohamad Adam B, Haniff J. Validation of the Malay version of Berlin questionaire to identify Malaysian patients for obstructive sleep apnea. *Malaysian Family Physician*. 2013;8(1):03–09.
[6] Burgess LJ, Maritz FJ, Le Roux I, Taljaard JJ. Use of adenosine deaminase as a diagnostic tool for tuberculous pleurisy. *Thorax*. 1995;50(6):672–674.
[7] Premsenthil M, Salowi MA, Bujang MA, Kueh A, Siew CM, Sumugam K, et al. Risk factors and prediction models for retinopathy of prematurity. *Malays J Med Sci*. 2015;22(5):57-63.
[8] Baeres M, Herkel J, Czaja AJ, Wies I, Kanzler S, Cancado ELR, et al. Liver disease: Establishment of standardised SLA/LP immunoassays: specificity for autoimmune hepatitis, worldwide occurrence, and clinical characteristics. *Gut*. 2002;51(2):259-64.
[9] http://www.ncss.com/software/pass/procedures/ [Last accessed on 03 June 2016].
[10] Mohamad AB, Nurakmal B. Sample size guideline for correlation analysis. *World Journal of Social Science Research*. 2016;3(1):37-46.
[11] Law M, Yang S, Wang H, Babb JS, Johnson G, Cha S, et al. Glioma Grading: Sensitivity, Specificity, and Predictive Values of Perfusion MR Imaging and Proton MR Spectroscopic Imaging Compared with Conventional MR Imaging. *AJNR Am J Neuroradiol*. 2003;24(10):1989-98.
[12] Choplin NT, Lundy DC. The sensitivity and specificity of scanning laser polarimetry in the detection of glaucoma in a clinical setting. *Ophthalmology*. 2001;108(5):899-904.
[13] Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. *BMJ: British Medical Journal*. 2003;327(7424):1144-46.
[14] Shea JA, Berlin JA, Escarce JJ, Clarke JR, Kinosian BP, Cabana MD, et al. Revised estimates of diagnostic test sensitivity and specificity in suspected biliary tract disease. *Arch Intern Med*. 1994;154(22):2573-81.
[15] Erbel R, Daniel W, Visser C, Engberding R, Roelandt J, Rennollet H. Echocardiography in diagnosis of aortic dissection. *The Lancet*. 1989;333(8636):457–61.
[16] Nori S, Rius-Díaz F, Cuevas J, Goldgeier M, Jaen P, Torres A, et al. Sensitivity and specificity of reflectance-mode confocal microscopy for in vivo diagnosis of basal cell carcinoma: A multicenter study. *Journal of the American Academy of Dermatology*. 2004;51(6):923–30.
[17] 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: Proceedings of the International Conference Statistics Sciences Business Engineering. 2009;2012:15.
[18] Bujang MA, Sa'at N, Joys AR, Mohamad Ali M. An audit of the statistics and the comparison with the parameter in the population. AIP Conference Proceedings, 1682, 050019. 2015.

PARTICULARS OF CONTRIBUTORS:
1. Biostatistics Unit, National Clinical Research Centre, Ministry of Health, Malaysia.
2. Biostatistics Unit, National Clinical Research Centre, Ministry of Health, Malaysia.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Mr. Mohamad Adam Bujang,
Biostatistics Unit, National Clinical Research Centre, Ministry of Health, Malaysia.
E-mail: adam@crc.gov.my

FINANCIAL OR OTHER COMPETING INTERESTS: None.

Date of Submission: **Dec 02, 2015**
Date of Peer Review: **Jan 25, 2016**
Date of Acceptance: **Jul 09, 2016**
Date of Publishing: **Oct 01, 2016**