Sampling methods and sample size determination analytical review

[1]Abingwa John Patrick, [2]Abingwa Hebron Emanuel

[1]Department of Surgery
[1]Busitema University, Faculty of Health Sciences
[2]School of computing and mathematics
[2] Kampala International University

***Abstract*- Research data measures events and indicates exposure of these events. Also, statistics is used to determine if any association detected in the sample is actually real. Majority of the cases will have some association regardless of the magnitude of the association.**

Statistical approaches exist in applied research which are used to measure frequency. These frequency measures are rates, ratios and proportions. Sampling methods, on the contrary, are used for studies to improve measures accurately, in a limited time and at a bargain.

The preference of sampling and sample size determination are quite essential in contractual research investigations to draw right conclusions. If the sample size is very small, even a well performed study may fail to detect important associations, or may estimate associations imprecisely. Equivalently, when the sample size is vast, the study would be more complex and may lead to inaccuracy in results.

Methods for sample size estimation and achieving power analysis, bank chiefly on the design and the main measure of the study. There are well laid out procedures for calculating sample size for various study designs and different outcome measures.

Besides, there are other procedures for calculating the sample size for two methods of drawing statistical inferences from the study results on the basis of confidence interval method and test of significance method. In the recent past there has been an upsurge in the growth of journals and the number of publications in survey-based studies has greatly increased.

However many of these studies, offer limited pertinent sampling methods. It was therefore important to give solitary fountain of information about the sampling and sample size determination to the researchers working in various fields. This paper, hence, gives a review of usually used terms and methods that are appropriate for specific research goals, sample size determination, and describe some sampling methods for research study goals.

Background

Sampling is a procedure that facilitates data to be collected from a small number of individuals or population within a community, organization or program and this data is then used to draw conclusions about a wider population

The notion of sampling is as old as Bible mentioned several times in luke2:1-2, Numbers3:4, Judges 21:9. In 1786, Pierre Simon Laplace determined the population of France, using a sampling method, along with ratio estimation. He also computed probabilistic estimates of the error in test statics. Alexander Ivanovich Chuprov introduced sample surveys to Imperial Russia in the 1870s

The two main advantages of sampling are the faster data collection and lower cost. Each observation measures one or more properties of observable subjects distinguished as independent individuals. In business research, medical research, agriculture research, sampling is widely used for gathering information about a population **(1)**.

Sampling Methods

The recipe for choosing subjects on which data are to be made has been expressed in literature, the following points should be taken note of when considering choosing subjects for sampling.

Investigations may be carried out on an entire group or a representative taken out from the group, whenever a sample is chosen, it should be done randomly. While selecting the samples the heterogeneity within the group should be kept in mind and proper recipe should be applied.

Some commonly described sample designs in the literature include, purposive sampling, random sampling, and quota sampling Random sampling can be of different types (2).

Purposive Sampling

In this method, sampling units are chosen according to the purpose. Purposive sampling provides biased estimate and because of this, it is not statistically recognized. This method can be used only for some particular purposes.

Depending on the research objectives, various purposive sampling methods can be employed and these include; Maximum variation or heterogeneous sampling, Homogeneous sampling, Typical case sampling, Extreme case sampling, Critical case sampling and Expert sampling (3).

Random Sampling
In this method of sampling, each unit included in the sample will have certain pre assigned probability of inclusion in the sample. This method of sampling provides a better estimate of parameters in research studies compared to purposive sampling
Every single individual in the sampling frame has a known and equal probability of being selected into the sample. It is the ideal and recognized single stage random sampling (4).

Lottery Method of Sampling
There are different ways to draw a random sample. The lottery method is the most common method used. Here, each member or subject of the population at hand is assigned a unique number. The numbers are then thoroughly mixed, like if you put them in a bowl and shook it.

Then, without looking, the researcher selects any numbers (n). The population members or subjects that are assigned that number are then included in the sample (5).

By Using Random Number Table
Most statistics and research books contain a table of random numbers as a part of the appendices. A random number table typically contains 10,000 random digits between 0 and 9 that are arranged in groups of 5 and given in rows.

In the table, all digits carry equal probability and this probability is not by presence of other digits in a row (6).

Simple Random Sampling
In simple random sampling method, each unit included in the sample has equal chance of inclusion in the sample. This technique provides the unbiased and better estimate of the parameters if the population is homogeneous (7).

Stratified Random Sampling
Stratified random sampling is a useful method for data collection if the population is heterogeneous. In this method, the entire heterogeneous population is divided into a number of homogeneous units called strata, each of these units is then sampled randomly.

The sample size in each stratum varies according to the relative significance of the stratum in the population. The method of the drawing this stratified sample is known as Stratified Sampling.

In a stratified sample, the sampling error depends on the population variance within stratum but not between the strata. Stratified random sampling also defined as where the population embraces a number of distinct categories, the frame can be organized by these categories into separate "strata." Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected (7; 8).

Cluster Sampling
Cluster sampling is a sampling method where the entire population is divided into groups or clusters and a random sample of these clusters are selected. All observations in the selected clusters are included in the sample.

Cluster sampling is a sampling technique used when "natural" but relatively homogeneous groupings are evident in a statistical population. Cluster sampling is generally used when the researcher cannot get a complete list of the units of a population they wish to study but can get a complete list of groups or 'clusters' of the population.

This sampling method may well be more practical and economical than simple random sampling or stratified sampling. Compared to simple random sampling and stratified sampling, cluster sampling has advantages and disadvantages.

For example, given equal sample sizes, cluster sampling usually provides less precision than either simple random sampling or stratified sampling. However, contact costs between clusters are high; cluster sampling may be more cost-effective than the other methods (9).

Systematic Random Sampling
In this method of sampling, the first unit of the sample selected at random and the subsequent units are selected in a systematic way. If there are N units in the population and n units are to be selected, then R = N/n (the R is known as the sampling interval). The first number is selected at random out of the remainder of this R (Sampling Interval) to the previous selected numbers (10).

.**Multistage Random Sampling**
Multistage random sampling, units are selected at various stages. The sampling designs may be either same or different at each stage. Multistage sampling method also known as cluster sampling, it involves the use of samples that are to some extent of clusters.

The principle advantage of this sampling method is that it permits the available resources to be concentrated on a limited number of units of the frame, but the sampling error will be increased (11).

Quota sampling
In quota sampling, the population is segmented into mutually exclusive sub-groups, just as stratified sampling. Then judgment is used to select the subjects or units from each segment based on a specified proportion.

The second step makes the method one of non-probability sampling. In quota sampling, the selection of the sample is non-random. For example interviewers might be tempted to interview those who look most helpful.
Samples may be biased because not everyone gets a chance of selection. This random element is greatest (12).

Independent Sampling
Independent samples are those samples selected from the same population, or different populations, which have no effect on one another. That is, no correlation exists between the samples (13).

SAMPLE SIZE FOR RESEARCH PURPOSES
The sample size should considerately be determined so that it will be sufficient to draw valid and generalized conclusions. The determination of a sufficient sample size needs specific data about the problems under study in the population being studied.

Including, the sub categories of sample require for analysis, variation, precision, availability and cost of the studies. The data collected during the study from these samples is to be recorded on pre-designed schedule or the data collecting tool. The design the tool depends on the objectives and facilities for analysis.

Sample size determination is the technique of choosing the number of observations to be included in a sample. The sample size is an important feature of any study in which the aim is to make inferences about the population from a sample.

In general, the sample size used in a study is determined based on the cost of data collection, and based on sufficient statistical power. In advanced studies there may be several different sample sizes involved in the study: for example, in a survey sampling if the population is heterogeneous involving stratified sampling there would be different sample sizes for each population.

In census, data are collected through complete counting; hence the sample size is equal to the population size.
A study may be divided into different study groups. In a census, data are collected through complete counting; hence the sample size is equal to the population size.

A study may be divided into different experimental groups, there may be different sample sizes for each research group. Larger sample sizes generally lead to increased precision when estimating unknown parameters. Several fundamental facts of mathematical statistics describe this phenomenon, including the law of large numbers and the central limit theorem. Generally sample sizes may be chosen in three different ways as follows.

Cost base
Include those items readily available or convenient to collect. A choice of small sample sizes, though sometimes necessary, can result in wide confidence intervals or risks of errors in statistical hypothesis testing (14).

Variance base
Using a target variance for an estimate to be derived from the sample eventually obtained

Statistical power base
Using a target for the power of a statistical test applied once the sample is collected.
Sample sizes are judged based on the quality of the resulting estimates, sample size may be assessed based on the power of a hypothesis test **(15)**.

Significance of the Sample Size
In a comparative study, the means or proportions of some characteristic in two or more comparison groups are observed.
A statistical test is then applied to determine the significant difference between the means or proportions observed in the different groups.
Sample size is important principally due to its effect on statistical power. Statistical power is the chance that a statistical test will indicate a significant difference when there truly is one.
Statistical power is analogous to the sensitivity of a diagnostic test and one could mentally substitute the word sensitivity‖ for the word power‖ during statistical conclusions
In a comparative study of two groups of individuals, the power of a statistical test must be sufficient for detection of a statistically significant difference between the two groups if a difference is truly present.
This issue becomes important if the study results were to demonstrate no statistically significant difference. If such a negative result were to occur, there would be two possible interpretations.
The first interpretation is that the results of the statistical test are correct and that there truly is no statistically significant difference (a true-negative result).
The second interpretation is that the results of the statistical test are erroneous and that there is actually an underlying difference, but the study was not powerful enough (sensitive enough) to find the difference, yielding a false-negative result. In statistical terminology, a false-negative result is known as a type II error.
An adequate sample size gives a statistical test enough power so that the first interpretation is much more plausible than the second interpretation (that a type II error occurred) in the event no statistically significant difference is found in the study

It is well known that many published clinical research studies possess low statistical power owing to inadequate sample size or other design issues (16).

Determination of the sample size for parameters

A sample size generally depends on five study design parameters: minimum expected difference or also known as the effect size, estimated measurement variability, desired statistical power, significance criterion, and whether a one- or two-tailed statistical analysis is planned

Basis of Minimum Expected Difference or size effects

The minimum expected difference is made smaller, the sample size needed to detect statistical significance increases. The setting of this parameter is subjective and is based on clinical judgment or experience with the problem being investigated.

For example, suppose a study is designed to compare a standard diagnostic procedure of 85% accuracy with a new procedure of unknown but potentially higher accuracy. Suppose the investigator believes that it would be a clinically important improvement if the new procedure were 95% accurate. Therefore, the investigator would choose a minimum expected difference of 10% (0.10).

Estimated Measurement Variability

This parameter is represented by the expected standard deviation in the measurements decide within each comparison group.
As statistical variability increases, the sample size needed to detect the minimum difference increases. Ideally, the estimated measurement variability should be determined on the basis of preliminary data collected from a similar study population.
A review of the literature can also provide estimates of this parameter. If preliminary data are not available, this parameter may have to be estimated on the basis of subjective experience, or a range of values may be assumed.
A separate estimate of measurement variability is not required when the measurement being compared is a proportion (in contrast to a mean), because the standard deviation is mathematically derived from the obvious trade-off with the number of individuals that can feasibly be investigated, given the usually fixed amount of time and resources available to conduct a research study
In randomized controlled trials, statistical power is customarily set to a number greater than or equal to 0.80, with many clinical trial experts now advocating a power of 0.90.
This parameter is the maximum P value for which a difference is to be considered statistically significant. As the significance criterion is decreased, the sample size needed to detect the minimum difference increases. The statistical significance criterion is customarily set to 5 percent **(17)**.

One- or Two-tailed Statistical Analysis

In a some cases, it may be known before the investigation that any difference between comparison or experimental groups is possible in only one direction.
In such cases, use of a one-tailed statistical analysis, which would require a smaller sample size for detection of the minimum difference than would a two-tailed analysis, may be considered.
The sample size of a one-tailed study design with a given statistical significance criterion—for example, α—is equal to the sample size of a two-tailed design with a significance criterion of 2α, all other parameters being equal
Criteria for good Sample Size
In addition to the purpose of the study and population size, three criteria usually will need to be specified to determine the appropriate sample size: the level of precision, the level of confidence or risk, and the degree of variability in the attributes being measured **(18)**.
Each of these is reviewed below

The Level of Precision

The level of precision, sometimes called sampling error, is the range in which the true value of the population is estimated to be.
This range is often expressed in percentage points (e.g., ±5 percent) in the same way that results for political campaign polls are reported by the media.
Thus, if a researcher finds that 60% of farmers in the sample have adopted a recommended practice with a precision rate of ±5%, then he or she can conclude that between 55% and 65% of farmers in the population have adopted the practice **(18)**.

The Confidence Level

The risk level of confidence level is based on ideas of Central Limit Theorem. The key idea in the Central Limit Theorem is that when a population is repeatedly sampled, the average value of the attribute obtained by those samples is equal to the true population value.
Further, the values obtained by these samples are normally distributed about the true value, with some samples having a higher value and some obtaining a lower value than the true population value. In a normal distribution, approximately 95% of the sample values are within two standard deviations of the true population value. This confidence interval is also known as risk of error in the statistical hypothesis testing
In other words, this means that if a 95% confidence level is selected, 95 out of 100 samples will have the true population value within the range of precision specified.

There is always a probability that the sample obtain by the researcher or investigator does not represent the true population value. Such samples with extreme values are represented.
This risk is reduced for 99% confidence levels and increased for 90% or lower levels of confidence **(18)**.

Degree of Variability
The third criterion, the degree of variability in the attributes being investigated, refers to the distribution of attributes in the population.
The variables with more homogeneous population, the smaller the sample size required. If the more heterogeneous population, the larger the sample size required to obtain a given level of precision.
For example, a proportion of 50% indicates a greater level of variability than either 80% or 20%. This is because 80% and 20% indicate that a large majority do or do not, respectively, have the attribute of interest.
Because a proportion of .5 indicates the maximum variability in a population, it is often used in determining a more conservative sample size, that is, the sample size may be larger than if the true variability of the population attribute were used
Strategies for Determining Sample Size
There are many approaches to determining the sample size. These include using a census for small populations, imitating a sample size of similar studies, using published tables, and also applying formulas to calculate a sample size (19).

Using a Census for Small Populations
One approach is to use the entire population as the sample. Although cost considerations make this impossible for large populations, a census is more attractive for small populations (e.g., 200 or less). A census eliminates sampling error and provides data on all the individuals in the population. In addition, some costs such as questionnaire design and developing the sampling frame are "fixed," that is, they will be the same for samples of 50 or 200. Therefore, entire population will have to be sampled in small populations to achieve a desirable level of precision **(20)**.

Using A Sample Size Of A Similar Study
Using a Sample Size of a Similar Study Another approach is to use the same sample size as those of studies similar to the one you plan. Without reviewing the procedures employed in these studies you may run the risk of repeating errors that were made in determining the sample size for another study. However, review of the literature in this discipline can provide guidance about "typical" sample sizes which are used **(21)**.

Using Published Tables
A third way to determine sample size is to rely on published tables, which provide the sample size for a given set of criteria. Sample sizes that would be necessary for given combinations of precision, confidence level and variability presented two tables for the selection of sample size (Table-1 and Table-2). Please note two things. First, these sample sizes reflect the number of obtained responses and not necessarily the number of surveys mailed or interviews planned. Second, the sample sizes in Table 2 presume that the attributes being measured are distributed normally or nearly so. If this assumption cannot be met, then the entire population may need to be surveyed

Table 1. Sample size for □3%, □5%, □7% and □10% Precision Levels Where Confidence Level is 95% and P=.5.

Size of Population	Sample Size (n) for Precision (e) of:			
	□3%	□5%	□7%	□10%
500	a	222	145	83
600	a	240	152	86
700	a	255	158	88
800	a	267	163	89
900	a	277	166	90
1,000	a	286	169	91
2,000	714	333	185	95
3,000	811	353	191	97
4,000	870	364	194	98
5,000	909	370	196	98
6,000	938	375	197	98
7,000	959	378	198	99
8,000	976	381	199	99
9,000	989	383	200	99
10,000	1,000	385	200	99
15,000	1,034	390	201	99
20,000	1,053	392	204	100
25,000	1,064	394	204	100
50,000	1,087	397	204	100
100,000	1,099	398	204	100
>100,000	1,111	400	204	100

a = Assumption of normal population is poor (Yamane, 1967). The entire population should be sampled.

Size of Population	Sample Size (n) for Precision (e) of: Population		
	□5%	□7%	□10%
100	81	67	51
125	96	78	56
150	110	86	61
175	122	94	64
200	134	101	67
225	144	107	70
250	154	112	72
275	163	117	74
300	172	121	76
325	180	125	77
350	187	129	78
375	194	132	80
400	201	135	81
425	207	138	82
450	212	140	82

Table 2. Sample size for ±5%, ±7% and ±10% Precision Levels Where Confidence Level is 95% and P=.5

Using Formulas to Calculate a Sample Size

Glenn tables can provide a useful guide for determining the sample size, you may need to calculate the necessary sample size for a different combination of levels of precision, confidence, and variability.

The fourth approach to determining sample size is the application of one of several formulas was used to calculate the sample sizes in Table 1 and Table 2.

Some techniques for Calculation of sample size by Kish's Formular

1. Required Sample sizes for hypothesis tests by Cohen's d and Power

Calculating the sample size required to yield a certain power for a test, given a predetermined Type I error rate α. As follows, this can be estimated by pre-determined tables for certain values, by Mead's resource equation, or, more generally, by the cumulative distribution function:

□ The desired statistical power of the trial, shown in column to the left.

□ Cohen's d (=effect size), which is the expected difference between the means of the target values between the experimental group and the control group, divided by the expected standard deviation

Table 3: Sample sizes for hypothesis tests by Cohen's d and Power

Cohen's d	POWER							
	0.25	0.50	0.60	0.70	0.80	0.90	0.95	0.99
0.20	84	193	246	310	393	526	651	920
0.50	14	32	40	50	64	85	105	148
0.80	06	13	16	20	26	34	42	58

***2.* Determination of sample size for laboratory animal study based on Mead's resource equation**

Mead's resource equation is often used for estimating sample sizes of laboratory animals, as well as in many other laboratory experiments.

It may not be as accurate as using other methods in estimating sample size, but gives a hint of what is the appropriate sample size where parameters such as expected standard deviations or expected differences in values between groups are unknown or very hard to estimate (2; 22)

All the parameters in the equation are in fact the degrees of freedom of the number of their concepts, and hence, their numbers are subtracted by 1 before insertion into the equation.

The equation is:

$E = N - B - T$

Where:

N is the total number of individuals or units in the study (minus 1)
B is the blocking component, representing environmental effects allowed for in the design (minus 1)
T is the treatment component, corresponding to the number of experimental groups (including control group) being used, or the number of questions being asked (minus 1)

E is the degrees of freedom of error component, and shall be somewhere between 10 and 20.
For Example, if a study using laboratory animals is planned with four treatment groups (T=3), with eight animals per group, making 32 animals total (N=31), without any further stratification (B=0), then E would equal 28, which is above the cutoff of 20, indicating that sample size may be a bit too large, and six animals per group might be more appropriate

3. Determination of sample size by cumulative distribution function
Let Xi, i = 1, 2, ...,n be independent observations taken from a normal distribution with unknown mean μ and known variance σ2.
Let us consider two hypotheses, a null hypothesis:
H0: $\mu = 0$

And an alternative hypothesis:

H1: $\mu = \mu^*$
For some smallest significant difference $\mu^*>0$. This is the smallest value for which we care about observing a difference.

Now, if we wish to reject H0 with a probability of at least 1-β when H1 is true (i.e. a power of 1-β), and second reject H0 with probability α when H0 is true, then we need the following:
If Zα is the upper α percentage point of the standard normal distribution, then
P x > Zασ / √ n H0] = α

And hence, 'Reject H0 if our sample average sample mean is more than ***Zασ*** / √ **n**
is a decision rule for 1-tailed test..
Now we wish for this to happen with a probability at least 1-β when H1 is true. In this case, our sample average will come from a Normal distribution with mean μ*. Therefore we require
P [x > Zασ / √ n H1] = 1 – β where x = sample mean

Through careful manipulation, this can be shown to happen when
n ≥ [{Φ-1 (1-β) + Zα} / (μ* / σ)

Where Φ is the normal cumulative distribution function**.**

Formula for Calculating A Sample for Proportions

Cochran developed the equation to yield a representative sample for proportions of large sample.
n0 = Z2 p q / e2
Which is valid where n0 is the sample size, Z2 is the abscissa of the normal curve that cuts off an area α at the tails (1 - α equals the desired confidence level is 95%), **e** is the desired level of precision, p is the estimated proportion of an attribute that is present in the population, and q is 1-p.

The value for Z is found in statistical tables which contain the area under the normal curve.
To illustrate, suppose we wish to evaluate a state-wide Extension program in which farmers were encouraged to adopt a new practice

. Assume there is a large population but that we do not know the variability in the proportion that will adopt the practice; therefore, assume p=.5 (maximum variability). Furthermore, suppose we desire a 95% confidence level and ±5% precision.

The resulting sample size is n0 = Z2 p q / e2
Finite Population Correction for Proportions (If small population)
If the population is small then the sample size can be reduced slightly.

This is because a given sample size provides proportionately more information for a small population than for a large population.
The sample size (n0) can be adjusted as **n = n0 / [1 + {(n0 – 1) / N}]**
Where n is the sample size and N is the population size.
Suppose our evaluation of farmers' adoption of the new practice only affected 2,000 farmers. The sample size that would now be necessary is given as
n = n0 / [1 +{(n0 – 1) / N}]
This adjustment can substantially reduce the necessary sample size for small populations and also called the population correction

A Simplified Formula for Proportions
Yamane provides a simplified formula to calculate sample sizes. This formula was used to calculate the sample sizes in Tables 2 and 3 and is shown below. A 95% confidence level and P = .5 are assumed.
n = N / [1 + N (e)2]

Where n is the sample size, N is the population size, and e is the level of precision. When this formula is applied to the above sample, we get.
n = N / [1 + N (e)2]

Rao presented some another calculation for sample size under different circumstances in simple manner. These determinations are also more useful for medical or clinical research investigations
a. When it is a field survey to estimate the prevalence rate of specific event or cases or disease the sample size is calculated by the formula
n = 4 p q / L2

Where n is the required sample size, p is the approximate prevalence rate for which the survey is to be conducted. The knowledge of this is to be obtained from previous surveys or from pilot **survey. q = 1 – p and L is the permissible error in the estimate.**

Table 4: Similarly, calculated sample size for different levels.

Prevalence		Permissible error in the estimate 5%	10%
p (%)	1-p = q (%)	Required Sample	Required Sample
0.5	99.5	318400	79600
01	99.0	158400	39600
05	95.0	30400	7600
10	90.0	14400	3600
25	75.0	4800	1200
50	50.0	1600	400

When conducting research investigation on quantitative data, the sample size is calculated by the given formula
n = tα2s2 / ε2
Where n is the desired sample size, s is the standard deviation of observations, ε is the permissible in the estimate of mean and tα is the value of at 5%level of significance

For illustration
If from pilot it is known the mean is 12gm.% with 1.5 gm.% standard deviation and permissible error 0.5 gm.%. So t0.05 = 2.0

Therefore required sample size
n = [(2.0)2 x (1.5)2] / (0.5)2 = 36
(c) In a clinical trials usually there will be two groups one experimental and the other control group. In order to estimate the size of sample for each group, the difference in the response rates of the two groups is to be taken in to consideration and the sample size is estimated from the following formula
n = 2 tα2s2 / d2
where n is the required sample size for each group, s is the pooled standard deviation of the two groups and d is anticipated smallest difference in the estimates for the two groups and tα is usually taken as 5 % level of significance.

For illustration
If d is the smallest anticipated difference in the rise of mean between two groups is 2%, pooled standard deviation s = 3.0 gm.% and t0.05= 2
Therefore, required sample size
n = [2 x (2)2 x (3.0)2] / (2)2 = 18
The appropriate sample size for a population-based survey is determined largely by three factors: the estimated prevalence of the variable of interest, chronic malnutrition in this instance level of confidence and the acceptable margin of error.
In the similar manner sample size can be calculated based on margin of error in confidence interval especially for estimation of population mean.

Z x (s /√n) where **s** is the standard deviation. If fluctuations in the estimate of population mean is **ε**
Z x (s /√n)<ε Therefore, n = [Z2x S2] / ε2
For calculation if standard deviation 0.4 gm and fluctuation in the estimated mean is 3 gm with 98% confidence interval **n =** [(2.326)2 x (3)2] / (0.4)2 = 304.3 size will therefore minimum sample be n= 305

Concepts to Minimize the Sample Size
The statistical tests that incorporate the use of continuous values are mathematically more powerful than those used for proportions, given the same sample size.

In a radiological diagnosis is expressed in terms of a binary result, such as the presence or absence of a disease, it is natural to convert continuous measurements into categories.

For example, the size of a lesion might be coded as ―small‖ or ―large.‖ For a sample of fixed size, the use of the actual measurement rather than the proportion in each category yields more power **(16; 23)**.

Use of More Precise Measurements
In the investigation any way to increase the precision or decrease the variability of the measurement process should be sought. For some types of research, precision can be increased by simply repeating the measurement.

More complex equations are necessary for studies involving repeated measurements in the same individuals but the basic principles are similar **(24)**.

Use of Paired Measurements
The paired t test are statistically more powerful for a given sample size than are unpaired tests because in paired tests, each measurement is matched with its own control.

For example, instead of comparing the average lesion size in a group of treated patients with that in a control group, measuring the change in lesion size in each patient after treatment allows each patient to serve as his or her own control and yields more statistical power.

The additional power and reduction in sample size are due to the standard deviation being smaller for changes within individuals than for overall differences between groups of individuals

Use of Unequal Group Sizes
Sample size is statistically most efficient if the two groups are equal in size, benefit is still gained by studying more individuals, even if the additional individuals all belong to one of the groups.

For example, it may be feasible to recruit additional individuals into the control group even if it is difficult to recruit more individuals into the non control group.

More complex equations are necessary for calculating sample sizes when comparing means and proportions of unequal group sizes **(25)**.

Expansion of the Minimum Expected Difference.
Study is a preliminary. The results are a minimum expected difference that has been specified and unnecessarily small, larger expected differences could be justified, especially if the planned study could be used to justify a more ambitious follow-up study of a larger number of individuals and a smaller minimum difference.

Some Recent Reports Emphasizing Sample Size
Some important studies which can be used to emphasize the significance of sample size include the determination of sample size for animal studies, sample size calculations for cluster randomized controlled trials (CRCT) for fixed number of clusters, calculation of sample size for medical research and sample size and power analysis in medical research.
Macfarlane derived sample size determination for research projects for medical sciences whereas studied the sample size calculations for trials in health services research
presented systematically the outline sample size formulae including required number of randomization units, detectable difference and power for cluster randomized control trials (CRCT) with a fixed number of clusters, to provide a concise summary for both binary and continuous outcomes and also extensions to the case of unequal cluster sizes were provided.

This study concluded that CRCT with a fixed number of clusters might mean that the study would not be feasible and lead to the notion of a minimum detectable difference (or a maximum achievable power) irrespective of how many individuals were included within each cluster (26).

Macfarlane has described the sample size calculations for research projects as an essential part of a study protocol for submission to ethical committees, research funding bodies and some peer reviewed journals.

In this study, it is conclude that sample size calculation is an important part of study design and a professional statistician is the best person to ask for help when planning a research project.

However, researchers must be prepared to provide the necessary information in order that the sample size can be determined

Wood presented the current orthodox way of estimating sample size for a trial is through a power calculation based on a significance test.

This study carries the assumption that test should be the center piece of the statistical analysis.

However, it is increasingly the case that confidence intervals are preferred to significance tests in summarizing the results of trials particularly in health services research.

He believes that the way sample size is estimated should reflect this change and focus on the width of the confidence interval rather than on the outcome of a significance test.

Such a method of estimation is described and shown to have additional advantages of simplicity and transparency, enabling a more informed debate about the proposed size of trials.

Researchers must calculate the sample size before starting of any animal study. It should be adequate enough to detect a small significant difference between the groups. In this study also reported that small sample size is not only responsible for the insignificant result but also for the less power of the study. Calculation of sample size involves complex statistics but it can be simplified to help the researchers who are not from statistical background.

Dell and coworkers have described the methodology of sample size determination for use in animal base experimental research. They calculated the sample size for single group experiment, continuous variable, sample size for repeated studies and for time to an event

Sample size for single group experiment **(n) = log β / log p** where β is the probability of Type II error (usually 0.10 or 0.05) and p represents the proportion of the animals are not infected

Second case for continuous variable:
In this case, a simple formula derived from the formula for the *t*-test can be used to compute sample size when power, significance level, size of difference in means (called the effect), and variability or standard deviation of the population means are specified.

Sample Size for Continuous Variable **(n) =1 + 2 C (S / d)2**
Where *C* is dependent on values chosen for significance level (α) and power (1-β) and also defined the Constant C is Dependent on the Value of α and 1-β.
Third case for repeated studies: In this case n is derived from the paired *t*-test equation
Sample size (n) = 2 + C (S/d)2

where (S/d)2 is multiplied by C in paired studies, rather than 2C, indicating that a paired study is more powerful than a comparison of two independent means, as occurs in sample size calculations of continuous variables

In this last case, the researchers could estimate the proportion of the control group that would exhibit the event and can state a difference that must be detected between the control group and the experimental group. The smaller this difference, the more animals will be needed
.
Thus, given estimates for proportion of the control group exhibiting the event (**pc**) and the desired proportion of the experimental group exhibiting the event (pe), then
Sample size (n) = C [(pc qc + pe qe) / d2] + (2 /d) + 2 Where qc = 1 – pc, q*e* = 1 – pe and d = |pc – pe|. d is the difference between pc and pe, expressed as a positive quantity

Sampling theory, thus, is undoubtedly an important aspect of applied and scientific research investigations. Generally different sampling methodologies help to draw the good sample or better representative for estimation of parameters.

Sample size is also more important to increase the precision of results, minimize the variability and for generalization of results with interpretation. In this paper for determination of the sample sizes for different types of research investigation are discussed. The above approaches to determining sample size have assumed that a simple random sample is the sampling design assumed that a simple random sample is the sampling design.

Another consideration with sample size is the number needed for the data analysis. If descriptive statistics are to be used, e.g., mean, frequencies, then nearly any sample size will suffice. In addition, an adjustment in the sample size may be needed to accommodate a comparative analysis of subgroups.

Sudman suggested that a minimum of 100 elements were required for each major group or subgroup in the sample and for each minor subgroup, a sample of 20 to 50 elements was necessary. According to Kish 30 to 200 elements are sufficient when the attribute is present 20 to 80 percent of the time if the distribution approaches normality. On the other hand, skewed distributions can result in serious departures from normality even for moderate size samples.

Finally, the sample size determination techniques provide the number of responses that need to be obtained. Many researchers commonly add 10% to the sample size to compensate for persons that the researcher is unable to contact. The sample size is also often increased by 30% to compensate for no-response. Thus, the number of mailed surveys or planned interviews can be substantially larger than the number required for a desired level of confidence and precision

CONCLUSION AND RECOMMENDATION

Research investigation with the help of appropriate research designs provides the unbiased estimates of parameters through unbiased estimates the health status of community can be monitored.

In the other hand research investigation is part of a wider development of any nation with regard to finance, education, public health, and agriculture, etc. that are indicators of better life of human beings.

Modern contractual research based on better living management is complex, requiring competent set of skills such as medical, social, technological, mathematical, statistical etc. in advanced research with the help of suitable statistical tools and sampling research methods which provides unbiased estimates of indicators, conclusions and their interpretations based on good samples with appropriate sample size.

The present report emphasized the significance of sampling and determination of sample size in statistical research.

The most common methods for sampling and sample size determination employed in recent statistical studies based on normal distribution, confidence interval (risk of error in the statistical hypothesis testing) and permissible error in the estimate.

An investigator or researcher can calculate the appropriate sample size according to design of the study. Attention to detail of a sample size will hopefully result into a more meaningful study whose results and interpretation will eventually receive a high priority for publication.

REFERENCES:

1. Sep, Ali Zulfiqar and S Bala Bashkir India J..Anaesth 2016 and 60(9):662-669.doi:10.4103/00195049, 190623.
2. Ajay S.Singh Micah Masuka, International Journal of Economics, commerce and Management, Nov2014, VLII, United Kingdom,http://ijecm.co.uk/.
3. Robinson, R.S(2014).Purposive SamplingIn:Michalos,A.C.(eds) Encyclopedia of Quality of Life and Well-Being Research.Springer,Dordrecht.https//doi.org/10.1007/978-007-0753-2337.
4. Shagofah Noor, Jawad Golzar, Omid Tajik,International Journal of Education and Language studies 2022 DOI:10.22034/ijels.2022.162982.
5. Bhardwaj P.Types of sampling in research.J Pract Cardiovasc Sci 2019, 5:157-63.
6. Bhardwaj P.Types of sampling in research.J Pract Cardiovasc Sci 2019, 5:157-63.
7. Mulisa, F.(2022), International Journal of Academic research in Education,8(1), 75-78. DOI:10.17985/ijare.1225214.
8. Hamed Taherdoost International Journal of Academic research in Management (IJARM) 5.No 2, 2016 page 18-27.ISSN:2296-1747@Helvetic Editions LTD,Switzerland www.elvedit.com.
9. Iqbal Jeelan M, Faizan Danish and Mansha Gul Biostatistics and Biometrics Open Access Journal ISSN:2633 Volume 5 Issue5-March 2018.DOI:10.19080/BBOAJ.2018.05.555673.
10. Iqbal Jeelan M, Faizan Danish and Mansha Gul Biostatistics and Biometrics Open Access Journal ISSN:2633 Volume 5 Issue5-March 2018.DOI:10.19080/BBOAJ.2018.05.555673.
11. Philip Sedgwick, Institute for Medical and Biomedical Education.BMJ2015 and July2015), 351:h4155doi:10.1136/bmj.h4155 (Published31.
12. Keming Yang and Ahmad Banamah, Sage Journals, Volume 19 Issue1 https:// doi.org/10.5153/sro.3.
13. Keming Yang and Ahmad Banamah, Sage Journals, Volume 19 Issue1 https:// doi.org/10.5153/sro.3.
14. Chinelo Blessing ORIBHABOR, Chioma A.ANYANWU.Federal university Dutsun-Ma Journal of.

15. Simarjeet Kaur International Journal of Current research Vol9, Issue03, pp.48365-4837, March 2017 at://www http.com.
16. Browner WS, Newman TB, Cummings SR, Hulley SB (2001): Estimating sample size and power.
17. Areti Angeliki, Dan Jackson, Ralf Bender, Wileyonlinelibrary DOI:10.1002/jrsm.1164.
18. Ganng Chen, Robert W.Cox, Daniel R Glen , PubMed Central DOI and 10.1002/hbm.24399.
19. Ajay S Singh and Masuku, International journal of Commerce and Management ISSN 23480286.
20. Statistics, WHO/ESCAP Training Manaual on Disability.
21. Avijit Azra, Nithya Gogtay Indian journal of Dematology DOI:10.4103/0019-5154.190119.
22. Wan Nor Arifin, Wan Mohd Zahiruddin PubMed Central, DOI: 10.21315/mjms2017.24.5.11 2017.
23. KK Gupta, JP Atri, A Singh et al PubMed Central DOI: 0.4103/1658-354X.174918 2016.
24. Shiyang Ma, Tianying Wang BMC Medical Research , DOI: 10.1186/s12874-023-01893w 2023.
25. Rosner, B. Scientific Research journal 6th Edition 2000.
26. Louangrath, Pi ISSN: 2415-0371 DOI: 10.5281/zenodo.3877623 2019.