

Design And Determination Of The Sample Size In Medical Research

¹Arif Habib, ²Ayman johargy, ³Khalid Mahmood, ⁴Humma

¹Investigator, RRIUM, Naseembagh campus-University of Kashmir.

²Department of Microbiology, Faculty of Medicine, Umm Al-Qura University.

³Scientist-IV, CCRUM, Department of AYUSH, Ministry of Health & Family Welfare, Government of India.

⁴Research officer, RRIUM, Naseembagh campus-University of Kashmir.

Abstract: This paper is designed as a tool that a researcher could use in planning and conducting quality research. This is a review paper which gives a discussion of various aspects of designing consideration in medical research. This paper covers the essentials in calculating power and sample size for a variety of applied study designs. Sample size computation for survey type of studies, observation studies and experimental studies based on means and proportions or rates, sensitivity – specificity tests for assessing the categorical outcome are presented in detail. Over the last decades, considerable interest has been focused on medical research designs and sample size estimation. The resulting literature is scattered over many textbooks and journals. This paper presents these methods in a single review and comments on their application in practice.

Keywords: Types of studies, sample size determination, inferential statistics, sensitivity, specificity, odd ratios.

I. Background:

In medical research, we attempt to study the multitude of constantly changing and interrelated biologic processes that comprise human physiology. In order to make meaningful conclusions from the abundance of physiologic data available, we need to carefully consider the design of our investigations. We must meticulously define our study hypotheses, patient population, and research methods for our conclusions to be valid and our study a useful addition to the medical literature.

Before data collection ever begins, therefore, much of the work involved in medical research. Failure to exercise such attention to detail and planning may result in faults in study design that are subsequently propagated throughout the study and impact on each step of the research process. No amount of statistical manipulation can correct for errors and biases introduced by a poorly designed study. The goal in designing any research study is to avoid systematic errors and biases as it is much easier to correct flaws in study design prior to beginning a study rather than during or after concluding a study.

The planning of a good research has many aspects. First step is to define the problem and it should be operational. Second step is to define the study type and the appropriate subjects and controls. Meticulously, one has to define the inclusion and exclusion criteria, which should take care of all possible variables which could influence the observations and the units which are measured. The study design must be clear and the procedures are defined to the best possible and available methodology. Based on these factors, the study must have an adequate sample size, relative to the goals and the possible variabilities of the study.

Let us discuss inferential statistics which has two parts: 1. Estimation of population parameter and 2. Testing of hypothesis. According to the type of medical research, any one of them can be adopted. The estimation method is used in prevalence/descriptive studies and the testing of hypothesis is used for cohort/case control/clinical trials. Descriptive studies are designed to describe the occurrence of disease by time, place and person. Descriptive study deals with estimation of population parameters. Two commonly used parameters are the mean (measure of central tendency) and the proportion. Using estimation method, the best estimates for population characteristics such as prevalence, incidence, mean, standard deviation, etc. can be found out. By testing the hypothesis, correctness of whatever values or any relationship or association between variables derived from estimation can be verified. These are the two requirements for the analysis of data in medical research. Before the testing of the hypothesis, one must confirm the type of normality of the data so that the type of the test (parametric or non parametric) can be decided. Violation of this rule will result in wrong conclusion. Once the correct test is selected, the next important step is to determine the sample size.

It may be note that sample size calculations may not be required for some pilot or exploratory studies. A pilot study is a preliminary study intended to test the feasibility of a larger study, data collection methods, collect information for sample size calculations, and therefore should always have a main study to which it leads. In other words, pilot studies cannot exist on their own, but only in relation to larger studies with the aim to facilitate the larger studies.

II. Analytical study:

An analytical study is one in which action will be taken on a cause system to improve the future performance of the system of interest. An analytical study focuses on prediction. Because of the temporal nature of improvement, the theory and methods for analytical studies are a critical component of the science of improvement. We search for the determinants of health outcomes, first, by relying on descriptive epidemiology to generate hypotheses about associations between exposures and outcomes. Analytic studies are then undertaken to test specific hypotheses. Samples of subjects are identified and information about exposure status and outcome is collected. The essence of an analytic study is that groups of subjects are compared in order to estimate the magnitude of association between exposures and outcomes. The two main types of analytic study design are the Experimental studies (clinical studies) and the observational studies, although there are several variations on these general designs.

Experimental studies involve the administration of an intervention or treatment to two or more groups of patients with attention being directed to identifying the impact the intervention has on a particular outcome. Experimental studies, through demonstrating a response to therapy, may prove “causation”. A **clinical trial** is an experimental study that involves a clinically applicable therapy. Clinical trials generally involve the comparison of an experimental drug or therapy against a control (standard therapy or placebo). The Food and Drug Administration categorizes these into **Phase I**, **Phase II**, and **Phase III** trials. A **Phase I** trial is the initial introduction of a drug or treatment to humans. It identifies any associated toxicity and may begin to determine dosing levels and efficacy. Phase I trials are small, usually involving only 10 to 20 patients. A **Phase II** trial demonstrates the effectiveness and safety of a drug or treatment at specific dosages. These closely monitored trials involve 100 to 200 patients. After demonstrating a treatment’s efficacy, an expanded **Phase III** trial is performed to identify the specific indications and uses of the drug or treatment as well as more precise information.

We can also subdivide research studies by the time course of data collection. **Longitudinal studies** involve the analysis of data collected over an extended period of time and are thus able to identify the changing nature of a disease process. **Prevalence studies**, on the other hand, involve the analysis of data obtained at a single point in time and are useful in identifying the presence of disease in a particular patient population. **Retrospective studies** are longitudinal studies that look back in time to evaluate the factors affecting a particular variable of interest. **Prospective studies** are longitudinal studies that look forward in time to analyze the impact of an intervention or disease process on the outcome variable(s) of interest. Retrospective studies are frequently easier to perform, but are more subject to potential errors and statistical biases. Prospective studies are usually more difficult to perform, more expensive, and more time consuming, especially if the disease of interest is rare. They provide more evidence for causality, however, than does a retrospective study.

Experimental studies prospectively evaluate the efficacy of an intervention or treatment against another treatment. They can be either **uncontrolled** or **controlled**. **Uncontrolled studies** describe the effect of an intervention or treatment in a single group of patients. If the natural course of a disease process is predictable, a control group is less important. A study on the value of intravascular fluids in hypotensive shock, for example, would not necessarily need a control group. Without fluids, a hypotensive patient will almost certainly die (the natural course of the disease is predictable). An uncontrolled study design would therefore be appropriate in this circumstance.

If the natural course of disease is to improve spontaneously, however, a control group is important. We might, for example, perform an uncontrolled study to evaluate the effect of a drug on the common cold. When we analyze our results, however, we may interpret patient improvement as being due to our drug when, in reality, the patients were likely to improve regardless of our therapy. In this situation, we need to be able to compare the patients who received the drug with patients who did not to be able to make valid conclusions regarding the clinical benefit of the drug. Uncontrolled studies are also subject to the “**Hawthorne effect**”, which is the tendency of patients to change their behaviour as a result of a study. In our hypothetical study on the common cold, for example, some patients, wishing to please their doctors, might report fewer symptoms and therefore appear improved. Because their attention was focused on their colds as a result of the study, they might also take care to get more sleep, eat properly, or drink more fluids. Any of these factors might result in improvement in their symptoms which would appear to be due to our drug, but, in fact, would not be. A controlled study would thus be necessary to account for these potential confounding variables.

“**Regression towards the mean**” is another potential error to consider in an uncontrolled study. This refers to the fact that there is a certain degree of error associated with any physiologic measurement we make. Suppose, for example, that we decide to study the effect of a drug on blood pressure and include only those patients with systolic blood pressures (SBP) over 150 mmHg. Some patients will truly have a SBP over 150 mmHg. Others, however, will have an erroneously high initial SBP due to measurement error and will subsequently have normal SBP measurements. In an uncontrolled study, we might interpret this apparent decrease in mean SBP (a “regression towards the mean”) as being due to the treatment drug. In reality, however, it is due to the fact that

our sample population does not accurately represent the population we wished to study (i.e., true hypertensive patients). A controlled study theoretically evenly distributes the effect of regression towards the mean between the study and control groups, and negates its confounding effect on study conclusions. A final error that may occur with uncontrolled trials is the "placebo effect". Some patients who receive a drug may believe that the drug is making them better when it is not. In an uncontrolled trial, such as our common cold study where the outcome is largely subjective, it may be difficult to separate placebo effect from true clinical improvement.

This is not to say that uncontrolled studies have no role in medical research. In situations where the outcome without therapy is clearly predictable, an uncontrolled trial may be perfectly appropriate. Uncontrolled studies are a frequent first step in the preliminary evaluation of a therapy to determine safety and technical information. The results of these preliminary studies may then suggest the need for larger controlled clinical trials.

Controlled studies compare the impact of an intervention on two or more groups of patients. They compare the treatment group which receives the intervention to a control group which does not or which receives a different intervention that is frequently the current standard of treatment. Properly designed controlled studies equally distribute the potential sources of error discussed for uncontrolled studies between both the treatment and control groups. The effect of these errors will then tend to "cancel" each other out in the statistical analysis. Controlled studies are therefore statistically stronger than uncontrolled trials and are frequently necessary to adequately prove causality. Consider a study in which we wish to demonstrate whether crystalloid or colloid administration is superior in reversing hypotension in the same patients. It is not immediately obvious whether crystalloid or colloid is the better choice of resuscitation fluid. To perform this study, therefore, we would need to perform a controlled study where we can compare the effect of one intervention (crystalloid) with another (colloid) to determine which therapy is superior. **E.g.** To evaluate a new back school, patients with lower back pain were randomly allocated to either the new school or to conventional occupational therapy. After 3 months they were questioned about their back pain, and observed lifting a weight by independent monitors.

In the ideal controlled study, both the treatment and control groups are exactly the same except for the intervention to be received. In this way, we can attribute any differences that we detect between the groups to the intervention alone. We also minimize the effect of statistical bias and confounding variables on the validity of our conclusions. In reality, however, it is impossible to create two groups which are exactly alike, and the goal of designing a study is to minimize the group differences as much as possible.

Observational studies involve the surveillance of one or more groups of patients to determine the effect of various patient characteristics on one or more outcome variables. These patients, however, receive no intervention or treatment. Observational studies demonstrate the "association", if any, between the patient characteristics and outcome variables of interest. **Case-series, case-control, cross-sectional, and cohort studies** are all examples of observational studies.

Case-series studies are simply descriptive reports illustrating observations of interest on one or more, usually consecutive, patients. There are no control patients involved or initial hypotheses presented, although such studies frequently result in hypotheses that lead to further studies. Because they describe the disease course of the patients involved, they are longitudinal in design.

Case-control studies are retrospective studies that identify a disease or outcome variable and look back in time to determine the risk factors that led to the disease. They compare a control population (patients without the disease) with a case population (patients with the disease) to identify factors that make the two groups different and may account for the occurrence of the disease. **E.g.** To investigate the relationship between egg consumption and heart disease, a group of patients admitted to hospital with myocardial infarction were questioned about their egg consumption. A group of age and sex matched patients admitted to a fracture clinic were also questioned about their egg consumption using an identical protocol.

Cross-sectional studies are prevalence studies which collect data at a discrete point in time in order to answer particular questions about the status of the population at that instant. Disease prevalence studies, surveys, questionnaires, and meta-analyses are examples of cross-sectional studies. Cross-sectional studies are also frequently used to show "association" and suggest future studies. **E.g.** A survey was conducted of all nurses employed at a particular hospital. Among other questions, the questionnaire asked about the grade of the nurse and whether she was satisfied with her career prospects.

Cohort studies are longitudinal studies which follow a group of patients with a common characteristic and collect data prospectively to answer questions related to the outcome variables of interest. This is the strongest of the observational study designs in that it is the least subject to statistical biases and errors in data recall and collection due to its prospective nature. Studies that are intended to determine the natural history of a particular disease process, **e.g.** to investigate the relationship between certain solvents and cancer; all employees of a particular organisation were questioned about their exposure to an industrial solvent, and the amount and length

of exposure measured. These subjects were regularly monitored, and after 10 years a copy of the death certificate for all those who had died was obtained.

III. Sample Size:

Study design has a major impact on the sample size. Descriptive studies need hundreds of subjects to give acceptable confidence interval for small effects. Experimental studies generally need lesser sample while the cross-over designs needs one-quarter of the number required compared to a control group because every subject gets the experimental treatment in cross-over study. An evaluation studies in single group with pre-post type of design needs half the number for a similar study with a control group. A study design with one-tailed hypothesis requires 20% lesser subjects compared to two-tailed studies. Non-randomized studies needs 20% more subjects compared to randomized studies in order to accommodate confounding factors. Additional 10 - 20% subjects are required to allow adjustment of other factors such as withdrawals, missing data, lost to follow-up etc.

The “outcome” expected under study should be considered. There are 3 possible categories of outcome. The first is a simple case where 2 alternatives exist: Yes/no, death/alive, vaccinated/not vaccinated, etc. The second category covers multiple, mutually exclusive alternatives such as religious beliefs or blood groups. For these 2 categories of outcome, the data are generally expressed as percentages or rates. The third category covers continuous response variables such as weight, height, blood pressure, VAS score, IL6, TNF-a, homocysteine etc, which are continuous measures and are summarized as means and standard deviations. The statistical methods appropriates the sample size based on which of these outcomes measure is critical for the study, for example, larger sample size is required to assess the categorical variable compared to continuous outcome variable.

In the medical studies, many parameters of interest are compute in order to assess the effect size of healthcare interventions based on 2×2 contingency table and categorical variables. If we look for example at diagnostic studies, we have the next parameters that can be express as a proportion: sensitivity, specificity, positive predictive value, negative predictive value, false negative rate, false positive rate, probability of disease, probability of positive test wrong, probability of negative test wrong, overall accuracy, probability of a negative test, probability of a positive test For the therapy studies experimental event rate, control event rate, absolute risk in treatment group, and absolute risk on control group are the proportion-like parameter which can be compute based on 2×2 contingency table. If we look at the risk factor studies, the individual risk on exposure group and individual risk on nonexposure group are the parameters of interest because they are also proportions. Whatever the measure used, some assessment must be made of the trustworthiness or robustness of the findings, even if we talk about a p-value or confidence intervals. The confidences intervals are preferred in presenting of medical results because are relative close to the data, being on the same scale of measurements, while the p-values is a probabilistic abstraction.

Calculation of the number of patients or **sample size** necessary to perform a study should be one of the initial steps in study design. If such calculations (also known as **power analysis**) are not performed and insufficient data are collected, we may not have the statistical power to make accurate conclusions and we will have performed a useless study. At the same time, if we study more patients than are necessary to demonstrate a statistically significant difference, we will have performed a more expensive and time consuming study than was necessary.

Generic expression for calculating sample size

$$\text{Sample size} = \frac{(\text{power, Inverse function of significance level})}{(\text{absolute difference})^2}$$

To calculate the sample size necessary for a particular study, we must have to be very clear of the study type and answer of the requirements.

The significance level: or alpha, is the probability of making a Type I error (i.e., rejecting the null hypothesis when there is no difference). The chosen level of significance sets the likelihood of detecting a treatment effect when no effect exists (leading to a so-called “false-positive” result) and defines the threshold “P value”. Results with a P value above the threshold lead to the conclusion that an observed difference may be due to chance alone, while those with a P value below the threshold lead to rejecting chance and concluding that the intervention has a real effect. Most researchers use significance at level 5% (that is, P = 0.05) or 1% (P = 0.01). This means the investigator is prepared to accept a 5% (or 1%) chance of erroneously, that is, they wish to have 95% or 99% confidence that their conclusions are accurate and that they are not committing a Type I error.

Statistical power: quantifies the ability of a study to detect a clinically significant difference when present. Power is defined as $1 - \beta$, where β is the probability of making a Type II error (i.e., rejecting the alternate hypothesis when there is a difference). This is usually chosen to be 80%. By definition, a study power set at 80% accepts a likelihood of one in five (that is, 20%) of missing such a real difference. Thus, the power for large trials is occasionally set at 90% to reduce to 10% the possibility of a so-called “false-negative” result. It is important to remember that as we decrease the desired significance level or increase the desired power, we must increase our sample size.

In other words, the difference between 2 groups in a study will be explored in terms of estimate of effect, appropriate confidence interval, and P value. The confidence interval indicates the likely range of values for the true effect in a population while P value determines how likely it is that the observed effect in the sample is due to chance. A related quantity is the statistical power of the study, is the probability of detecting a predefined clinical significance. The ideal study is the one, which has high power. This means that the study has a high chance of detecting a difference between groups if it exists, consequently, if the study demonstrates no difference between the groups, the researcher can reasonably confident in concluding that none exists. The ideal power for any study is considered to be 80%.

In research, statistical power is generally calculated with 2 objectives. i) It can be calculated before data collection based on information from previous studies to decide the sample size needed for the current study. ii) It can also be calculated after data analysis. The second situation occurs when the result turns out to be non-significant. In this case, statistical power is calculated to verify whether the non-significance result is due to lack of relationship between the groups or due to lack of statistical power.

Statistical power is positively correlated with the sample size, which means that given the level of the other factors viz. alpha and minimum detectable difference, a larger sample size gives greater power. However, researchers should be clear to find a difference between statistical difference and scientific difference. Although a larger sample size enables researchers to find smaller difference statistically significant, the difference found may not be scientifically meaningful. Therefore, it is recommended that researchers must have prior idea of what they would expect to be a scientifically meaningful difference before doing a power analysis and determine the actual sample size needed. Power analysis is now integral to the health and behavioral sciences, and its use is steadily increasing whenever the empirical studies are performed.

Effect size: The difference that is considered to be clinically important is determined by clinical experience. A treatment which results in an increase in cardiac index of 200 mL is probably not clinically important whereas an increase of 1000 mL is. The smaller the clinically important difference, the more difficult it will be to prove statistically, and the larger the sample size necessary. Similarly, the larger the difference, the easier it will be to prove, and the smaller the sample size necessary. The obvious question is how to know the difference in a study, which is not conducted. If available, it may be useful to use the effect size found from prior studies. Where no previous study exists, the effect size is determined from literature review, logical assertion, and conjecture.

Variance or standard deviation: The variance or standard deviation for sample size calculation is obtained either from previous studies or from pilot study. Larger the standard deviation, larger is the sample size required in a study. For example, in a study, with primary outcome variable is TNF-a, needs more subjects compared to a variable of birth weight, 10-point Vas score etc. as the natural variability of TNF-a is wide compared to others.

Effect of compliance is another factor that directly affects the sample size. So, it should be calculated correctly.

The compliance adjustment formula is as follows: Adjusted sample size $n_1 = \frac{n}{[c_1 - c_2 - 1]^2}$ per group

equals where n is the original sample size, and c_1, c_2 are the average compliance rates per group. In addition to the above factors, other factors that affect the sample size include consideration for unequal allocation, effect of important clinical treatment, etc. One of the most important decisions to make before calculating a sample size is to define the effect of important clinical treatment, δ (delta), which should not be confused with a statistical significance of the treatment effect – neither one implies the other and the distinction between them is important.

IV. Withdrawals, missing data and losses to follow-up:

Sample size calculated is the total number of subjects who are required for the final study analysis. There are few practical issues, which need to be considered while calculating the number of subjects required. It is a fact that all eligible subjects may not be willing to take part and may be necessary screen more subjects than the final number of subjects entering the study. In addition, even in well-designed and conducted studies, it is unusual to finish with a dataset, which is complete for all the subjects recruited, in a usable format. The reason could be subject factor like- subjects may fail or refuse to give valid responses to particular questions, physical measurements may suffer from technical problems, and in studies involving follow-up (eg. Trials or cohort studies), there will be some degree of attrition. The reason could be technical and the procedural problem- like

contamination, failure to get the assessment or test performed in time. It may, therefore, necessary to consider these issues before calculating the number of subjects to be recruited in a study in order to achieve the final desired sample size.

Example, say in a study, a total of n number of subjects are required in the end of the study with all the data being complete for analysis, but a proportion (q) are expected to refuse to participate or drop out before the study ends. In this case, the following total number of subjects (n^*) would have to be recruited to ensure that the final sample size (n) is achieved:

$$n^* = \frac{n}{1 - q}, \text{ where } q \text{ is the proportion of attrition and is generally } 10\%,$$

The proportion of eligible subjects who will refuse to participate or provide the inadequate information will be unknown at the beginning of the study. Approximate estimates is often possible using information from similar studies in comparable populations or from an appropriate pilot study.

V. Sample size estimation for proportion in survey type of studies:

A common goal of survey research is to collect data representative of population. The researcher uses information gathered from the survey to generalize findings from a drawn sample back to a population, within the limits of random error. The general rule relative to acceptable margins of error in survey research is 5 - 10%.

The sample size can be estimated by using formula

$$n = \frac{(Z_{\alpha/2})^2 \times P(1 - P) D}{E^2},$$

Where P is the prevalence or proportion of event of interest for the study, E is the Precision (or margin of error) with which a researcher want to measure something. Generally, E will be 10% of P and $Z_{\alpha/2}$ is normal deviate for two-tailed alternative hypothesis at a level of significance; for example, for 5% level of significance, $Z_{\alpha/2}$ is 1.96 and for 1% level of significance it is 2.58. D is the design effect reflects the sampling design used in the survey type of study. This is 1 for simple random sampling and higher values (usually 1 to 2) for other designs such as stratified, systematic, cluster random sampling etc, estimated to compensate for deviation from simple random sampling procedure. The design effect for cluster random sampling is taken as 1.5 to 2. For the purposive sampling, convenience or judgment sampling, D will cross 10. Higher the D , the more will be sample size required for a study. Simple random sampling is unlikely to be the sampling method in an actual filed survey. If another sampling method such as systematic, stratified, cluster sampling etc. is used, a larger sample size is likely to be needed because of the "design effect. In case of impact study, P may be estimated at 50% to reflect the assumption that an impact is expected in 50% of the population. A P of 50% is also a conservative estimate; Example: Researcher interested to know the sample size for conducting a survey for measuring the prevalence of HT in certain community. Previous literature gives the estimate of an HT at 20% in the population to be surveyed, and assuming 95% confidence interval or 5% level of significance and 10% margin of error, the sample size can be calculated as follow as;

$$n = \frac{(1.96)^2 \times 0.20(1 - 0.20) \times 1}{(0.1 \times 0.20)^2} = 1537 \text{ for a simple random sampling design. Hence, sample size of } 1537$$

is required to conduct community-based survey to estimate the prevalence of Hyper Tension. Note- E is the margin of error, in the present example; it is $10\% \times 0.20 = 0.02$.

To find the final adjusted sample size, allowing non-response rate of 10% in the above example, the adjusted sample size will be $1537/(1-0.10) = 1537/0.90 = 1708$.

Sample size calculation, when mean is the parameter of study:

The confidence interval contains an estimate, above or below a margin of error. The margin of error for a 95% confidence interval is 1.96 times the standard error. It shows the accuracy of the guess and is based on the variability of the estimate.

Let E denote the margin of error. Then

$$n = \frac{(1.96)^2 \sigma^2}{E^2}$$

For 99% confidence interval

$$n = \frac{(2.58)^2 \sigma^2}{E^2}$$

e.g. The mean pulse rate of a population is believed to be 72 per minute with a standard deviation of 10 beats. Calculate the minimum sample size required to verify this if allowable error is 2 beat at 2% risk.

$$n = \frac{(1.96)^2 \sigma^2}{E^2} \rightarrow = \frac{(1.96)^2 10^2}{2^2} \rightarrow \frac{3.84 \times 100}{4} = 96.$$

$$n = \frac{(2.58)^2 \sigma^2}{E^2} \rightarrow = \frac{(2.58)^2 10^2}{2^2} \rightarrow \frac{6.65 \times 100}{4} = 166.$$

Sample Size calculation, when proportion is the parameter of study:

For 95% confidence

$n = \frac{(1.96)^2 PQ}{E^2}$, where P is the population proportion and $Q = 1 - P$. Also if E is given as a percentage,

then it is to be taken as the percentage of p .

E.g. Giardiasis among children prevalence rate was 20% before the specific treatment and adoption of other measures. Find out the size of sample required to find the prevalence rate now if available error is 20% at 5% risk.

$$n = \frac{(1.96)^2 PQ}{E^2}, \quad E = 20 \times \frac{10}{100} = 2$$

$$n = \frac{(1.96)^2 20 \times 80}{2^2}, \rightarrow = \frac{3.8 \times 1600}{4} = 1520$$

Calculate sample size for a sensitivity of a Test:

Based on the literature review identify the sensitivity and specificity of the diagnostic test being study-

For sensitivity

$T_+ + F_- = (Z_{\alpha/2})^2 \frac{\{S_n (1 - S_n)\}}{E^2}$, where T_+ is true positive, F_- is false negative S_n is sensitivity and E accuracy which is usually taken 0.05.

$n S_n = \frac{T_+ + F_-}{p}$, where p is the prevalence of the disease in test population.

For Specificity

$F_+ + T_- = (Z_{\alpha/2})^2 \frac{\{S_p (1 - S_p)\}}{E^2}$, where F_+ is false positive, T_- is true negative S_p is specificity.

$n S_p = \frac{F_+ + T_-}{(1 - p)}$.

E.g. Let sensitivity be 85% and prevalence be 28% , using above formula the sample size is as follows

$T_+ + F_- = (1.96)^2 \frac{\{0.85 (1 - 0.85)\}}{0.05^2} = 195.9$

$n S_n = \frac{195.9}{0.28} = 699.79$ can be taken as **700**.

Now take specificity as 70% and prevalence 28%

$F_+ + T_- = (1.96)^2 \frac{\{0.70 (1 - 0.70)\}}{0.05^2} = 322.72$ can be taken as **323**.

Note: if the researcher is interested in both sensitivity and specificity, than take the higher number. In the above example we can take 700 sample size.

Determination of sample size for experimental studies:

a) Formula for difference in proportion:

$$n = \frac{(Z_{\beta} + Z_{\alpha/2})^2 \times 2\bar{P}(1 - \bar{P})}{E^2}$$

b) Formula for difference in means:

$$n = \frac{(Z_{\beta} + Z_{\alpha/2})^2 2\sigma^2}{E^2}$$

where $\bar{P}(1 - \bar{P})$ is a measure of variability (similar to standard deviation) and $\bar{P} = \frac{(P_1 + P_2)}{2}$.

$E = p_1 - p_2$ is the effect size (the difference in means or proportions).

σ is the standard deviation of the outcome variable.

c) If unequal number in each group:

Suppose a researcher wish to randomize n patients to the experimental group and sn to the to the control group where $0 < s < n$. i.e. one would like to study only half as many experimental as control group subjects. Then (ignoring the continuity correction) the required formula of sample size is

$$n = \frac{\left\{ \left[(s+1) \bar{P}_S (1 - \bar{P}_S) \right]^{1/2} Z_{\alpha/2} + \left[s P_e (1 - P_e) P_c (1 - P_c) \right]^{1/2} Z_{\beta} \right\}^2}{s E^2},$$

where $\bar{P}_S = \frac{P_e + s P_c}{s + 1}$.

Example:- Investigators anticipate a 5-year death rate among control patients of about 40% . They view a reduction of this mortality rate to 30% among experimentally treated patients as clinically important. They would like to detect such a reduction with 80 per cent power and with a two-tailed test at the 5 per cent significance level. Thus $P_c = 0.40$, $P_e = 0.30$, $\bar{P} = (0.40 + 0.30)/2 = 0.35$, $Z_{\alpha/2} = 1.96$, $Z_{\beta} = 0.84$.

First we calculate $\bar{P}_S = \frac{0.30 + 2 \times 0.40}{2 + 1} = 0.36.6$, we can take 0.37 and using formula

$$n = \frac{\left\{ \left[(2+1) 0.37 (1 - 0.63) \right]^{1/2} 1.96 + \left[2 (0.30) (1 - 0.30) 0.40 (1 - 0.40) \right]^{1/2} 0.84 \right\}^2}{2 (0.30 - 0.40)^2} = 122.84,$$

Thus the required number of patients required in the control group is 2 (123)=246 and in a experimental group is 123 patients. Total sample size of a trail will be 246 + 123 = 369.

Determination of sample size for observational studies:

a) when difference in proportions (for binary exposure):

$$n = \frac{(\bar{p})(1 - \bar{p})(Z_{\alpha/2} + Z_{\beta})^2}{E^2} \left[\frac{r+1}{r} \right], \text{ where } E = P_1 - P_2 \text{ and } r \text{ is ratio of controls to cases.}$$

In the analysis stage, calculate the frequency of each of the measured variables in each of the two groups. As a measure of the strength of the association between an exposure and the outcome, case-control studies yield the odds ratio. An **odds ratio** is the ratio of the odds of an exposure in the case group to the odds of an exposure in the control group.

Example: Let the researcher want to detect an odd ratio (OR) of 2.5 or greater for 80% power and 1.96 level of significance. Let $r=1$ i.e equal number of cases and controls. Consider the proportion exposed in the control group is 30%, to get proportion of cases exposed, than

$$P_{CE} = \frac{OR \times P_{CS}}{1 + (OR - 1) P_{CS}}$$

$$P_{CE} = \frac{(2.5) \times (0.30)}{1 + (2.5 - 1) (0.30)} = \frac{0.75}{1.45} = 0.51$$

Average proportion exposed = $(0.51 + 0.30/2) = 0.40$

$$n = \frac{(0.35)(1 - 0.35)(1.96 + .84)^2}{(0.40 - 0.30)^2} \left[\frac{1+1}{1} \right] = \frac{3.5672}{0.01} = 356.7 \text{ or } 357$$

$n = 357$, so 357 for cases and 357 for controls

Sample size for this study is 714.

b) When difference in means (for continuous exposure):

$$n = \frac{\sigma^2 (Z_{\alpha/2} + Z_{\beta})^2}{E^2} \left[\frac{r+1}{r} \right],$$

where σ is the standard deviation of outcome variable.

Example:- For 80% power and 95% significance level, $r = 1$ (equal no. of cases and controls) with difference 5.0 and deviation 10.0.

$$n = \frac{\sigma^2 (Z_{\alpha/2} + Z_{\beta})^2}{E^2} \left[\frac{r+1}{r} \right],$$

$$n = \frac{10^2 (7.84)^2}{5^2} [2] = 63,$$

$n = 63$, so 63 for cases and 63 for controls

Sample size for this study is 126.

c) Sample size for Cohort study:

$$n' = \frac{\left[Z_{\alpha/2} \sqrt{(r+1) \bar{p} \bar{q}} - Z_{\beta} \sqrt{r p_1 q_1 + p_2 q_2} \right]^2}{r E^2}$$

$$n_1 = \frac{n'}{4} \left[1 + \sqrt{\frac{2(r+1)}{n' r |E|}} \right]^2$$

Where $n' = n_1$ and $n_2 = r n_1 \rightarrow r = \frac{n_2}{n_1}$ ratio of cases to control and $\bar{p} = \frac{(p_1 + r p_2)}{r+1}$. Also $q = 1 - p$.

$E = p_2 - p_1$ is the effect size (the difference in means or proportions).

Example:- Over weight have high risk of Diabetic Mellinitus . From literature review identify the rate of disease among those with and with out of the risk factor.

Ratio of unexposed vs exposed 1:1

Proportion of sample from no risk (normal) population = 50%.

Proportion of sample from at risk (over weight) population = 50%

p_1 = true proportion of DM in no risk (normal) population = 7%

p_2 true proportion of DM in at risk (over wt.) population = 32%

By using above formula we have

$$n' = \frac{\left[1.96 \sqrt{(1+1)(0.195)(0.805)} + 0.842 \sqrt{1(0.07)(0.93) + (0.32)(0.68)} \right]^2}{1(0.32 \times 0.07)^2} = \frac{2.3898}{0.0625} = 38.23$$

$$n_1 = \frac{38.23}{4} \left[1 + \sqrt{\frac{2(r+1)}{38.23 \times 1 \times 0.25}} \right]^2 = 45.88 \text{ or } 46.$$

$$n_1 = 46, \quad n_2 = r n_1 = 1 \times 46 = 46.$$
$$n = n_1 + n_2 = 46 + 46 = 92$$

VI. Discussions:

Carefully and well planned Medical research will result in relevant and socially useful results. Planning has several parts, such as well defined relevant research hypothesis, objectives, subjects must be selected from appropriate population, and instruments should be reliable, carefully undergone through best possible procedures and other guidelines. It requires the collaboration of a specialist who has a good scientific knowledge in the art and practice of medical statistics. Conducting a study that has little chance of answering the hypothesis at hand is a misuse of time and valuable resources and may unnecessarily expose participants to potential harm or unwarranted expectations of therapeutic benefits. As scientific and ethical issue go hand-in-hand, the awareness of determination of minimum required sample size and application of appropriate sampling methods are extremely important in achieving scientifically and statistically sound results.

It is also recommended that a professional statistician must be included in the research protocol, projects; scientific studies etc from beginning to the end of the study because statistician and medical professionals have lot in the common. Physician examine the patients take samples assess symptoms run diagnostics and write up reports for patient records where as statisticians helps in planning, handle data examine the data for sampling run statistical diagnostics like graphical representation applying statistical tools to check significance and write up reports/results .

A Sample size calculation is an essential step in research protocols and is a must to justify the size of clinical studies in papers, reports etc. Nevertheless, one of the most common errors in paper reporting clinical trials is a lack of justification of the sample size, and it is a major concern that important therapeutic effects are being missed because of inadequately sized studies. Using an adequate sample size along with high quality data collection efforts will result in more reliable, valid and generalizable results, it could also result in saving resources. Information on the scheduled duration of the study, any adjustment for non-compliance and any other issues that formed the basis of the sample size calculation should be included. For continuous outcomes, in particular (eg, blood pressure), assumptions made about the distribution or variability of the outcome should be explicitly stated.

Finally the sample size formulas provide the number of responses that need to be obtained. Many researchers commonly add 10% to the sample size to compensate for the persons that the researcher is unable to contact by some reasons. Thus, the number of sample size must be increased in such studies, so that desired level of confidence and precision may be achieved.

VII. Conclusion:

We first adopt new technologies in order to be more efficient or to improve what we are producing. But as those technologies mature, they drive fundamental changes in the way we approach and solve problems. Just as modern computing has transformed the ways we do data management and analysis, so is it finally transforming how we do statistical planning. Successful resolution of the sample size problem requires the close and honest collaboration of statisticians and subject-matter experts. Understanding what kind of study has been done is a prerequisite to thoughtful reading of research. Sample size planning is often important, and almost always difficult. It requires care in eliciting scientific objectives and in obtaining suitable method prior to the study. Sample size problems are context-dependent. Better designing tools will allow us to better customize our research work to fit our proposed designs and our conjectures for the underlying infinite datasets. Doing this earnestly, honestly and creatively leads to better study designs and thus better research outcomes.

Conflict of Interests: - The authors do not have any conflict of interest arising from the study.

Notations:

n	Sample size
E	Margin Error
σ	Standard Deviation
P	Population proportion, $Q = 1 - P$
r	Ratio of controls to cases
$Z_{\alpha/2}$	Level of significance (Typically 1.96 at 5%)
Z_{β}	Desired power (Typically 0.84 for 80%)
$(\bar{P})(1 - \bar{P})$	Measure of Variability (Similar to standard deviation)

$$\bar{P} = \frac{P_1 + P_2}{2}$$

References

- [1] Hintze JL. Kaysville, Utah, USA: 2008. Power analysis and sample size system (PASS) for windows User's Guide I. NCSS.
- [2] Jones S R, Carley S & Harrison M. "An introduction to power and sample size estimation" *Emergency Medical journal* 2003; 20; 453-458,2003.
- [3] GebSKI V, Marschner I, Keech AC. Specifying objectives and outcomes for clinical trials. *Med J Aust* 2002; 176(10): 491-2.
- [4] Wittes J. Sample size calculations for randomized controlled trials. *Epidemiol Rev.* 2002;24:39-53.[PubMed].
- [5] Lui KJ, Cumberland WG. Sample size determination for equivalence test using rate ratio of sensitivity and specificity in paired sample data. *Control Clin Trials* 2001;22(4):373-89. 23.
- [6] Donner A, Klar N. Design and analysis of cluster randomization trials in health research. London, UK: Arnold Publishing Co, 2000.
- [7] Gore SM. Statistics in question. Assessing clinical trials--trial size. *Br Med J (Clin Res Ed)* 1981;282(6277):1687-9. 5. Boen JR, Zahn DA. *The Human Side of Statistical Consulting*. Belmont, CA: Lifetime Learning Publications, 1982.
- [8] Blackwelder, W. C. 'Proving the null hypothesis in clinical trials', *Controlled Clinical Trials*, 3, 345-354 (1982).
- [9] Fleiss, J. L. *Statistical Methods for Rates and Proportions*, Wiley, New York, 1981.
- [10] Lachin, J. M. 'Introduction to sample size determination and power analysis for clinical trials', *Controlled ofclinical Trials* 2, 93-114 (1981).
- [11] Friedman, L. M., Furberg, C. D. and DeMets D. L. *Fundamentals of clinical Trials*, John Wright, Boston, 1981.
- [12] Schwartz, D., Flamant, R. and Lellouch, J. *Clinical Trials*, Academic Press, London, 1980.
- [13] Altman DG. Statistics and ethics in medical research: III How large a sample?. *Br Med J* 1980;281(6251): 1336-8.
- [14] Freiman, J. A., Chalmers, T. C., Smith, H. Jr. et al. 'The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: survey of "negative" trials', *New England Journal of Medicine*, 299, 690-694 (1978).
- [15] Lachin, J. M. 'Sample size determinations for r x c comparative trials', *Biometrics*, 33, 315-324 (1977).
- [16] Feinstein, A. R. *Clinical Biostatistics*, C. V. Mosley, Saint Louis, 1977.
- [17] Colton, T. *Statistics in Medicine*, Little, Brown, Boston, 1974.
- [18] Cox, D. R. *The Analysis of Binary Data*, Methuen, London, 1970.