Subscription Agency
Exploring and Distributing Knowledge Globally

International Journal of Research in Pharma & Pharmaceutical Sciences
Volume 2 Issue 1, 2022

Sample Size Determination as an Important Statistical Concept in Medical Research

Nnodim Johnkennedy[1*] and Nwaokoro Joakin Chidozie[2]
[1]Department of Medical Laboratory Science, Faculty of Health Science, Imo State University, Owerri, Nigeria
[2]Department of Public Health, Federal University of Technology, Owerri, Imo State, Nigeria

Corresponding Author
Email Id: johnkennedy23@yahoo.com

ABSTRACT
Sample size is one of the important considerations at the planning phase of medical research, but researchers are often faced with difficulties of calculating valid sample size. Many researchers frequently use inadequate sample size and this invariably introduces errors into the final findings. Many reviews on sample size estimation have focused more on specific study designs which often present technical equations and formula that are boring to statistically naïve medical researchers. Hence, these reviews on sample size estimation formula may provide basic understanding and principles to achieve valid sample size estimation.

Key words: *Sample size, determination statistical concept, Medical research.*

INTRODUCTION
Sample size is a statistical concept used for defining the number of individuals included in a research study to represent a population. The sample size references the total number of respondents included in a study, and the number is often broken down into sub-groups by demographics such as age, gender, and location so that the total sample achieves represents the entire population [1]. Indeed getting the right sample size is one of the most irrelevant factors in statistical analysis. If the sample size is too small, it will not give valid results or adequately represent the realities of the population being studied. But, while larger sample sizes give smaller margins of error and are more representative, a sample size that is too large may significantly increase the cost and time taken to conduct the research [2].This considerations to put in place when determining sample size include confidence interval and confidence levels [3].

Confidence Interval (Margin of Error)
Margin of Error also called confidence interval is the amount of error that your survey's findings can tolerate. Confidence intervals measure the degree of uncertainty or certainty in a sampling method and how much uncertainty there is with any particular statistic [4]. In simple terms, the confidence interval tells you how confident you can be that the results from a study reflect what you would expect to find if it were possible to survey the entire population being studied. The confidence interval is usually a plus or minus (±) figure. For example, if your confidence interval is 5 and 90% percent of your sample picks an answer, you can be confident that if you had asked the entire population, between 85% (90-5) and 95% (90+5) would have picked that answer [5].

Confidence Level

Confidence Level**,** tells you how sure you can be of your results. The confidence level refers to the percentage of probability, or certainty that the confidence interval would contain the true population parameter when you draw a random sample many times. It is expressed as a percentage and represents how often the percentage of the population who would pick an answer lies within the confidence interval. For example, a 95% confidence level means that should you repeat an experiment or survey over and over again, 95 percent of the time, your results will match the results you get from a population [6].

The larger the sample size, the more confident you can be that their answers truly reflect the population. In other words, the larger your sample size for a given confidence level, the smaller your confidence interval [7].

Standard Deviation

Another critical measure when determining the sample size is the standard deviation, which measures a data set's distribution from its mean. Standard deviation is the measure of dispersion or variability in the data. While calculating the sample size an investigator needs to anticipate the variation in the measures that are being studied. In calculating the sample size, the standard deviation is useful in estimating how much the responses one receive will vary from each other and from the mean number, and the standard deviation of a sample can be used to approximate the standard deviation of a population [8].

The higher the distribution or variability, the greater the standard deviation and the greater the magnitude of the deviation. For instance, once you have already sent out your survey, how much variance do you expect in your responses? That variation in responses is the standard deviation [10].

Population Size

Population Size is the total number of people you are choosing your random sample from. This can, for example, be the total number of beneficiaries or the number of women living in a given district. A population is the entire group that you want to draw conclusions about. It is from the population that a sample is selected, using probability or non-probability samples. The population size may be known, or unknown, but there's a need for a close estimate, especially when dealing with a relatively small or easy to measure groups of people [11].

As demonstrated through the calculation below, a sample size of about 385 will give you a sufficient sample size to draw assumptions of nearly any population size at the 95% confidence level with a 5% margin of error, which is why samples of 400 and 500 are often used in research. However, if you are looking to draw comparisons between different sub-groups, for example, provinces within a country, a larger sample size is required [12].

Importance of Sample Size Determination in Medical Research

Essentially, sample sizes are used to represent parts of a population chosen for any given survey or experiment. Both internal and external validities of the research are ensured with an accurately estimated sample size that based on previous studies or evidences. If representativeness in a study is accurately determined, it ensures that it measured the population attributes it tends to study. In human and animal experiment, sample size is a central issue for ethical reasons. Insufficient sample size will produce scientific inference with small power [13]. This will expose subjects to potentially harmful treatments without advancing knowledge. On the other hand, oversized experiments will recruit an unnecessarily

large number of subjects into the study. This will in turn expose them to unnecessary harmful treatment. The volunteer in the study will be needlessly troubled without the study adding significant contribution to scientific knowledge [14].

Dynamics of Sample Size Determination

Some researcher's stated four categories of sample size determination depending on the aim and procedure involved, which are sample size determination, sample size justification, sample size adjustment and sample size re-estimation. Sample size determination requires the actual calculation using scientific assumption and evidence to achieve desired statistical significance of valid and reliable outcome [15]. This is the usual method which requires attributes such as prevalence, proportion and means from previous studies. Predetermined assumptions for validity and reliability such as power of study, level of significance and design effect may be needed in sample size estimation. Sample size justification is necessary when a sample size is already chosen. It becomes expedient for the researcher to provide a 'statistical justification' for the selected sample size. Usually, a small size of the population will be recruited initially due to budgetary constraints or for medical consideration [16].

Calculation of Sample Size

Sample size is involves determining the number of observations or replicates that should be included in a statistical sample. It is an essential aspect of any empirical study requiring that inferences be made about a population based on a sample. Essentially, sample sizes are used to represent parts of a population chosen for any given survey or experiment. To carry out this calculation, set the margin of error, **ε**, or the maximum distance desired for the sample estimate to deviate from the true value. To do this, use the confidence interval equation above, but set the term to the right of the ± sign equal to the margin of error, and solve for the resulting equation for sample size, **n**. The equation for calculating sample size is shown below.

$$\text{Unlimited population: } n = \frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2}$$

$$\text{Finite population: } n' = \frac{n}{1+\frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2 N}}$$

Where,
***z** is the z score*
***ε** is the margin of error*
***N** is the population size*
***p̂** is the population proportion*

For example: Determine the sample size diabetics in Owerri with 95% confidence, and a margin of error of 5%.

Assume a population proportion of 0.5, and unlimited population size. Remember that **z** for a 95% confidence level is 1.96. Refer to the table provided in the confidence level section for **z** scores of a range of confidence levels.

$$n = \frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2}$$

$$n = \frac{1.96^2 \times 0.5(1-0.5)}{0.05^2} = 384.16$$

Thus, for the case above, a sample size of at least 385 diabetics would be necessary.

How to determine the sample size using a sample calculation formula known as the Andrew Fisher's Formula.

1) Determine the population size (if known).
2) Determine the confidence interval.
3) Determine the confidence level.
4) Determine the standard deviation (a standard deviation of 0.5 is a safe choice where the figure is unknown)
5) Convert the confidence level into a Z-Score. This table shows the z-scores for the most common confidence levels:[17]

Confidence level	z-score
80%	1.28
85%	1.44
90%	1.65
95%	1.96
99%	2.58

6) Put these figures into the sample size formula to get your sample size.

$$\text{Necessary Sample Size} = \frac{(\text{Z-score})2 \times \text{StdDev} \times (1\text{-StdDev})}{(\text{margin of error})2}$$

Here is an example calculation:
Say you choose to work with a 95% confidence level, a standard deviation of 0.5, and a confidence interval (margin of error) of ± 5%, you just need to substitute the values in the formula:

((1.96)2 x .5(.5)) / (.05)2
(3.8416 x .25) / .0025
.9604 / .0025
384.16

Your sample size should be 385.

CONCLUSION

In conclusion, there are several available software like survey software to help with this calculation. You only need to input the confidence level, population size, the confidence interval, and the perfect sample size is calculated for you. Sample size estimation is an important step in conducting a valid and generalisable research.

REFERENCES

1) Kirby A, Gebski V, Keech AC. Determining the sample size in a clinical trial. Med J Aust. 2002;177:256–7.
2) Larsen S, Osnes M, Eidsaunet W, Sandvik L. Factors influencing the sample size, exemplified by studies on gastroduodenal tolerability of drugs. Scand J Gastroenterol. 1985;20:395–400.
3) Sandelowski, M. Sample size in qualitative research. Research in Nursing & Health, 1995, 18, 179–183
4) Francis, J J.; Johnston, M; Robertson, C; Glidewell, L; Entwistle, V; Eccles,M P.; Grimshaw, J M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychology & Health.2010 25 (10): 1229–1245.
5) Fugard AJB; Potts HWW. Supporting thinking on sample sizes for thematic analyses: A quantitative tool. International Journal of Social Research Methodology.201518 (6): 669–684.
6) Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. BMC Med Res Methodol 2011;11:102.
7) Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform 2014;48:193-204.
8) Bartlett, J. E.; Kotrlik, J. W.; Higgins, C. Organizational research: Determining appropriate sample size for survey research.Information Technology, Learning, and Performance Journal.2001;19 (1): 43–50.
9) Ichihara K, Boyd JC. An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537–51
10) Jennen-Steinmetz C, Wellek S. A new approach to sample size calculation for reference interval studies. Stat Med 2005; 24:3199–212.
11) Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013;35:121-6.
12) Guest, G; Bunce, A; Johnson, L (2006). How Many Interviews Are Enough?. Field Methods.2006;18: 59–82.
13) Kasiulevicius V, Šapoka V, Filipaviciute R. Sample size calculation in epidemiological studies. Gerontologija 2006;7:225-31.
14) Hazra A, Gogtay N. Biostatistics series module 5: Determining sample size. Indian J Dermatol 2016;61:496-504.
15) Zhong B. How to calculate sample size in randomized controlled trial? J Thorac Dis 2009;1:51-4.
16) Shintani A. Sample Size Estimation and Power Computation on Paired or Skewed Continuous Data; 2006. p. 1-15.
17) Noordzij M, Tripepi G, Dekker FW, Zoccali C, Tanck MW, Jager KJ. Sample size calculations: Basic principles and common pitfalls. Nephrol Dial Transplant 2010;25:1388-93.