EDITORIAL

Ann Ibd. Pg. Med 2023. Vol.21, No.1

THE BASICS OF SAMPLE SIZE ESTIMATION: AN EDITOR'S VIEW

K. I Egbuchulem

Division of Paediatric Surgery, Department of Surgery, University College Hospital, Ibadan.

The aim of this editorial is to highlight the necessary information needed; and the basic steps in estimating the minimum sample size for common study designs among resident doctors and clinicians alike.

Concepts such as sample size determination, sample size justification, sample size adjustment and re-estimation will be elucidated.

The statistical theory for sample size estimation is based on certain assumptions such as:[1]

1. The population from which the sample is drawn is infinitely large hence it will be cumbersome to study such a population.
2. The sample is selected by a simple random sampling method using a design effect.

It is noteworthy that, too few subjects make estimates unreliable and imprecise, and a study with such is poorly powered to detect the desired difference or effect.[2]

On the other hand, too many subjects amount to waste of resources with increasing risk of type I error.[2]

It is true that most clinical and hospital-based studies are quantitative research and not qualitative studies hence that will be the focus in sample size estimation.

Sample size estimation is a compromise between statistical requirements (power) and what is feasible. Such samples must be selected to obtain information which is reliable, precise, with narrow confidence interval and from which valid conclusions about the larger population can be drawn.[3]

Before setting out to estimate sample size, this research question must be answered. How many subjects do I really need to study?

To answer this question the researcher must first answer other questions that provide information about what they expect to achieve from the study.

That is, the researcher provides the parameters like the objectives and primary outcome measures that will be used for determining the sample size.

Requirements for sample size determination

1. The minimum sample size (n) of a study depends on the objective of the study. Is the primary outcome measure a quantitative or numerical variable versus categorical or qualitative variable?

The objectives of any medical research can be broadly broken down into two.

- Estimation of certain population parameters, e.g., prevalence studies.
- Comparisons of estimates of certain population characteristics between different groups (hypothesis testing).

2. Nature of the study design: e.g., observational versus experimental.
3. Number of resources available, e.g., grant funding. This tries to justify the sample size based on available resources and how common or rare the study participants are (sample size justification).
4. Plan for the statistical technique.
   a. To determine the number of participants needed to achieve a specified statistical power.
   b. It needs a good grasp of hypothesis testing in most instances.

However, power does not apply to descriptive studies and no hypothesis testing is required.

The power of a statistical test is the probability that the test will reject the null hypothesis when the null hypothesis is false.

The chances of a Type II error occurring decreases as the power of a test increases and vice versa.

The probability of a Type II error occurring is referred to as the false negative rate (β). Therefore, power is $1 - \beta$, and is the ability to identify a difference if it truly exists.

Z_{1-} for sample size estimation at different values of power (1-)

That is, a power of 60% will give a $Z_{1-\beta}$ of 0.25 and 99% will give a $Z_{1--\beta}$ of 2.33.

NB: The middle column is the power (1-β) of the study.

β	1-β	$Z_{1-\beta}$
0.500	0.500	0.000
0.400	0.600	0.250
0.300	0.700	0.530
0.200	0.800	0.840
0.150	0.850	1.030
0.100	0.900	1.280
0.050	0.950	1.650
0.025	0.975	1.960
0.010	0.990	2.330

5. Level of significance (α), which is usually set at 5% (0.05). It also has an inverse correlation with the confidence interval set for that study.[4]

Z~ for sample size estimation for selected values of ~

A one-sided hypothesis testing will make use of the values on the second column, e.g., a study to establish that scrotal exploration is superior to scrotal imaging in determining testicular torsion in teenagers. On the other hand, a two-sided hypothesis will make use of the third column, which is usually favored in most studies. E.g., there is no difference or there is a difference in the uptake rate of testicular torsion among teenagers with either scrotal exploration or scrotal imaging as an investigative modality.

	One-sided	Two-sided
α	$Z_{1-\alpha}$	$Z_{1-\alpha/2}$
0.1	1.28	1.65
0.05	1.65	1.96
0.03	1.96	2.24
0.01	2.33	2.58

6. The level of confidence (1 - α): This helps to address the question; how confident you want to be that sample estimates are as accurate as you wish. A 95% confidence interval (CI) is a popular convention.[3]

7. The degree of precision (d): The degree to which you want the sample estimate to deviate from the true population value. In determining the choice of precision (d) for a prevalence study, below is a guide:[5]
   - d = 5% (0.05) if prevalence (p) is between 10% and 90%.
   - If p < 10% (0.1), try d = 0.5 p.
   - If p > 90% (0.9), try d = 0.5 (1-p).

8. Anticipated drop-out/attrition/non-response rate. **Sample size adjustment** is usually made by accounting for the dropout rate especially for longitudinal studies or the non-response rate in cross sectional studies.
9. Is the sampling from a finite population, e.g., < 10,000?
10. And other special considerations such as multiple primary outcome measures, et cetera.

Approaches to sample size determination

a. Use of software, e.g.,
   i. Epi info. http://www.cdc.gov/Epiinfo/.
   ii. Gpower sample size determination software 3.1.
   iii. STATA.
   iv. Sample size tables, e.g., http://www.brixtonhealth.com/pepi4windows.html (WinPepi).
   v. Power and sample size: http://biostat.mc.vanderbilt.edu/wiki/main/powerSamplesize.
   vi. For qualitative research, a software like N vivo can be used.
b. Previous studies. One can use the most recent prevalence (p) from previous studies with similar design and study population.
c. If not available, an assumption of p = 0.5 or using 50% of the population.
d. Manual calculation. If a range of p is available, values close to a prevalence of 0.5 gives the higher sample size estimate.
e. A pilot study can be done before the definitive study to estimate prevalence.
f. If the sampling technique is total sampling, it therefore means there is no need for sample size determination, hence the whole population is studied.

Sample size estimation for descriptive qualitative studies

An example is the estimation of the prevalence of circumcision mishap among circumcisionists within a Local Government, in Ibadan.

Here, no hypothesis is tested. However, to estimate the proportion of persons with a such peculiar characteristic in the population, one needs to state the following:[5]

- Sampling error, d (how close to the proportion of interest the estimate is desired to be (e.g., 5%)
- Confidence level (95%) = Z_{α}

The minimum sample size n, is given by

$$n = \frac{(Z_{\alpha})^2 \times pq}{d^2}$$

Where p is the prevalence or proportion, q = 1 – p

Example: A health officer in a local government area wishes to estimate the prevalence of congenital anomalies in a remote community with over 10,000 people and the present prevalence rate in University College Hospital (UCH), Ibadan is estimated to be 5%. What is the minimum sample size required if he is willing to accept an absolute error of 1% at a 95% confidence level for the estimate?

It invariably means at least 1,825 subjects are needed for this study.

Sample size for a descriptive quantitative outcome

An example is estimating the serum level of albumin at presentation in UCH as a predictor of outcome in patients with perforated typhoid ileitis.

Here one must state:[5]

- The degree of precision (the amount of sampling error), d.
- The standard deviation (s) of the distribution of the characteristic in the population.
- The confidence level, which is usually 95%.
- The minimum sample size, n is given by $n = \frac{(Z_\alpha)^2 \times S^2}{d^2}$

Where: $Z_\alpha = 1.96$, at a standard normal deviate of 95% CI

A House Surgeon wishes to estimate the mean random blood glucose of patients going for surgery on a busy theater day. Preliminary information indicated that this mean is about 150 mg% with a SD of 32 mg%. If the sampling error of up to 5 mg% is to be tolerated, how many subjects should be assayed to obtain an estimate at the 95% confidence level.

Thus, the study needs at least 158 subjects.

Hypothesis testing for a single population proportion[6]

For instance, to investigate whether the current prevalence of HIV among children with tuberculosis differs from 5.2% (assumed cut-off for generalized epidemic of tuberculosis – HIV co-infection)

$$n = \frac{\{Z_{1-\alpha/2}\sqrt{p_0(1-p_0)} + Z_{1-\beta}\sqrt{p_a(1-p_a)}\}^2}{(p_a - p_0)^2}$$

Previous surveys have shown that the rate of post-circumcision urethro-cutaneous fistula among neonates in Atobatele community is about 25%. How many children should be studied in a new survey if it is desired to be 80% sure of detecting a prevalence of 20% or less at the 5% level of significance.

From the above the $P_0 = 25\%$; $p_a = 20\%$, hence the minimum sample size n is about 441.

Hypothesis testing for a single population mean.[6]

For instance, to investigate whether the current mean systolic BP in a cohort of Wilms tumor patients differs from 100 mmHg.

$$n = \frac{\{Z_{1-\alpha/2} + Z_{1-\beta}\}^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

For example, a paediatric surgeon wishes to estimate the mean penile length of male babies born in Aanuoluwapo hospital in New Zealand. How large a sample of birth records should be retrieved if he wants an estimate with a precision of 1.0cm with 99% confidence? Assume that a reasonable estimate of standard deviation for penile length is 1.0cm.

Comparison of two proportions[6]

An example is comparing the outcome of hypospadias repair in two teaching hospitals in Nigeria using fistula rate as a common denominator.

$$n = \frac{\{Z_{1-\alpha/2}\sqrt{2\bar{p}(1-\bar{p})} + Z_{1-\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)}\}^2}{(p_1-p_2)^2}$$

where

$$\bar{p} = \frac{p_1 + p_2}{2}$$

Comparison of two proportions when one sample is K times as large as the other[6]

$$n = \frac{\{Z_{1-\alpha/2}\sqrt{pq(1+1/k)} + Z_{1-\beta}\sqrt{p_1(1-p_1)+p_2(1-p_2)/k}\}^2}{(p_1-p_2)^2}$$

where

$$n_1 = kn_2$$

$$p = \frac{p_1 + kp_2}{1+k}$$

$$q = 1 - p$$

Suppose it has been estimated that the rate of malnutrition is 500 per 1000 children in public schools and 250 per 1000 in private schools. How many children should be selected from each type of school to determine whether there is a difference in the level of malnutrition at a significance level of 5% and a power of 80%?

Interventional studies

An example is randomized clinical trial (RCT), not randomized controlled trial as sometimes stated by researchers.

Here, we compare the proportion of participants with a specified outcome in a group that received treatment A and the proportion of participants with the same outcome in a group that received treatment B.[6]

There are different formulae based on whether it is an equality, equivalence, superior, or non-inferiority interventional study.

It also depends on whether it is a one-tailed, two-tailed parallel, or two-tailed crossover RCT.

The sample size in each group depends on:

- P_0 = proportion of subjects in the control group who are expected to exhibit the outcome of interest.
- P_1 = proportion of subjects in the treatment group who are expected to exhibit the outcome of interest (usually set relative to p_0).
- Level of significance = α.
- Power = 1-β.
- Anticipated attrition rate = f.

The examples below are for two-tailed parallel equality RCT, which is the most common study design for RCTs.

Sample size for RCT – with qualitative outcome.

$$n = \frac{2 * (Z_{\alpha} + Z_{1-\beta})^2 * p(1-p)}{(p_0 - p_1)^2}$$

- Where $p = (p_0+p_1)/2$

Or

$$n = \frac{(Z_{\alpha}+Z_{1-\beta})^2 * [p_0(1-p_0)+p_1(1-p_1)]}{(p_0-p_1)^2}$$

The current one-year discontinuation rate for a certain type of contraceptive method is 60%. A study is being designed to see whether the one-year discontinuation rate claimed to be 50% for a new method significantly differs from the old. How many subjects should the investigator recruit to achieve 80% power of detecting a difference at 95% confidence level? Assume no subject was lost to follow up.

Here, $P_0 = 0.6$; $p_1 = 0.5$; $Z_{1-\beta} = 0.84$; $Z_{\hat{a}} = 1.96$

About 520 subjects are needed to be recruited for each of the two methods.

Assuming an Attrition rate of 10% (i.e., f = 0.1)

- Adjusted Sample Size, N = n/(1-f) = 520/0.9 ≃ 578

Sample size for RCT – with quantitative outcome.

$$n = \frac{2 * (Z_{\alpha} + Z_{1-\beta})^2 * \sigma^2}{(\mu_0 - \mu_1)^2}$$

Where μ_0 and μ_1 are the mean value in each group and σ is the standard deviation of the measurement.

How many Wilms tumor patients with mild hypertension will be needed for a trial to detect an average difference of 10 mmHg in systolic blood pressure between a group who receives a new hypertensive drug and a control group? Assume standard deviation of SBP of 5 mmHg, 90% power and 5% level of significance.

Cohort Studies

Here, we compare the proportion of subjects with a specified outcome who are exposed to a potential health risk (or benefit) [exposed group] and the proportion of subjects with the same outcome who are not exposed to the potential health risk [unexposed group].[6]

There are two groups:

- p_0 = proportion of subjects in the unexposed group who are expected to exhibit the outcome of interest.
- p_1 = proportion of subjects in the exposed group who are expected to exhibit the outcome of interest (usually set relative to p_0).

Sample size estimation for cohort studies

$$n = \frac{2 * (Z_{\alpha} + Z_{1-\beta})^2 * p(1-p)}{(p_0 - p_1)^2}$$

- Where $p = (p_0+p_1)/2$
- Or

$$n = \frac{(Z_{\alpha}+Z_{1-\beta})^2 * [p_0(1-p_0)+p_1(1-p_1)]}{(p_0-p_1)^2}$$

Case-control studies

We compare the exposure histories of individuals who have had the outcome (cases/disease) and those who have not had the outcome (controls).[6]

The sample size formula is like that for RCT and cohort studies.

- Two quantities are also needed:
  - Proportion of cases exposed (p_1).
  - Proportion of controls exposed (p_0).

The two are related such that given an estimate of proportion of controls exposed (p_0) and the odds ratio (OR), an estimate of p_1 can be obtained from

- $p_1 = p_0 * OR / [1+p_0 * (OR-1)]$
- Where OR is for the outcome among cases versus controls.

SPECIAL CASES

1. **Unequal group sizes**
   Especially if the outcome is rare, it needs sample size adjustment.
   The ratio of controls to cases (r) may be specified and the sample size is adjusted by a factor (n) = (r+1)/(2*r).
   - Adjusted sample size, $N^* = \frac{N(1+n)^2}{4n}$

N is the sample size calculated assuming all arms are equal.

2. **More than two comparison groups**
   For three groups, one needs to estimate the sample sizes for three independent comparisons and take the maximum for the study size.
   For more than three groups, there are special methods for determining such sample sizes and a biostatistician will be of help.

3. **Non-consent, missing response, dropout, withdrawal from study**
   Here, re estimation of the sample size is done.
   - $N^* = \frac{N}{(1 - P)}$

P = Proportion; N = initial estimated sample size.

4. **More than one primary outcome**
   Sometimes, there is more than one end point, e.g., survival, response rate, quality of life, et cetera. The most important is designated as the primary outcome and can be used for sample size estimation.
   When there is more than one important outcome measure.
   - Repeat the sample size estimation for each outcome and select the largest as the sample size required.
   - There is danger with multiple significance testing as it increases type 1 error. The significance level may also need to be adjusted.

5. **Other considerations - Finite population**
   The finite population correction (fpc) must be applied when the size of population of interest is not large (usually < 10,000), this is n'.[6]

$$n' = \frac{n}{1 + \frac{n}{N}}$$

Where n = initial estimated sample size and N = size of population of interest.

It is more applicable for descriptive studies.

CONCLUSION

In estimating sample size, it is assumed that the outcome measure has an underlying normal distribution in the study population. Therefore,

- Sample size increases when the difference to detect is small, e.g., $\mu_1 - \mu_2$.
- Sample size increases with increasing power of the study, and $Z_{1-\beta}$ also has a positive correlation with the power of the study.
- When the significance level (α) is low, the $Z\alpha$ becomes high.
- When the d (error margin) is low, the sample size increases with a higher degree of precision.[4]

Sample size estimation is both statistics and pragmatics. It is possible to make a change and improve on the practice of biostatistics, however, as Cohen warns, "Don't look for a magic alternative … It does not exist".[7]

A researcher must have a statistician as a friend and some pertinent questions such as identifying the primary outcome measure for the sample size to be appropriately estimated must be answered.

When in doubt about the sample size and statistical analysis of a study, consult a biostatistician early enough. "To consult the statistician after the experiment is done may be no more than asking him to perform a postmortem examination: he may be able to say what the experiment died of".[8]

However, it is also good to know the basics of statistics even when you have a biostatistician to assist you, hence this editorial.

REFERENCES

1. **Lemeshow S** and Lwanga SK. Adequacy of Sample Size in Health Studies. 1990. WHO, Geneva.
2. **Kirkwood BR**, Sterne JAC 2003 Essential medical statistics, 2nd edition Blackwell Publishing Ltd, London.
3. **Adedokun BO.** P – value and Confidence intervals – Facts and Farces. Ann. of Ib. Postgrad. Med. 2008; 6 (1): 33.
4. **Egbuchulem KI**. How confident is the Confidence interval. Editorial. *Ann Ibd. Pg. Med* 2022; 20 (2): 101 -102.

5. **Naing L.**, Winn T., and Rusli BN. Practical issues in calculating the sample size for prevalence studies. *Archives of Orofacial Sciences, 2006. 1: 9 – 14.*
6. **Bamgboye EA** 2008. A Companion of Medical Statistics. 2nd ed. Folbam Publishers Ibadan.
7. **Cohen J.** Things I Have Learned (So Far). Amer Psychologist 1990; 45:997-1003.
8. **Fisher RA.** Ronald Fisher Quotes. Brainy Quote.com. BrainyMedia Inc, 11, May 2023. Presented at the First session of the Indian Statistical Conference, Calcutta, 1938. httpsP:// www.brainquote.com/quotes/ronald_fisher_ 186022.