International Journal of Health Sciences and Research
Vol.13; Issue: 2; February 2023
Website: www.ijhsr.org
Research Method *ISSN: 2249-9571*

Contextual Role of Absolute and Relative Precision in Estimation of Sample Size for Single Proportion in Health Research

Sachit Ganapathy[1], Deepthy MS[1], Akash Mishra[1]

[1]Department of Biostatistics, Jawaharlal Institute of Post Graduate Medical Education and Research (JIPMER), Puducherry, India

Corresponding Author: Sachit Ganapathy

DOI: https://doi.org/10.52403/ijhsr.20230210

ABSTRACT

Background: Sample size estimation is one of the key components in the initial stage of any health research. The validity of an observational study in estimating prevalence of an outcome of interest is primarily determined by the precision of the estimate. This is generally motivated by the method of sample size estimation. The choice of incorporating absolute precision or relative precision in estimating the sample size of proportions has been a grey area for many years. The objectives of the study were to investigate the role of relative and absolute precision in sample size estimation of proportion and also to provide an easy guide for estimation of sample size using relative and absolute precision using real life examples.
Materials and Methods: Sample sizes for different proportions using varying levels of relative and absolute precision were estimated and the variations in the sample size using both methods were graphically plotted.
Results: Sample size decreases exponentially with increase in anticipated prevalence in the case of relative precision whereas for absolute precision, it follows a bell-shaped curve.
Conclusions: The current study provides scenarios where and how absolute and relative precision can be used. Also, the relation between absolute and relative precision is provided.

Keywords: *sample size, prevalence, health research, relative precision, absolute precision*

INTRODUCTION

In health research, statistical inference is made about the outcome variable of interest, by estimating the population parameter with a degree of certainty. There are a few important considerations while making such an inference. One of the most important considerations is the statistical precision with which the estimation has been carried out. The statistical precision of sample statistic is the closeness with which it can be expected to approximate the relevant population value.[1] When we consider the sampling distribution of a sample statistic, the standard error of the statistic determines the statistical precision. At a fixed level of significance and a constant variability in the data, the only way of increasing the statistical precision is by inflating the sample size.

In both observational and interventional studies, based on the research question, study design and analysis plan, the required minimum sample size is estimated. For descriptive studies with an intention of estimating the population parameter, the major components which determines the sample size are level of significance, standard deviation and margin of error. For analytical studies, with a test of hypothesis as a part of the study, power of the test is an additional component that is required to estimate the sample size. In this study, our

main aim is to consider the estimation of sample size for estimating the prevalence/proportion of a population parameter.

MATERIALS & METHODS

Requirements and formulas for sample size estimation of single proportion

The components required to calculate this sample size in this scenario are, level of significance (α), anticipated prevalence/ proportion (p) and precision (d). For a given level of significance and proportion, precision can be of two types, absolute precision and relative precision.

Absolute precision is the total percentage units that is permitted on either side of the estimate.[2] It gives us the margin of error in absolute units of the estimated proportion. For example, estimating the prevalence of gestational Diabetes mellitus amongst pregnant women in South India, the estimate was found to be 16.2%[3] with 10% absolute precision. This means that the confidence interval of the point estimate can vary between 6.2% - 26.2% for the predetermined '∝' level of significance. The determination of sample size for estimating proportion with absolute precision is given by:

$$n \geq Z^2_{1-\alpha/2} \frac{P(1-P)}{d^2}$$

where,

Z is the standard normal variate for a two tailed test at α level of significance,

P is the anticipated proportion and

d is the absolute precision

For example, the minimum sample size required to estimate the prevalence of gestational Diabetes mellitus amongst pregnant woman in South India whose anticipated prevalence is 16.2% at 5% level of significance with an absolute precision of 10% comes out to be 53 subjects.

Relative precision is the percentage units of the true value that is permitted on either side of the estimate. It gives us a margin of error in relative units to the estimated proportion of the outcome variable. For example, estimating the prevalence of obesity amongst adolescents in India, the estimate was found to be 2.2%[4] with 25% precision relative to the estimate. This means that the confidence interval of the point estimate can vary between 1.1% to 3.3% for the predetermined '∝' level of significance. The determination of sample size for estimating proportion with relative precision is given by:

$$n \geq Z^2_{1-\alpha/2} \frac{(1-P)}{(Pd)^2}$$

where,

Z is the standard normal variate for a two tailed test at α level of significance,

P is the anticipated proportion and

Pd is the relative precision.

For example, the minimum sample size required to estimate the prevalence of cancer in South India whose anticipated prevalence is 2.2% at 5% level of significance with a relative precision of 25% of anticipated prevalence comes out to be 2,733 subjects.

Relationship between sample sizes by relative and absolute precision

The relationship between the sample sizes calculated using relative and absolute precision can be established by comparing the formulae. The relationship is given by:

$$n_r = P^2 n_a$$

Here "n_r" is the sample size estimated by using relative precision, "n_a" is the sample size estimated by using absolute precision and "*P"* is the anticipated prevalence of the condition of interest. As "*P"* tends towards 1, the sample size nearly coincides for both absolute and relative precision.

Finite population correction

Finite population correction (FPC) is a factor considered in the estimation of required sample size for the study when the sampling fraction from a population is high. FPC is given by the formula,

$$FPC = \sqrt{\frac{N-n}{N-1}}$$

Sampling techniques are carried out with an assumption that our population of interest is very large (infinite). But when the population of interest is small, and sampling fraction of sample size and population size is large, the

standard error of the sample mean is overestimated as central limit theorem does not hold in these conditions[7]. When the sampling fraction is small, FPC ~1 and therefore FPC can be ignored. When the sampling fraction goes beyond 5-10%, it is important to consider FPC in estimation of sample size.

Adjusting for design effect

In a cross-sectional study, the variance of the sample proportion is estimated with an assumption that the samples were chosen with simple random sampling (SRS) method. In a situation where cluster sampling method is used instead, the strength of association between the clusters, also called as intra-class correlation (ICC) must be considered for estimation of sample size. This effect of sampling method used for estimation of sample size is called as design effect[8]. Design effect is given by,

$$Design\ effect = 1 + \rho(n-1)$$

where ρ is the ICC.

The variance of sample proportion in cluster sampling is given by,

$$V_{cluster}(\hat{p}) = \frac{pq}{k * m}[1 + \rho(m-1)]$$

RESULT

Sample size for absolute and relative precision

The estimation of sample size gives different values for different anticipated prevalence and margin of error. It is illustrated with the worked-out estimation of sample size for different values of proportion and margin of error as given in Table 1.

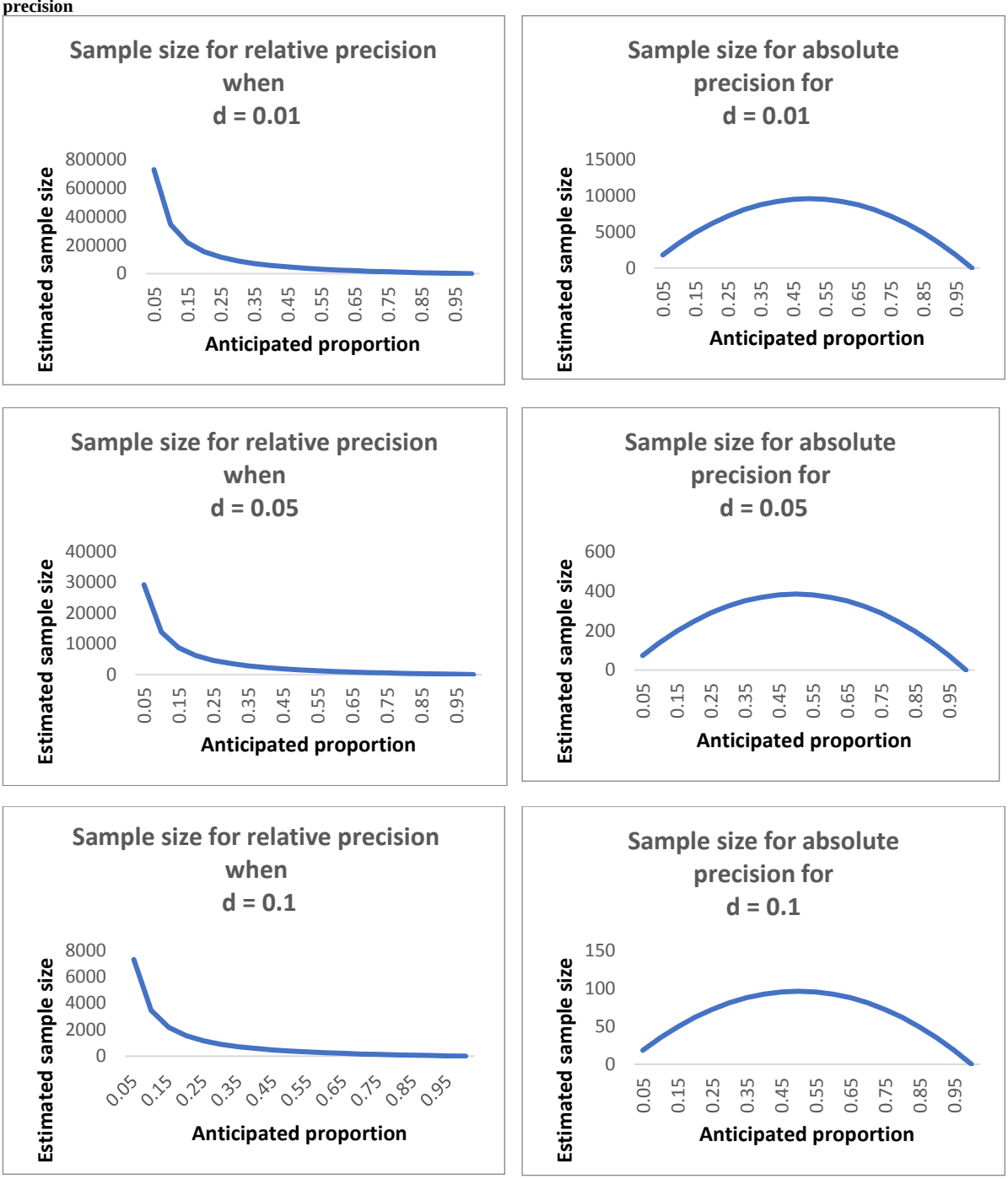
We can observe that the sample size varies significantly between that estimated from absolute precision and relative precision for a smaller anticipated prevalence. But as this value of "*P*" keeps on increasing, the two sample sizes tends to get closer to each other. Therefore, our rules to apply the type of precision for different situation are consistent with this observation. To visualize how the sample size varies with the increase in the anticipated prevalence for relative and precision refer Figure 1.

We observe that the sample size decreases exponentially with increase in anticipated prevalence in the case of relative precision. The sample size calculated from absolute precision, on the contrary follows a bell-shaped curve.

Table 1: The estimated sample size tables for varying levels of relative and absolute precision in a study for estimation of single proportion

Proportion	d = 0.01		d = 0.05		d = 0.10	
	Relative precision	Absolute precision	Relative precision	Absolute precision	Relative precision	Absolute precision
0.05	729904	1825	29196	73	7299	18
0.1	345744	3457	13830	139	3457	35
0.15	217691	4898	8708	196	2177	49
0.2	153664	6147	6147	246	1537	61
0.25	115248	7203	4610	289	1152	72
0.3	89637	8067	3586	323	896	81
0.35	71344	8740	2854	350	713	87
0.4	57624	9220	2305	369	576	92
0.45	46953	9508	1879	381	470	95
0.5	38416	9604	1537	385	384	96
0.55	31431	9508	1258	381	314	95
0.6	25611	9220	1025	369	256	92
0.65	20686	8740	828	350	207	87
0.7	16464	8067	659	323	165	81
0.75	12805	7203	513	289	128	72
0.8	9604	6147	385	246	96	61
0.85	6779	4898	272	196	68	49
0.9	4268	3457	171	139	43	35
0.95	2022	1825	81	73	20	18
1	0	0	0	0	0	0

Figure 1: Graphs depicting the change in sample size for varying levels of anticipated prevalence with relative and absolute precision



DISCUSSION

An exponential decrease in sample size was observed with an increase in anticipated prevalence for relative precision. The reason for this is because sample size in relative precision is proportional exponentially on the prevalence of the condition of interest. As the anticipated prevalence increases, the sample size from relative precision nearly coincides with that of absolute precision. Also, for an increase in absolute precision, sample size followed a bell-shaped inverted curve. The reason is due to the inherent structure of the sampling distribution of the sample statistic. We observe that the sample size depends heavily on the variance of the statistic, which is maximum when the anticipated prevalence is 50%.

Although relative precision accounts for a smaller margin of error, it inflates the sample

size which may not be feasible at all times. There are determinantal rule for the use of one of them in estimating the sample size but, the following three simple rules can be followed to make the choice between them:

When the disease condition is very rare, anticipated prevalence of the condition of interest is small and therefore opting relative precision is a better option. When the anticipated prevalence is small (say <10%), we need a large sample to get enough people with the condition of interest in the study. Considering relative precision gives us a larger sample size with a smaller margin of error, it is a better choice. Also, for a small prevalence, the lower limit of the confidence interval may go to an irrelevant negative lower-bound value which is logically flawed for practical application.[5]

When the anticipated prevalence of the condition of interest is moderate, opting absolute precision is a better option. For a moderate prevalence (say between 10% to 90%), the sample size estimated from absolute precision will be sufficient to get the required estimate. In terms of the feasibility of sample size also absolute precision will give a better prospect.

When the disease condition is very common (say >90%), anticipated prevalence of the condition of interest is high and therefore opting absolute or relative precision does not make much difference. But researchers must opt for a smaller margin of error as sample size will not inflate much with higher anticipated prevalence. Also, choosing a smaller margin of error will make sure the upper limit of confidence interval will not go beyond 1 which is logically flawed for practical applications.

CONCLUSION

This easy guide will help the budding researchers and clinicians for the estimation of sample size for proportions.

Declaration by Authors

Ethical Approval: Not Applicable

Acknowledgement: None

Source of Funding: None

Conflict of Interest: The authors declare no conflict of interest.

REFERENCES

1. Cohen J. Statistical power analysis for the behavioral sciences. Routledge; 2013 May 13. http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf
2. Chadha VK. Sample size determination in health studies. NTI Bulletin - 2006, 42/3&4, 55 – 62. https://ntiindia.kar.nic.in/ntibulletin/ntibulletin%20volume%2042-3&4/pages/pdf/03_mainpaper_sampling.pdf
3. Seshiah V, Balaji V, Balaji MS, Paneerselvam A, Arthi T, Thamizharasi M, Datta M. Prevalence of gestational diabetes mellitus in South India (Tamil Nadu): A community-based study. JAPI. 2008 May; 56:329-3. https://www.japi.org/t2f48494/prevalence-of-gestational-diabetes-mellitus-in-south-india-tamil-nadu-a-community-based-study
4. Goyal RK, Shah VN, Saboo BD, Phatak SR, Shah NN, Gohel MC, Raval PB, Patel SS. Prevalence of overweight and obesity in Indian adolescent school going children: its relationship with socioeconomic status and associated lifestyle factors. The Journal of the Association of Physicians of India. 2010 Mar 1;58:151-8. https://www.japi.org/s2c47454/prevalence-of-overweight-and-obesity-in-indian-adolescent-school-going-children-its-relationship-with-socioeconomic-status-and-associated-lifestyle-factors
5. Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. Archives of orofacial Sciences. 2006; 1:9-14. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.504.2129&rep=rep1&type=pdf
6. Arya R, Antonisamy B, Kumar S. Sample size estimation in prevalence studies. The Indian Journal of Pediatrics. 2012 Nov 1;79(11):1482-8. https://link.springer.com/article/10.1007/s12098-012-0763-3
7. Cochran WG. Sampling techniques. John Wiley & Sons; 2007 Nov 21. https://hwbdocuments.env.nm.gov/Los%20

Alamos%20National%20Labs/General/14447.pdf

8. Alimohamadi Y, Sepandi M. Considering the design effect in cluster sampling. Journal of Cardiovascular and Thoracic Research. 2019;11(1):78. https://www.ncbi.nlm.nih.gov/labs/pmc/articles/PMC6477104/

How to cite this article: Sachit Ganapathy, Deepthy MS, Akash Mishra. Contextual role of absolute and relative precision in estimation of sample size for single proportion in health research. *Int J Health Sci Res.* 2023; 13(2):63-68. DOI: *https://doi.org/10.52403/ijhsr.20230210*

******