EDITORIAL

How to determine Sample Size: the Design of Sample Size in Health Studies

Eihab Abdel Rahim Eldawi Ahmed [1], Dr. Mohamed Salih Mahfouz [2] and Isam Omer Fdul [1]

1. Faculty of Economics –Department of Statistics –University of Gezira– MSc in Statistics – Email:Eihabmer@gmail.com
2. Department of Family and Community Medicine – Faculty of Medicine – Jazan University – PhD in Statistics – Email: mm.mahfouz@gmail.com.
3. Faculty of Economics –Department of Statistics – University of Gezira – MSc in Population and Development

Abstract:

Sample-size determination is often an important step in planning a health study—and it is usually a difficult one. Among the important hurdles to be surpassed, one must obtain an estimate of one or more error variances, and specify an effect size of importance. This paper offers some suggestions for successful and meaningful sample-size determination. Also discussed is the possibility that a sample size may not be the main issue and that the real goal is to design a high-quality study. Finally, criticism is made of some ill-advised shortcuts relating to testing power and sample size.

***Key words*:** Sample size, proportion, mean, confidence interval, precision, null hypothesis, power, Type I and Type II errors.

Introduction

In experimental researche in health studies , the choice of an appropriate sample size is the most important factor to consider in the design of experimental researches. Studies that are too small may fail to detect important effects on the outcomes of interest, or may estimate those effects too imprecisely. Studies that are too large are a waste of resources. For such an important issue, there is a great amount of published literature. Important general references include Mace (1964), Kraemer and Thiemann (1987), Cohen (1988), Desu and Raghavarao (1990),Lipsey (1990), Shuster (1990), and Odeh and Fox (1991). There are

.

EDITORIAL

numerous articles, especially in biostatistics journals, concerning sample-size determination for specific tests. Also of interest are studies of the extent to which sample size is adequate or inadequate in published studies; see Freiman et al. (1986) and Thornley and Adams (1998). In Al-Gezira University Alnoury (2002) produced a quick manual for calculating sample size for reproductive health studies.

There is a growing amount of software for sample-size determination, including nQuery Advisor (Elashoff, 2000), PASS (Hintze, 2000), UnifyPow (O'Brien, 1998), and Power and Precision (Borenstein et al., 1997). Web resources include a comprehensive list of power-analysis software (Thomas, 1998) and online calculators such as Lenth (2000). Wheeler (1974) provides some useful approximations for use in linear models; Castelloe (2000) gives an up-to-date overview of computational methods. May be the most frequently asked question concerning sampling is, "What sample size do I need?" The answer to this question is influenced by a number of factors, including the purpose of the study, population size, the risk of selecting a "bad" sample, and the allowable sampling error.

The objective of this paper is to illustrate some statistical concepts that are necessary for sample size determination, and to review the most common methods of sample size determination in health studies. Also criticism is made of some ill-advised shortcuts relating to testing power and sample size. Before calculating sample size it is necessary to go through the following concepts: types of outcome measure, p.value etc. effect size, population etc... Traditionally, data collected in a research study is submitted to a significance test to assess the viability of the null hypothesis.

Types of outcome Measures:

The statistical methods for sample size determination depend on which type of outcome is expected. The three most common types of outcomes in case of surveys/ studies / trials are:

i. Proportions: For example, in a trial of a new measles vaccine, an outcome measure of interest may be the proportion of vaccinated subjects who develop high levels of antibodies.

ii. Means: For example, in a trial of an antimalarial intervention, it may be of interest to compare the mean packed cell volume (PCV) at the end of the malaria season among those in the intervention group and those in the comparison group.

iii. Rates: For example, in a trail of multi-drug therapy for leprosy, the incidence rates of relapse following treatment may be compared in the different study groups under consideration.

P-value

The p-value, provided by the significance test and used to reject the null hypothesis, is a function of three factors: size of the observed effect, sample size, and the criterion required for significance (alpha). A power analysis, executed when the study is being planned, is used to anticipate the likelihood that the study will yield a significant effect and is based on the same factors as the significance test itself.

The P value or calculated probability is the estimated probability of rejecting the null hypothesis (H_0) of a study question when that hypothesis is true. The null hypothesis is usually an hypothesis of "no difference" e.g. no difference between blood pressures in group A and group B. Define a null hypothesis for each study question clearly before the start of your study. The only situation in which you should use a one sided P

EDITORIAL

value is when a large change in an unexpected direction would have absolutely no relevance to your study. This situation is unusual; if you are in any doubt then use a two sided P value. The term significance level (alpha) is used to refer to a pre-chosen probability and the term "P value" is used to indicate a probability of posterior alpha.

The alternative hypothesis (H_1) is the opposite of the null hypothesis; in plain language terms this is usually the hypothesis you set out to investigate. For example, question is "is there a significant (not due to chance) difference in blood pressures between groups A and B if we give group A the test drug and group B a sugar pill?" and alternative hypothesis is " there is a difference in blood pressures between groups A and B if we give group A the test drug and group B a sugar pill". If your P-value is less than the chosen significance level then you reject the null hypothesis i.e. accept that your sample gives reasonable evidence to support the alternative hypothesis. It does NOT imply a "meaningful" or "important" difference; that is for you to decide when considering the real-world relevance of your result.

The choice of significance level at which you reject H_0 is arbitrary. Conventionally the 5% (less than 1 in 20 chance of being the 5% (less than 1 in 20 chance of being wrong), 1% and 0.1% ($P < 0.05$, 0.01 and 0.001) levels have been used. These numbers can give a false sense of security. In the ideal world, we would be able to define a "perfectly" random sample, the most appropriate test and one definitive conclusion. We simply cannot. What we can do is try to optimize all stages of our research to minimize sources of uncertainty. When presenting P values some groups find it helpful to use the asterisk rating system as well as quoting the P value:

$P < 0.05$ * $P < 0.01$ ** $P < 0.001$

Most authors refer to statistically significant as $P < 0.05$ and statistically highly significant as $P < 0.001$ (less than one in a thousand chance of being wrong). The asterisk system avoids the woolly term "significant". Please note, however, that many statisticians do not like the asterisk rating system when it is used without showing P values. As a rule of thumb, if you can quote an exact P value then do. You might also want to refer to a quoted exact P-value as an asterisk in text narrative or tables of contrasts elsewhere in a report. At this point, a word about error. Type I error is the false rejection of the null hypothesis and type II error is the false acceptance of the null hypothesis.

The significance level (alpha) is the probability of type I error. The power of a test is one minus the probability of type II error (beta). Power should be maximized when selecting statistical methods. If you want to estimate sample sizes then you must understand all of the terms mentioned here.

The following table shows the relationship between power and error in hypothesis testing:

	Decision	
Truth	Accept H_0	Reject H_0
H_0 is true	Correct Decision P	Type I Error P

EDITORIAL

	1-alpha	Alpha (significance)
H_0 is false	Type II error P	correct decision P
	Beta	1-beta (power)

H_0 = Null hypothesis, P = Probability

If you are interested in further details of probability and sampling theory at this point then please refer to one of the general texts listed in the reference section.You must understand confidence intervals if you intend to quote P values in reports and papers. Statistical referees of scientific journals expect authors to quote confidence intervals with greater prominence than P values. Computed for a significance test, the p-value gives the proportion of cases (for a population in which the effect is null) that will yield a sample in which the effect is as large (or larger) than the observed effect. also the proportion of studies expected to result in a type 1 error.

Effect Size and Power Analysis

In clinical trails the most important principle is the homogenous of the selected members of the sample in terms of: age, prognosis, history of diseases and sex.The term *effect size* refers to the magnitude of the effect under the alternate hypothesis. The nature of the effect size will vary from one statistical procedure to the next (it could be the difference in cure rates, or a standardized mean difference, or a correlation co- efficient), but its function in power analysis is the same in all procedures.

The ***Effect Size*** should represent the smallest effect that would be of clinical or substantive significance, and for this reason, it will vary from one study to the next. In clinical trials, for example, the selection of an effect size might take into account the severity of the illness being treated (a treatment effect that reduces mortality by 1% might be clinically important, while a treatment effect that reduces transient asthma by 20% may be of little interest). It might take into account the existence of alternate treatments. (If alternate treatments exist, a new treatment would need to surpass these other treatments to be important). It might also take into account the treatment's cost and side effects. (A treatment that carried these burdens would be adopted only if the treatment effect was very substantial).

Power analysis gives power for a specific effect size. For example, the researcher might report that if the treatment increases the recovery rate by 15 percentage points, the study will have power of 80% to yield a significant effect. For the same sample size and alpha, if the treatment effect is less than15 percentage points, then the power will be less than 80%. If the true effect size exceeds 15 percentage points, then power will exceed 80%. While one might be tempted to set the "clinically significant effect" at a small value to ensure high power for even a small effect, this determination can't be made in isolation. The selection of an effect size reflects the need for balance between the size of the effect that we can detect and resources available for the study.

EDITORIAL

The ***true*** (population) effect size is not known. While the effect size used for the power analysis is assumed to reflect the population effect size, the power analysis is more appropriately expressed as, "If the true effect is this large, power would be …," rather than, "The true effect is this large, and therefore power is ….". This distinction is an important one.

Researchers sometimes assume that a power analysis cannot be performed in the absence of pilot data. In fact, it is usually possible to perform a power analysis based entirely on a logical assessment of what constitutes a clinically (or theoretically) important effect (formative research is essential). Indeed, while the effect observed in prior studies might help to provide an estimate of the true effect, it is not likely to be the true effect in the population if we knew that the effect size in these studies was accurate, there would be no need to run the new study.

Since the effect size used in power analysis is not the true population value, the researcher may decide to present a range of power estimates. For example (assuming that N = 93 per group and alpha=0.05,two-tailed),the researcher may state that the study will have power of 80% to detect a treatment effect of 20 points (30% versus 50%) and power of 99% to detect a treatment effect of 30 points (30% versus 60%).

Specifically, the larger the effect size used in the power analysis, the larger the sample size; the larger (more liberal) the criterion required for significance (alpha), the higher the expectation that the study will yield a statistically significant effect. These three factors, together with power, form a closed system — once any three are established, the fourth is completely determined. The goal of a power analysis is to find an appropriate balance among these factors by taking into account the substantive goals of the study, and the resources available to the researcher.

Sample Size Criteria

In addition to the purpose of the study and population size, three criteria usually will need to be specified to determine the appropriate sample size: the type and level of precision, the level of confidence or risk, and the degree of variability in the attributes being measured (Miaoulis and Michener, 1976). Each of these is reviewed below.

The Level of Precision

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

In his book Sampling Techniques, 3rd ed. (pp 72-74), William Cochran gives the example of an anthropologist who wishes to know the percentage of inhabitants of some island who belong to blood group O. He decides he needs to know this to within 5%. Why 5%? Why not 4% or 6%. we don't know. Neither does Cochran. Cochran asks! He strongly suspects that the islanders belong either to a racial type with a P

EDITORIAL

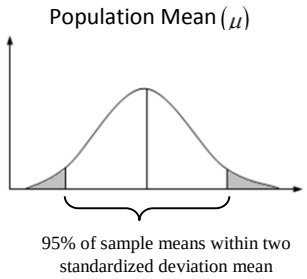
of about 35% or to one with a P of about 50%. An error limit of 5% in the estimate seemed to him small enough to permit classification into one of these types. He would, however, have no violent objection to 4 or 6% limits of error. Thus the choice of a 5 %limit of error by the anthropologist was to some extent arbitrary. In this respect the example is typical of the way in which a limit of error is often decided on. In fact, the anthropologist was more certain of what he wanted than many other scientists and administrators will be found to be. When the question of desired degree of precision is first raised, such persons may confess that they have never thought about it and have no idea of the answer.

The Confidence Level

The confidence or risk level is based on ideas encompassed under the Central Limit Theorem. The key idea encompassed in the Central Limit Theorem is that when a population is repeatedly sampled, the average value of the attribute obtained by those samples is equal to the true population value. Furthermore, the values obtained by these samples are distributed normally about the true value, with some samples having a higher value and some obtaining a lower score than the true population value. In a normal distribution, approximately 95% of the sample values are within two standard deviations of the true population value (e.g., mean).

In other words, this means that, if a 95% confidence level is selected, 95 out of 100 samples will have the true population value within the range of precision specified earlier (Figure1). There is always a chance that the sample you obtain does not represent the true population value. Such samples with extreme values are represented by the shaded areas in Figure 1. This risk is reduced for 99% confidence levels and increased for 90% (or lower) confidence levels.

Figure 1.



Degree of Variability

The third criterion, the degree of variability which in the attributes being measured refers to the distribution of attributes in the population. The more heterogeneous a population frame, the larger the sample size required to obtain a given level of precision. The less variable (more homogeneous) a population frame, the smaller the sample size. Note that a proportion of 50% indicates a greater level of variability than either 20% or 80%. This is because 20% and 80% indicate that a large majority do not or do, respectively, have the attribute of interest. Because a proportion of 0.5 indicates the maximum variability in a population, it is often used in determining a more conservative sample size, that is, the sample size may be larger than if the true variability of the population attribute were used.

EDITORIAL

Population Frame

How many people are there in the group your sample represents? This may be the number of people in a city you are studying, the number of people who buy new cars, etc. Often you may not know the exact population size. This is not a problem. The mathematics of probability proves the size of the population is irrelevant, unless the size of the sample exceeds a few percent of the total population you are examining. This means that a sample of 500 people is equally useful in examining the opinions of a state of 15,000,000 as it would a city of 100,000. For this reason, The Survey System ignores the population size when it is "large" or unknown. Population size is only likely to be a factor when you work with a relatively small and known group of people (e.g., the members of an association), and this depends on the choice of the anticipated population value.

Having examined the most important concepts that necessary for sample size calculations, interest now turns to the different strategies for sample size determination. The coming section discusses sample size determination.

Strategies for Determining Sample Size

There are several approaches to determining the sample size. These include using a census for small populations, imitating a sample size of similar studies, using published tables, and applying formulas to calculate a sample size. Each strategy is discussed below.

Using a Census for Small Populations

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

Using a Sample Size of a Similar Study

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

Please note two things. First, these sample sizes reflect the number of obtained responses, and not necessarily the number of surveys mailed or interviews planned (this number is often increased to compensate for no response). Second, the sample sizes in Table 2 presume that the attributes being measured are distributed normally or nearly so. If this assumption cannot be met, then the entire population may need to be surveyed.

Using Formulas to Calculate a Sample Size

EDITORIAL

Although tables can provide a useful guide for determining the sample size, you may need to calculate the necessary sample size for a different combination of levels of precision, confidence, and variability. The fourth approach to determining sample size is the application of one of several formulas, here the famous of them.

Formula for Calculating a Sample for Proportions

For populations that are large, Cochran (1963:75) developed the Equation 1 to yield a representative sample for proportions.

$$Eq.(1): \qquad n_0 = \frac{z^2 pq}{d^2}$$

Which is valid where n_0 is the sample size, Z is the abscissa of the normal curve that cuts off an area at the tails (1-equals the desired confidence level, e.g., 95%), d is the desired level of precision, p is the estimated proportion of an attribute that is present in the population, and q is 1-p. The value for Z is found in statistical tables which contain the area under the normal curve.

To illustrate, suppose we wish to evaluate a state-wide Extension program in which patients were encouraged to adopt a new practice.

Assume there is a large population but that we do not know the variability in the proportion that will adopt the practice; therefore, assume p=0.5 (maximum variability). Furthermore, suppose we desire a 95% confidence level and ±5% precision. The resulting sample size is demonstrated as following:

$$n_0 = \frac{z^2 pq}{d^2} = \frac{(1.96)^2 (0.5)(0.5)}{(0.05)^2} = 385\ patiens$$

Finite Population Correction for Proportions

If the population is small then the sample size can be reduced slightly. This is because a given sample size provides proportionately more information for a small population than for a large population. The sample size (n_0) can be adjusted using Equation 2 .

$$Eq.(2): \qquad n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

Where n is the sample size and N is the population size.

Suppose our evaluation of patient's adoption of the new practice only affected 4,000 patients. The sample size that would now be necessary is shown below

EDITORIAL

$$n = \frac{n_0}{1+\frac{(n_0-1)}{N}} = \frac{385}{1+(\frac{385-1)}{4000}}$$
$$= 351\ patients$$

As you can see, this adjustment (called the finite population correction) can substantially reduce the necessary sample size for small populations.

A Simplified Formula for Proportions

Yamane (1967:886) provides a simplified formula to calculate sample sizes. This formula is shown below. A 95% confidence level and P =0.5 are assumed for Equation 3.

$$Eq.(3): \qquad n = \frac{N}{1+Nd^2}$$

Where n is the sample size, N is the population size, and d is the level of precision. When this formula is applied to the above sample, we get

$$n = \frac{N}{1+Nd^2} = \frac{4000}{1+(4000)(0.05)^2}$$
$$= 364\ pat\,\text{int}\,s$$

Sample Size Requirements for Estimating a Mean or Mean Difference

The use of tables and formulas to determine sample size in the above discussion employed proportions that assume a dichotomous response for the attributes being measured. There are two methods to determine sample size for variables that are polytomous or continuous. One method is to combine responses into two categories and then use a sample size based on proportion (Smith, 1983). The second method is to use the formula for the sample size for the mean. The formula of the sample size for the mean is similar to that of the proportion, except for the measure of variability. The formula for the mean employs σ (the standard deviation) instead of (p×q), as shown in Equation 4.

$$Eq.(4): \qquad n = \frac{Z^2\sigma^2}{d^2} = \left(\frac{Z\sigma}{d}\right)^2$$

Where n is the sample size, z is the abscissa of the normal curve that cuts off an area at the tails, d is the desired level of precision (in the same unit of measure as the standard deviation), and σ is the standard deviation of an attribute in the population.

You should put considerable effort into getting a good estimate of the standard deviation of the variable you are studying since sample size calculations depend on this fact. Such estimates come from prior studies, pilot studies, and "Gestalt" (a combination of sources that contribute to knowledge about the variable).

EDITORIAL

The disadvantage of the sample size based on the mean is that a "good" estimate of the population standard deviation is necessary. Often, an estimate is not available. Furthermore, the sample size can vary widely from one attribute to another because each is likely to have a different variance. Because of these problems, the sample size for the proportion is frequently preferred.

Power

Power $(1-\beta)$is the probability of avoiding a type II error. A type II error occurs when you retain a false H0. Conventional practice is to determine the sample size that gives 80% power at the 0.05 level of significance (two- sided). This can be determined with:

$$Eq.(5) \quad n = \frac{16\sigma^2}{\Delta^2} + 1$$

(Dallal,1997,www.tufts.edu/~gdallal/SIZE.HTM) where

n: represents the required sample size per group

Δ: represents the expected mean difference (a difference worth detecting). For a one-sample t test $\Delta = \overline{X} - \mu$. For a paired-sample t test $\Delta = \mu_d$ .For an independent t test $\Delta = \mu_1 - \mu_2$. These values are NOT calculated but come from conjecture.

σ :represents the standard deviation of the variable as estimated by Sd.

For example, if you are trying to detect a mean difference of 5 for a variable with a standard deviation of 10, the required sample size per group

$$n = \frac{16\sigma^2}{\Delta^2} + 1 = \frac{16(10)^2}{(5)^2} + 1 = 65$$

When testing multiple treatments with ANOVA we suggest you focus on the most important post hoc comparison of means and use the above formula, perhaps using the square root of the *Mean Square Within* as your standard deviation.

Using Published Tables

The fourth way to determine sample size is to rely on published tables which provide the sample size for a given set of criteria. Table 1 and Table 2 present sample sizes that would be necessary for given combinations of precision, confidence levels, and variability. Both tables are based on the previously discussed equations (1 and 2).

EDITORIAL

Table 1. Sample size for ±3%, ±5%, ±7% and ±10% Precision Levels Where Confidence Level is 95% and P=0.5.

Size of	Sample Size (n) for Precision (e) of:			
Population	±3%	±5%	±7%	±10%
500	340	217	141	81
600	384	234	148	83
700	423	248	153	84
800	457	260	157	86
900	488	269	161	87
1,000	516	278	164	88
1500	624	306	173	90
2,000	696	322	179	92
2,500	748	333	182	92
3,000	787	341	184	93
4,000	842	350	187	94
5,000	879	357	189	94
6,000	906	361	190	95
7,000	926	364	191	95
8,000	942	367	191	95
9,000	954	368	192	95

EDITORIAL

10,000	1000	385	200	99
15,000	1034	390	201	99
20,000	1053	392	202	100
25,000	1064	394	202	100
50,000	1087	397	203	100
100,000	1099	398	204	100

Power of a t Test (Comparison of Means)

Let $\phi(z)$ represent the area under the curve to the left of z on a standard normal curve. For example, $\phi(0) = 0.50 = 50\%, \phi(0.75) = 0.7734 = \%77.34$

and $\phi(1.96) = 0.9750 = \%97.5$ (If this notation confuses you, draw a standard normal curve and place the z value on the x axis. The area under the curve to the left of the Z point represents the probability you want to know.), see figure 2.

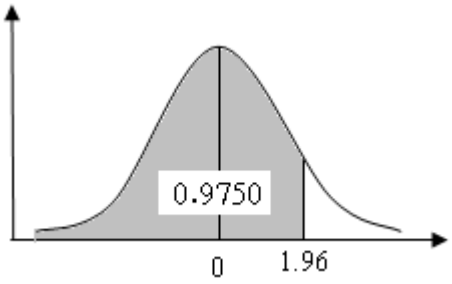


Figure 2

The power of a test $1-\beta$ comparing means is approximately equal to:

$$Eq(7)\ Power = 1-\beta = \phi\left(-1.96\frac{|d|\sqrt{n}}{\sigma}\right)$$

Where: d denotes the expected mean difference (or difference worth detecting).

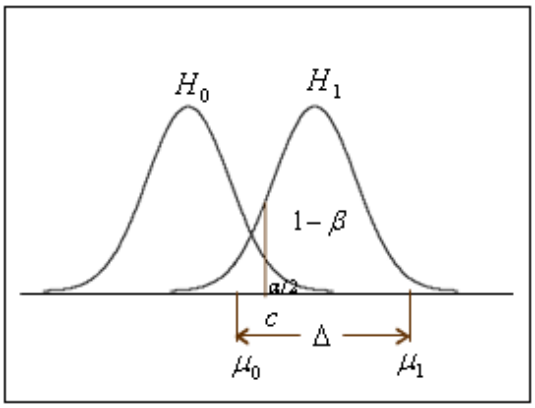
n denotes the per group sample size

σ denotes the standard deviation of the variable.

EDITORIAL

Equation (7) assumes a sampling distribution of no difference (H_0) and an alternative sampling distribution of a difference (H_1). Let critical value c determines the point at which you will reject H0. The power of the test is the area under the alternative curve beyond c (see Fig.3, below).

Figur.2



For example, a study of 25 pairs expects a mean difference of 1.5. The standard deviation of the paired difference is 4. Thus, the power of the test is

$$power = \phi\left(-1.96 + \frac{|1.5|\sqrt{25}}{2}\right)$$
$$= \phi(1.79) = \%96$$

Other Considerations

In completing this discussion of determining sample size, there are three additional issues. First, the above approaches to determining sample size have assumed that a simple random sample is the sampling design. More complex designs, e.g., stratified random samples, must take into account the variances of subpopulations, strata, or clusters before an estimate of the variability in the population as a whole can be made. Usually simple random samples are adjusted easily, by using a design effect multiplies by 2.

Another consideration with sample size is the number needed for the data analysis. If descriptive statistics are to be used, e.g., mean, frequencies, then nearly any sample size will suffice. On the other hand, a good size sample, e.g., 200-500, is needed for multiple regression, analysis of covariance, or log-linear analysis, which might be performed for more rigorous state impact evaluations. The sample size should be appropriate for the analysis that is planned.

In addition, an adjustment in the sample size may be needed to accommodate a comparative analysis of subgroups (e.g., such as an evaluation of program participants with nonparticipants). Sudman (1976) suggests that a minimum of 100 elements is needed for each major group or subgroup in the sample and for each minor subgroup, a sample of 20 to 50 elements is necessary. Similarly, Kish (1965) says that 30 to 200 elements are sufficient when the attribute is present 20 to 80 percent of the time (i.e., the distribution

approaches normality). On the other hand, skewed distributions can result in serious departures from normality even for moderate size samples (Kish, 1965:17). Then a larger sample or a census is required.

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 persons that the researcher is unable to contact. The sample size also is often increased by 30% to compensate for nonresponse. Thus, the number of mailed surveys or planned interviews can be substantially larger than the number required for a desired level of confidence and precision.

Discussions and Conclusion

Sample-size planning is often important, and almost always difficult. It requires care in eliciting scientific objectives and in obtaining suitable quantitative information prior to the study. Successful resolution of the sample-size problem requires the close and honest collaboration of statisticians and subject-matter experts.

It is a practical reality that sample size is not always determined based on noble scientific goals. Then it is important to evaluate the proposed study to see if it will meet scientific standards. Various types of changes to the study can be recommended if it turns out to be over- or under-powered. Sample-size problems are context-dependent. For example, how important it is to increase the sample size to account for such uncertainty depends on practical and ethical criteria. Moreover, sample size is not always the main issue; it is only one aspect of the quality of a study design. Besides the approaches discussed here, there are other respectable approaches to sample-size planning, including Bayesian ones and frequentist methods that focus on estimation rather than testing. While technically different, those approaches also require care in considering scientific goals, incorporating pilot data, ethics, and study design.

Over the last decade, statistical program packages have been developed for analyzing data from complex sample surveys. The best known of these is SUDAAN (from SUrvey DAta ANalysis), which is available as a stand-alone program or as an add-on to SAS. Lately, SAS has been adding this functionality to its own program with its SURVEYMEANS and SURVEYREG procedures.

References:

1-Alnoury A. H (2002) . Manual for Sample Size Determination in Reproductive Health Research, Population Studies Centre, Silver Star Press.

2-Bailar III and F. Mosteller , chap. 14, pp. 289–304, NEJM Books, Waltham, Mass.

3-Boen, J. R. and Zahn, D. A. (1982), The Human Side of Statistical Consulting, Lifetime Learning Publications, Belmont, CA.

4-Borenstein, M., Rothstein, H., and Cohen, J. (1997), Power and Precision, Biostat, Teaneck, NJ, Softwarefor MS-DOS systems.

EDITORIAL

5-Castelloe, J. (2000), "Sample Size Computations and Power Analysis with the SAS System," in Proceedingsof the Twenty-Fifth Annual SAS User's Group International Conference, Cary, NC, SAS Institute, Inc., Paper 265-25.

6-Cochran, W.G. (1977) Sampling Techniques,. Wiley Pub., 3rd edition, New York.

6- Cohen, J. (1988), Statistical Power Analysis for the Behavioral Sciences, Academic Press, New York, 2nd edn.

7-Dallal,Gerard, E.1997, ww.tufts.edu/~gdallal/SIZE.HTM

8- Desu, M. M. and Raghavarao, D. (1990), Sample Size Methodology, Academic Press, Boston.

9-Elashoff, J. (2000), nQuery Advisor Release 4.0, Statistical Solutions, Cork, Ireland, Software for MS-DOS systems.

10-Freiman, J. A., Chalmers, T. C., Smith, Jr., H., and Kuebler, R. R. (1986), "The Importance of Beta, the Type II Error, and Sample Size in the Design and Interpretation of the Randomized Controlled Trial:

Survey of 71 "Negative" Trials," in Medical Uses of Statistics, eds. J. C.

11-Hintze, J. (2000), PASS 2000, Number Cruncher Statistical Systems, Kaysville, UT, Software for MS-DOS systems.

12-Hoenig, J. M. and Heisey, D. M. (2001), "The Abuse of Power: The Pervasive Fallacy of Power Calculations in Data Analysis," The American Statistician, 55, 19–24.

13- Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley and Sons, Inc.

14-Kraemer, H. C. and Thiemann, S. (1987), How Many Subjects? Statistical Power Analysis in Research, Sage Publications, Newbury Park, CA.

15- Sudman, Seymour. 1976. Applied Sampling. New York: Academic Press.

16- Yamane, Taro. 1967. Statistics, An Introductory Analysis, 2nd Ed., New York: Harper and Row.