V. Sample Size

Juan O. Talavera, Rodolfo Rivas-Ruiz, Laura Paola Bernal-Rosales, Lino Palacios-Cruz

In clinical research it is impossible and inefficient to study all patients with a specific pathology; therefore, it is necessary to focus on a sample. Estimating the size of a sample warrants the stability of the results and allows for feasibility of the study to be foreseen, depending on cost and patient availability. The basic structure for estimating the sample size is based on the premise that tries to demonstrate —among other things— that the difference between two or more maneuvers in the subsequent state is real. For this, it is necessary to know the value of the expected difference (δ) and the dispersion measure of the data that gave rise to it (standard deviation), which usually are obtained from previous studies. Afterwards, other components are considered: α, which is percentage of type I error accepted in the claim that the difference between means is real, generally of 5 %; and β, which is the percentage of type II error accepted in the claim that the non-difference between means is real, generally from 15 to 20 %. These values are substituted in the formula or in some sample size estimation electronic program. Although summary and dispersion measures may vary according to the outcome measure and, consequently, the formula, the principle is the same.

Key words
sample size
confidence interval

This article was originally published in Rev Med Inst Med Seguro Soc 2011; 49 (3): 289-294 and it has been reviewed for this issue.

Introduction

In clinical research, it is impossible and inefficient to study all subjects affected by a specific pathology; therefore, when we read an article, the results it shows correspond to a portion of the entire population. The number of subjects included in a study is determined by a series of features that will be addressed later, but whose primary objective is to answer a question with the certainty that the obtained result is real. In addition to this, estimation of the sample size before starting a study allows for its feasibility to be considered depending on patient availability and cost. The lack of calculation in the sample size may cause an unnecessary expenditure of both financial and human resources. It is possible for study expenses to be unnecessarily increased due to a surplus number of subjects included in it, or for the investment made to turn out being fruitless when including an insufficient number of subjects to answer the research question.

The basic structure of the sample size estimation is based on the premise that tries to demonstrate that the observed difference between measurements made before and after the maneuver, or between two maneuvers in the subsequent state, is real and not due to random effects. This structure is the same regardless of the type of variables necessary to answer the research question. In other cases, the purpose is not demonstrating the veracity of a difference but rather to obtain the average value of a particular feature within a population, with a precision indicated by the upper and lower limits of the confidence interval (CI), which in most cases is requested to be 95 or 99 %.

Estimation for Two Groups

This purpose is exemplified when we try to demonstrate that blood pressure values are different with a certain drug versus another and that this difference is not due to casuality. To estimate the sample size, the first thing that is required in this exercise is the average ($\bar{x}$) of the diastolic blood pressure (DBP) values of the patients that took one drug (group *A*) or another (group *B*): assuming that the average DBP in group *A* is 90 mm Hg and in group *B* 85 mm Hg, then the difference between means will be 5 mm Hg, a value that represents the first component, which is identified as delta (δ).

Afterwards, it will be necessary to have some measurement of the variation of values within each group, since there will be patients with much lower and much higher pressures than the average; for example, from 60 to 112 mm Hg. This value will allow for the variation within each group to be observed and, at the same time,

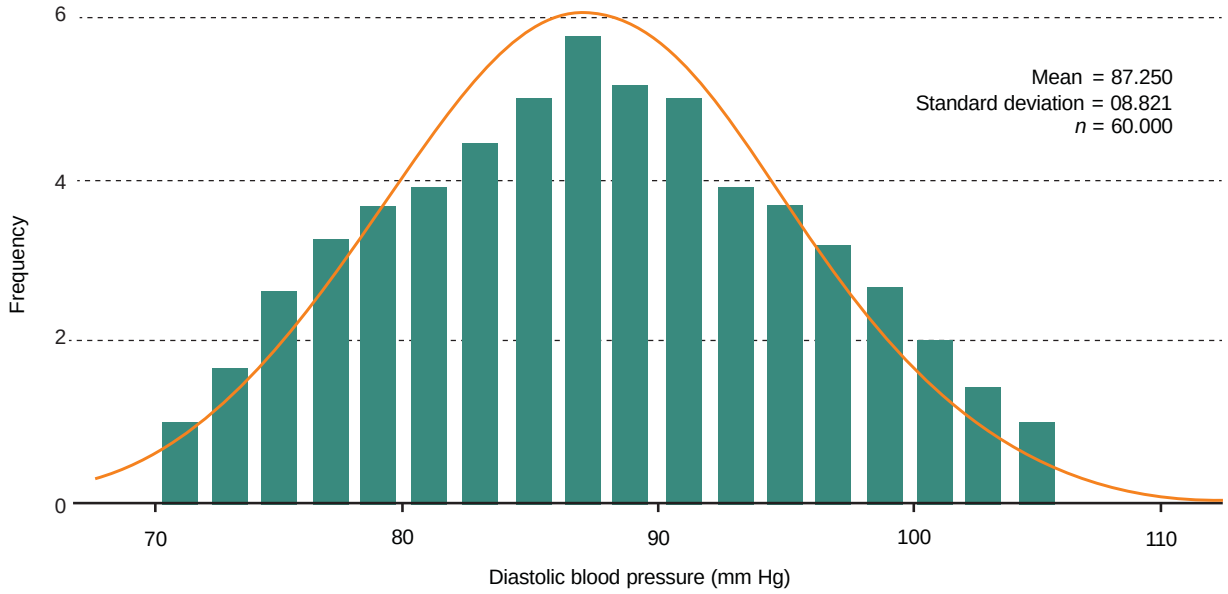

Figura 1 Total group of hypertensive patients under pharmacological treatment

to know if values between groups overlap excessively in relationship with the average difference. In a quantitative variable, as in the described model, the measure of dispersion is known as standard deviation (SD).

As is shown in Figure 1, the DBP average for the entire population is 87 mm Hg, with a standard deviation of 9 mm Hg, whereas in Figure 2a, DBP average in group A is 90 ± 9 mm Hg ($\bar{x} \pm$ SD) and DBP average (Figure 2b) in group B is 85 ± 8 mm Hg ($\bar{x} \pm$ SD). This means that the general population has an average of 87 mm Hg, but that its values in regards to two standard deviations range from 69 to 105 mm Hg ($\bar{x} \pm 2$ SD). In group A, with an average of 90 mm Hg, their values range from 72 to 108 mm Hg ($\bar{x} \pm 2$ SD), and in group B, with an average of 85 mm Hg, their values range from 69 to 101 mm Hg ($\bar{x} \pm 2$ SD). Average and variable of interest dispersion values are usually obtained from existing information in already published previous or preliminary studies.

Once we have a summary measure (average) and its corresponding measure of dispersion (DE), we have to consider:

1. To what degree of certainty do we want to demonstrate that the DBP difference between groups is real? When this point is not taken into account, we may incur in what is known as type I error: accepting that the difference is real without it being so.
2. To what degree of certainty do we want to demonstrate that the non-difference is real? When this point is not taken into account we may fall into what is known as type II error: accepting that the non-difference is real.

The certainty with which a difference is usually accepted to be real is at 95 % and this corresponds to an alpha value (α) of 0.05, indicating that once we establish that there is a difference in DBP values between groups, there is a 95 % of certainty that such difference is real and only a 5 % of error is accepted.

To accept that the non-difference found is real, we must have an initial pre-established capability to find significance when there is a difference, which is known as power and it is represented by the difference of 1 – beta (β). The accepted power value may vary from 80 to 95 %, which corresponds to a β-value of 20 to 5 % respectively.

At this point, all the components necessary for estimating the size of the sample are already available:

- δ: difference between the summary measures (in the example, it is the difference between the means).
- SD: measure of dispersion, which in the example is the standard deviation.
- Type I or α: error accepted in the claim that the difference between the means is real, usually of 5 % (0.05).
- Type II or β: error accepted in the claim that the non-difference between the means is real, generally ranging from 5 to 20 %.

Ignoring these different components usually causes that, at the end of the study, the size of the sample is insufficient and, thus, even if there is a clinically significant difference (≥ 10 %), no statistical difference is found ($p < 0.05$), which means insufficient power (< 80 %) and, therefore, a type II error.

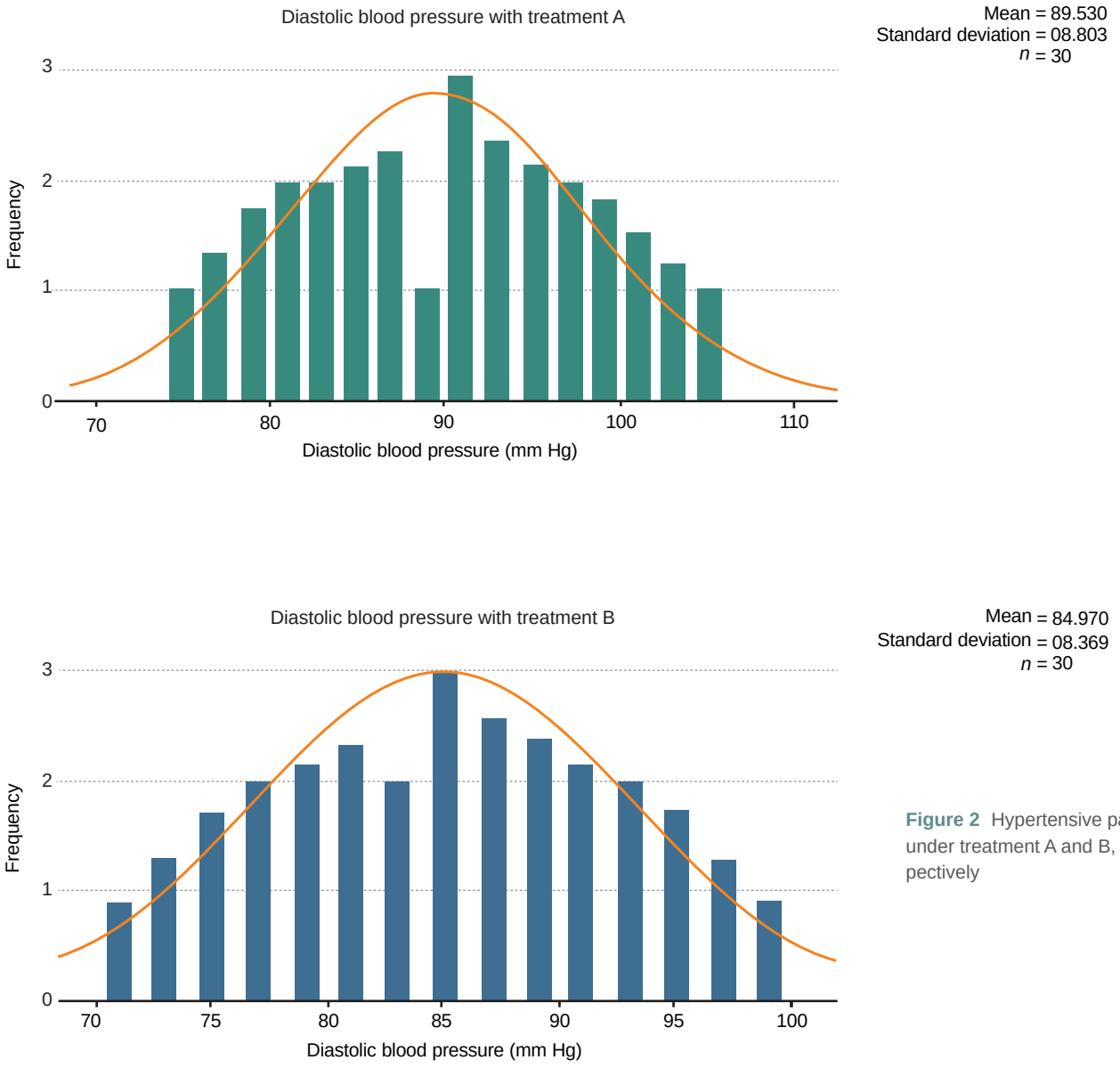

Figure 2 Hypertensive patients under treatment A and B, respectively

Mean Differences

With the above components, sample size is estimated using the formula of mean differences:

$$n = 2\left[\frac{(Z_\alpha - Z_\beta)\,DE}{\mu_1 - \mu_2}\right]^2$$

Where:

Z_α = value of *z* related to $\alpha = 0.05$ (extracted from reference tables)
Z_β = value of *z* related to $\beta = 0.20$ (80% power).
SD = standard deviation
μ_1 = group *A* mean
μ_2 = group *B* mean

According to the example, the substitution of values would be as follows:

Z_α = 1.96
Z_β = –0.84
SD = 9 mm Hg
μ_1 = 90 mm Hg
μ_2 = 85 mm Hg

And substituting in the formula:

$$n = 2\left[\frac{(1.96-(-0.84))9}{90-85}\right]^2 50.80 \approx 51$$

Therefore, it is necessary to include 51 patients in each group if obtaining 80 % of probabilities (80% power) is desired for the detection of a mean difference of 5 mm Hg or more between the two treatment groups.

Difference of Proportions

It is used when the outcome of interest is expressed in terms of proportions. Example: comparison of two groups of patients with overweight. The first group of patients receives medication and the second, dietary advice. If the outcome event is assessed after six months and measured as the proportion of patients who manage to normalize their weight (body mass index under 25), what is it required?

α = 0.05
β = 0.10
$\pi_1 - \pi_2$ = (difference of proportions) group 1 proportion minus group 2 proportion, which is clinically significant
SD = the formula for its determination is 1 – group proportion, which remains included within the global formula

The formula for the determination of the sample size for proportions difference is:

$$n = \left[\frac{Z_\alpha \sqrt{2\pi_1(1-\pi_1)} - Z_\beta \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)}}{\pi_1 - \pi_2} \right]^2$$

Where:

Z_α = (α = 0.05) 1.96
Z_β = ($\beta \approx 0.10 - 0.20$) ≈ –1.645, –0.84
π_1 = group 1 proportion
π_2 = group 2 proportion
$\pi_1 - \pi_2$ = difference between group 1 proportion – group 2 proportion, which is clinically significant

Assuming that for the study problem it would be expected that at six months, the group receiving drug therapy would succeed in 70 % of cases, whereas the group with dietary advice would succeed in 50 % of cases, the values would be replaced in the formula as follows:

$$n = \left[\frac{1.96 \sqrt{2 \times 0.70 \times 0.30} - (-1.645) \sqrt{(0.70 \times 0.30) + (0.50 \times 0.50)}}{0.70 - 0.50} \right]^2$$

$$n = \left[\frac{2.435}{0.20} \right]^2 = 12.18^2 = 148.35 \text{ subjects for each group}$$

This result must be rounded to the upper digit. Thus, the sample must include 149 subjects in each study group if 90 % of possibility (90 % power) is wanted for the detection of at least a difference of 20 % in the percentage of success in weight loss between the two treatment groups used as example.

Estimation for a Group

On the other hand, when the objective is to obtain the average value of a particular feature within a population, the sample size estimation requires the average value (proportion or mean) and its upper and lower limits indicated by the CI, which in most cases is requested at 95 or 99 %.

For a Proportion

To estimate the sample size for the prevalence or proportion of an event or feature, different components must be identified, starting with the summary measure (p_0), which corresponds to the expected proportion, and its precision (*d*), which is equivalent to half the amplitude of the CI. If we understand this section, we can solve the sample size formula based on the precision formula, which in turn comes from the estimation of the standard deviation of a proportion:

$$d = Z_\alpha \sqrt{\frac{p_0 \times q_0}{n}}$$

Solving for *n* yields:

$$n = \frac{Z_\alpha^2 \times p_0 \times q_0}{d^2}$$

In this case, $q_0 = (1 - p_0)$; therefore if we want to look for a prevalence (p_0) of 20 %, the q_0 value would be 1 – 0.2 = 0.8. Therefore, to make the calculation of the sample size for a proportion, the following must be considered:

- Precision (*d*, equal to ½ the amplitude of the CI), whose value is conferred by the investigator and corresponds to the degree of error that might be tolerated at each side of the mean; for example, for an error of 8 % based on the mean, its d2 would be 0.0064 (0.082 = 0.0064).
- Confidence, also known as Z_α corresponds to 1 – α.
- The p_0 value intended to be estimated.

Example: How many preterm infants will it be necessary to study in order to verify if the estimated prevalence of metabolic bone disease in a neonatal intensive care unit population is 20 %, considering an accuracy of 8 % and an α of 0.05 %?

With a confidence level of 95 % ($\alpha = 0.05$; $Z_\alpha = 1.96$), $Z_\alpha^2 = 3.8416$, which when solving:

$$N = (3.8416 \times 0.2 \times 0.8)/0.0064$$
$$N = 96.04$$

Therefore, the required sample size will be 97 children for an expected prevalence of 20 % with a CI ranging from 12 to 28 %.

As we can observe, the size of the sample will depend on the expected accuracy of the error based on the mean, so that for a narrower CI, a lower *d* is required; 0.08 and 0.04 values are generally used, with the latter being the most accurate (or the one with less error); therefore, a larger sample size will be required. Similarly, if a confidence level change from 95 to 99 % is desired, as requested in studies of genetic determinants, the sample size will increase. Table I shows some variation examples according to these parameters.

For a Mean

If the above is understood, it will be easy to understand the components for estimating the sample size for a mean. Similarly, the basis is the formula for the CI of the mean:

$$IC\ de\ 95\ \% = \bar{x} \pm Z_\alpha = \frac{DE}{\sqrt{n}}$$

In this case, precision (*d*) is calculated as follows:

$$d = Z_\alpha \frac{DE}{\sqrt{n}}$$

Therefore, the formula for the calculation of the sample size for estimating a mean is:

$$n = \frac{Z^2 \times DE^2}{d^2}$$

This formula requires the knowledge of Zα, SD and the desired *d*. Thus, the sample size for an expected mean depends on Zα (1.96 for $\alpha = 0.05$), on the standard deviation that has been observed in previous studies, as well as on the desired precision.

Final Considerations

It should be clear that the assumptions above are not the only ones for estimating the size of a sample, so that if we want to estimate it in order to demonstrate differences in cumulative incidence rates (Hazard risk ratio) or in units obtained in models such as Cox proportional hazards survival curves, the estimation is more complex since it considers the outcome over time; nevertheless, the basic concept is the same.

On the other hand, if the intention is controlling for multiple confounders or exploring multiple risk factors using a multiple logistic regression model, then it will be necessary using a number of events per variable, for which 10 to 20 subjects for each will be required in the smallest of the outcome groups (so that if mortality is 30 %, this is the smallest of the groups, since the remaining 70 % will survive).

Table I Different sample sizes according to different values of confidence level (α), prevalence (*p*) and precision (*d*)

a (Z_α)	p	d	n
0.05(1.960)	0.2	0.08	97
0.05(1.960)	0.2	0.04	385
0.01(2.576)	0.2	0.08	166
0.01(2.576)	0.2	0.04	664

Recommended readings

Cohen J. Statistical power analysis for the behavioural sciences. Second edition. New Jersey: Lawrence Earlbaum; 1988.
Dawson B, Trapp R. Basic and clinical biostatistics. Fourth edition. Lange Medical Books-McGraw-Hill; New York, USA; 2004.
Hulley SP, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma; 1993.
Portney LG, Watkins MP. Foundations of clinical research. Applications to practice. Appleton and Lange; Chicago, IL: 1993.

Feinstein AR. Principles of medical statistics. London, UK: Chapman and Hall-CRC; 2002.

For the calculation of sample size

Brixton Health. [Website]. Epicalc 2000. Available at http:// www.brixtonhealth.com/epicalc.html

EpiInfo 2000. Available at http://huespedes.cica.es/huespedes/epiinfo/espanol.htm

Department of Biostatistics, Vanderbilt University. [Website]. PS: Power and Sample Size Calculation. Available at http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize