

## Cálculo del tamaño de la muestra: un paso importante en el diseño investigativo

### Calculating the sample size: an important step in investigative research

Leonardo Elías Ordóñez Ordóñez, MD\* , Esther Sofía Angulo Martínez, MSc\*\*

#### RESUMEN

*Calcular el tamaño de la muestra que se necesita para aprobar o rechazar la hipótesis nula es una parte importante en el diseño investigativo. Para realizar este cálculo debemos definir el tipo de variable utilizada para medir el resultado, la diferencia que queremos determinar si existe (diferencia mínima clínicamente significativa), la desviación estándar de la variable de resultado (usualmente obtenida al revisar la literatura), el poder o potencia estadística (usualmente  $\geq 0.8$ ) y el valor del error tipo I (usualmente  $\leq 0.05$ ). Aunque son muchas las publicaciones que no incorporan este aspecto metodológico, cabe esperar que con la actual disponibilidad y accesibilidad a programas informáticos, el tamaño de muestra se calcule con más frecuencia en los trabajos de investigación en Otorrinolaringología. A través de un ejemplo se hace una explicación práctica sobre la forma de calcular el tamaño de la muestra utilizando un programa informático de buena aceptación.*

**Palabras clave:** *Tamaño de muestra, protocolo de investigación clínica. Diseño de investigación epidemiológica, diferencia mínima clínicamente significativa.*

---

\* Otológo-Otorrinolaringólogo. Clínica Universitaria Colombia, Hospital Universitario Clínica San Rafael.

\*\* Magíster en Infecciones y Salud en el Trópico. Epidemióloga.

#### Correspondencia

Leonardo Elías Ordóñez Ordóñez.  
Servicio de Otorrinolaringología.  
Hospital Militar Central.  
Transversal 3 # 49-00, tercer piso norte. Bogotá.  
E-mail: otoleor@gmail.com, kokhlias@gmail.com.

Recibido: 10/XI/2009

Aceptado: 20/XI/2009

## ABSTRACT

*Calculating the size of the sample needed to accept or reject the null hypothesis is an important part in investigative design. In order to figure out this calculation it is necessary to define the type of variable used to measure the outcome, the difference to be determined if any (minimal clinically significant difference), the standard deviation of the outcome variable (usually obtained by reviewing the literature), the statistical power (usually = 0.8) and the value of Type I error (usually = 0.05). Although many publications do not incorporate this methodological aspect, it may be expected that given the current availability and accessibility of statistical software, calculating the size of the sample shall be done on a more frequent basis in otolaryngology research works. Through an example, a practical explanation is given on how to calculate the size of the sample using a well- received computer program.*

**Key words:** *Sample Size, Clinical Protocols, Epidemiological Research Design, Minimal Clinically Important Difference.*

## INTRODUCCIÓN

Durante la realización de una investigación clínica un aspecto importante del diseño es el cálculo del tamaño de la muestra necesario para aceptar o rechazar nuestra hipótesis de trabajo. La relevancia científica de un estudio está determinada en parte por el tamaño de la muestra del mismo. Muchos artículos que a primera vista parecen tener gran relevancia clínica, terminan por no dar una respuesta definitiva a la pregunta de investigación debido a que su tamaño de muestra es inadecuado.

En un estudio sobre la calidad de los estudios de investigación publicados en las principales revistas de otorrinolaringología, se encontró que el tamaño de la muestra utilizado es inferior al encontrado en artículos de otras especialidades quirúrgicas (1). Esto nos deja con abundante literatura que no responde de forma clara a la pregunta de investigación planteada.

Los cálculos de tamaño de muestra tienen fórmulas matemáticas complejas (2); sin embargo con la posibilidad de utilizar programas informáticos esta tarea se reduce. Hay varios programas disponibles como SigmaStat, EpiInfo o PS (Power and Sample Size Calculations). En adelante vamos a continuar refiriéndonos a los cálculos utilizando el programa PS, ya que es fácil de utilizar, de amplia aceptación y su descarga es gratuita (disponible en: [http://biostat. mc. vanderbilt.edu/wiki/Main/PowerSampleSize](http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize)).

Vamos a utilizar un ejemplo para facilitar las explicaciones, supongamos que usted maneja de forma preferencial sus pacientes de parálisis unilateral de cuerda vocal en posición paramediana con inyección grasa (Procedimiento 1: P1); sin embargo, desea determinar si

en estos casos la tiroplastia de medialización con molde de silicona (Procedimiento 2: P2) puede dar resultados significativamente mejores. Decide emprender un estudio prospectivo para responder a esta inquietud, comparando los resultados del Procedimiento 1 (Grupo I), con los del Procedimiento 2 (Grupo II). Tan pronto como comienza a diseñar su estudio se encuentra con tres cuestiones importantes por resolver: 1. ¿Cómo va a medir el resultado? (¿qué variable utilizar?), 2. ¿Cómo definir “resultados significativamente mejores”? y 3. ¿Cuántos pacientes necesita para responder de forma categórica a su pregunta de investigación? Para responder a la primera pregunta hace una revisión de la literatura y encuentra que el tiempo máximo de fonación (TMF) medido en segundos a los 24 meses posoperatorios es una variable que considera adecuada para comparar los resultados (puede que usted considere que se requiere medir varias variables como la frecuencia fundamental, Jitter, Shimmer, hallazgos estroboscópicos y valoraciones preceptuales; en ese caso debe hacer los cálculos para cada variable por separado).

Para determinar que son “resultados significativamente mejores” se debe determinar lo que se ha llamado: la diferencia mínima clínicamente significativa (DMCS). La DMCS es diferente al concepto de significación estadística, es un concepto bastante subjetivo y que debe definir el investigador. Una aproximación es determinar cuál sería la diferencia que de encontrarse haría que cambiemos nuestra forma de hacer las cosas. En nuestro ejemplo, ¿cuál sería la diferencia en el TMF que haría que en adelante prefiramos el P2 sobre el P1? Nuevamente revisa la literatura y encuentra que el valor normal del TMF es de 20 segundos (3) y que hay un estudio en el cual se logró una mejora de 4.7 seg. con la inyección grasa (preoperatorio=4.5 y posoperatorio=9.2) (4). Usted considera que si se logra una mejoría con el P2 de 6.7 seg., es decir 2 seg. más que con el P1, sería una diferencia

que haría que en adelante usted prefiera la tiroplastia sobre la inyección grasa (es decir 2 seg. es lo que considera como la DMCS).

Determinar cuántos pacientes necesita para responder a su pregunta de investigación es el objeto de este artículo y vamos a discutir cuáles son los factores que debe tener en cuenta para hacer los cálculos.

## FACTORES QUE DETERMINAN EL TAMAÑO DE LA MUESTRA

### Tipo de variable utilizada para medir el resultado

Determinar el tipo de variable de acuerdo al nivel de medición es un aspecto fundamental para saber qué prueba estadística se va a utilizar en los análisis (5), y para hacer el cálculo del tamaño de la muestra se requiere conocer el tipo de prueba con que se analizarán los resultados. Recordemos que los tipos de variables de acuerdo al nivel de medición son: nominal, ordinal, interválica/proporcional (5). Los tamaños de muestra varían de acuerdo al tipo de variable, siendo mayores los tamaños de muestra (para una diferencia equiparable) para variables nominales > ordinales > interválicas/proporcionales. Si para un mismo resultado existen dos o más variables para medirlo de forma apropiada, se prefiere las interválicas/proporcionales > ordinales > nominales.

Para el ejemplo en cuestión, el TMF es una variable proporcional y la prueba que se utilizaría para comparar los resultados del P1 con el P2 sería una prueba t de Student de muestras independientes. Para iniciar los cálculos programas como el PS piden comenzar por seleccionar el tipo de prueba, Figura 1.

### Diferencia entre los grupos

Hace referencia a la diferencia en las medidas de tendencia central de cada grupo (puede ser medias o proporciones, dependiendo del tipo de variable). Este valor que se simboliza con la letra griega  $\delta$ , suele obtenerse de estudios previos en los que se haya hecho la misma (o muchas veces semejante) comparación, pero puede tomarse también como la DMCS. La naturaleza de la variable determina si la diferencia entre los grupos es grande o pequeña. De forma general, cuanto más grande es la diferencia entre los grupos, menor es el tamaño de muestra y viceversa. Si para un mismo resultado existen dos o más variables para medirlo de forma apropiada, se prefiere aquella que muestre las mayores diferencias entre los grupos, de forma que nos reduzca el tamaño de muestra del estudio. Para nuestro ejemplo vamos a usar la DMCS, de forma que  $\delta=2$ .

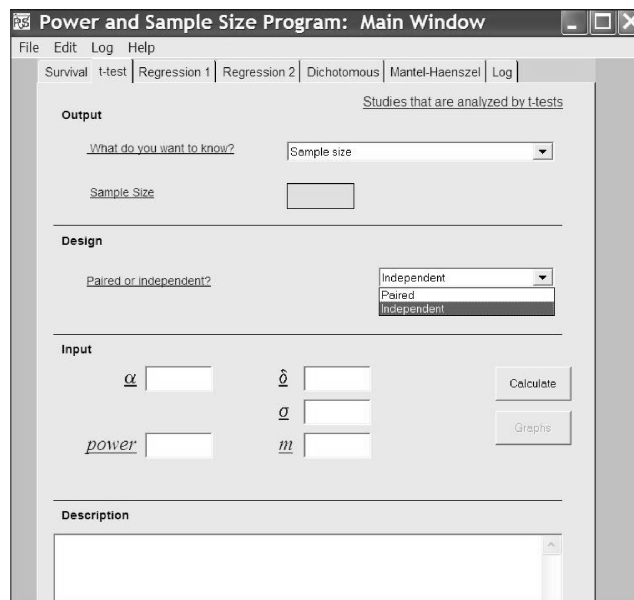


Figura 1. Ventana principal del programa PS, para iniciar el cálculo se selecciona el tipo de prueba ("t-test") y de diseño ("independent").

### Dispersión de los valores (desviación estándar)

La medida de dispersión de los valores alrededor de la medida de tendencia central se conoce como desviación estándar, y se simboliza con la letra griega  $\sigma$ . La naturaleza de la variable determina si la dispersión es grande o pequeña. De forma general, cuanto menor es la desviación estándar, menor será el tamaño de muestra. Este valor se obtiene de los resultados de estudio previos sobre el tema, para nuestro ejemplo  $\sigma$  es la desviación estándar del TMF, para el Grupo I=2.5 seg. (4) y el Grupo II=2.9 seg. (6), con lo que para nuestro ejemplo  $\sigma=2.7$ .

### Valor del error tipo I y poder de la muestra

Son valores que determinan que tantos resultados falsos positivos (error tipo I ó  $\alpha$ ) o falsos negativos (poder, p) están dispuestos a aceptar los investigadores. Cuando  $\alpha=0.05$ , queremos decir que estamos dispuestos a aceptar un resultado falso positivo el 5% de las veces. Cuando  $p=0.80$ , queremos decir que nuestros datos son lo suficientemente poderosos como para evitar un resultado falso negativo el 80% de las veces.

Si el valor de  $\alpha$  se disminuye o el poder deseado se aumenta, mayor será el tamaño de muestra y viceversa. Es común en la práctica clínica usar valores de  $\alpha \leq 0.05$  y de poder  $\geq 0.8$ . Para nuestro ejemplo  $\alpha=0.05$  y de  $p=0.8$ .

## Cálculo del tamaño de la muestra

Una vez definidos los anteriores aspectos se procede a realizar el cálculo del tamaño de la muestra. En la ventana del programa PS Usted selecciona el tipo de prueba estadística (para nuestro ejemplo prueba t de Student “t-test”) y el tipo de diseño (para nuestro ejemplo medidas independientes “Independent”) e ingresa los valores de  $\delta$ ,  $\sigma$ ,  $\alpha$  y p. Este programa nos pide un dato adicional, m, que es la proporción entre los pacientes del grupo control (Grupo I) y el grupo experimental (Grupo II). Para nuestro ejemplo Grupo I/Grupo II=1, es decir vamos a utilizar un paciente con inyección grasa por cada paciente con tiroplastia de medialización.

Ingresamos los anteriores valores y encontramos que el tamaño de muestra, n, es de 30, Figura 2. Esto significa que necesitamos 30 pacientes en el Grupo I y 30 pacientes en el Grupo II, para determinar si se presenta una diferencia de 2 segundos entre los Grupos I y II y poder rechazar la hipótesis nula de que no hay diferencias entre el Grupo I y Grupo II. Lo anterior con un poder  $\geq 0.8$  y un error tipo I,  $\alpha \leq 0.05$ . En la parte inferior de la ventana del PS se encuentra este texto explicativo para cada cálculo de tamaño de muestra que realicemos.

The screenshot shows a software window with a menu bar (File, Edit, Log, Help) and a toolbar (Survival, t-test, Regression 1, Regression 2, Dichotomous, Mantel-Haenszel, Log). The main area is titled "Studies that are analyzed by t-tests". It has four sections: "Output" with a dropdown for "What do you want to know?" set to "Sample size" and a text box for "Sample Size" containing "30"; "Design" with a dropdown for "Paired or independent?" set to "Independent"; "Input" with fields for  $\alpha$  (0.05),  $\delta$  (2),  $\sigma$  (2.7), power (0.8), and m (1), along with "Calculate" and "Graphs" buttons; and "Description" with a text box containing the following text: "We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 2.7. If the true difference in the experimental and control means is 2, we will need to study 30 experimental subjects and 30 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05."

Figura 2. Luego de ingresar los valores de  $\delta$ ,  $\sigma$ ,  $\alpha$ , p y m obtenemos el tamaño de muestra calculado. Observe el texto explicativo que aparece en el recuadro inferior de la ventana.

Para hacer cálculos de tamaños de muestra para otros tipos de pruebas se requiere información similar a la presentada en este ejemplo, con algunas variaciones, como se muestra en la Tabla 1 (7).

Para concluir, calcular el tamaño de la muestra es un paso esencial a la hora de diseñar una investigación. Una buena parte de la relevancia de un estudio depende de si consiguió el tamaño de muestra necesario para responder de forma clara a la pregunta de investigación. Infortunadamente se encuentran muchos artículos en los cuales no se calcula el tamaño de muestra necesario para aceptar o rechazar la hipótesis de trabajo. Uno de los factores que influye en la no inclusión de este aspecto metodológico es la complejidad de los cálculos; sin embargo, con la disponibilidad y accesibilidad a los programas informáticos la tarea se facilita y está al alcance de todos nosotros.

## BIBLIOGRAFÍA

1. Bhattacharyya N. Peer review: studying the major otolaryngology journals. *Laryngoscope* 1999; 109: 640-644.
2. Norman GF, Streiner DL. *Bioestadística*. 1st ed. Madrid: Mosby/Doyma Libros. 1996.
3. Fried MP. *The larynx*. 2nd ed. St. Louis: Mosby-Year Book. 1996; 119-120.
4. Umeno H, Shirouzu H, Chitose S, Nakashima T. Analysis of voice function following autologous fat injection for vocal fold paralysis. *Otolaryngol Head Neck Surg* 2005; 132: 103-107.
5. Angulo Martínez ES, Ordóñez LE. Cómo construir una base de datos. *Acta de Otorrinolaringología & Cirugía de Cabeza y Cuello* 2008; 36: 205-215.
6. Shin JE, Nam SY, Yoo SJ, Kim SY. Analysis of voice and quantitative measurement of glottal gap after thyroplasty type I in the treatment of unilateral vocal paralysis. *J Voice* 2002; 16: 136-142.
7. Neely JG, Karni RJ, Engel SH *et al*. Practical guides to understanding sample size and minimal clinically important difference (MCID). *Otolaryngol Head Neck Surg* 2007; 136: 14-18.

**Tabla 1**  
**Información necesaria para hacer cálculos de tamaños de muestra en investigación clínica**

Tipo de prueba estadística	Información requerida
<p><b>Prueba t de Student</b></p> <p><b>Variable dependiente:</b> interválica/proporcional  <b>Variable independiente:</b> nominal con dos posibles valores: Ej. Grupo I/Grupo II</p>	<ul style="list-style-type: none"> <li>• Diferencia entre las medias a ser detectada (<math>\delta</math>)</li> <li>• Desviación estándar de la media (<math>\sigma</math>)</li> <li>• Poder deseado (<math>p</math>, usualmente <math>\geq 0.80</math>)</li> <li>• Error tipo I (<math>\alpha</math>, usualmente <math>\leq 0.05</math>)</li> <li>• Relación Grupo I/Grupo II, <math>m</math></li> </ul>
<p><b>Prueba t de medidas repetidas (pareada)</b></p> <p><b>Variable dependiente:</b> interválica/proporcional  <b>Variable independiente:</b> nominal con dos posibles valores: Ej. pretratamiento/postratamiento</p>	<ul style="list-style-type: none"> <li>• Diferencia (cambio) entre las medias a ser detectada (<math>\delta</math>)</li> <li>• Desviación estándar esperada de la diferencia (<math>\sigma</math>)</li> <li>• Poder deseado (<math>p</math>, usualmente <math>\geq 0.80</math>)</li> <li>• Error tipo I (<math>\alpha</math>, usualmente <math>\leq 0.05</math>)</li> </ul>
<p><b>Proporciones dicotómicas</b></p> <p><b>Variable dependiente:</b> nominal con dos posibles valores: Ej. éxito/fracaso  <b>Variable independiente:</b> nominal con dos posibles valores: Ej. Grupo I/Grupo II</p>	<ul style="list-style-type: none"> <li>• Proporción esperada en el Grupo I (<math>P_0</math>)</li> <li>• Proporción esperada en el Grupo II (<math>P_1</math>)</li> <li>• Poder deseado (<math>p</math>, usualmente <math>\geq 0.80</math>)</li> <li>• Error tipo I (<math>\alpha</math>, usualmente <math>\leq 0.05</math>)</li> <li>• Relación Grupo I/Grupo II, <math>m</math></li> </ul>
<p><b>Proporciones con más de dos categorías</b></p> <p><b>Variable dependiente:</b> nominal con dos posibles valores: Ej. éxito/fracaso  <b>Variable independiente:</b> nominal con más de dos posibles valores: Ej. Grupo I/Grupo II/Grupo III</p>	<ul style="list-style-type: none"> <li>• Número estimado de observaciones en cada celda de la tabla de contingencia</li> <li>• Poder deseado (<math>p</math>, usualmente <math>\geq 0.80</math>)</li> <li>• Error tipo I (<math>\alpha</math>, usualmente <math>\leq 0.05</math>)</li> </ul>
<p><b>ANOVA</b></p> <p><b>Variable dependiente:</b> interválica/proporcional  <b>Variable independiente:</b> nominal con más de dos posibles valores: Ej. Grupo I/Grupo II/Grupo III</p>	<ul style="list-style-type: none"> <li>• Mínima diferencia detectable entre las medias (<math>\delta</math>)</li> <li>• Desviación estándar esperada del residuo (<math>\sigma</math>)</li> <li>• Número de grupos</li> <li>• Poder deseado (<math>p</math>, usualmente <math>\geq 0.80</math>)</li> <li>• Error tipo I (<math>\alpha</math>, usualmente <math>\leq 0.05</math>)</li> </ul>
<p><b>Correlaciones</b></p> <p><b>Variable dependiente:</b> interválica/proporcional  <b>Variable independiente:</b> interválica/proporcional</p>	<ul style="list-style-type: none"> <li>• Coeficiente de correlación esperado (<math>r</math>)</li> <li>• Poder deseado (<math>p</math>, usualmente <math>\geq 0.80</math>)</li> <li>• Error tipo I (<math>\alpha</math>, usualmente <math>\leq 0.05</math>)</li> </ul>

