Estimación del tamaño de la muestra en la investigación clínica y epidemiológica

Jaume Marrugat, Joan Vila, Marco Pavesi y Ferran Sanz

Unidad de Lípidos y Epidemiología Cardiovascular. Unidad de Informática Médica.
Instituto Municipal de Investigación Médica (IMIM). Barcelona.

«Dadme el número suficiente de pacientes y demostraré corno estadísticamente significativa cualquier diferencia por pequeña que sea». Esta paráfrasis de Arquímedes, cuando se refería al uso de la palanca para levantar un mundo, ilus-tra la situación con la que algunos investigadores se han tenido que enfrentar en la última década. En efecto, el beneficio terapéutico obtenido con algunos nuevos fármacos es a veces tan pequeño que para establecerlo con la suficiente seguridad se ha precisado un elevado número de pacientes. El estudio GUSTO (Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries) sobre la eficacia de la fibrinólisis en el infarto de miocardio[1] constituye un ejemplo reciente en el que fueron precisos más de 20.000 sujetos para demostrar como estadísticamente significativa una diferencia de mortalidad de poco menos de una unidad porcentual entre dos tratamientos trombolíticos (el 7,2% con estreptocinasa y el 6,3% con la administración rápida del activador del plasminógeno tisular). La situación opuesta, aquella en la que las diferencias clínica o biológicamente relevantes no se detectan como estadísticamente significativas como consecuen-cia de trabajar con tamaños muestrales insuficientes es, sin embargo, más frecuente[2,3].

Otros trabajos han abordado algunos aspectos parciales del cálculo del tamaño de la muestra con anterioridad[4,6]. El objetivo del presente artículo es revisar los conceptos y la metodología empleada para este cálculo en un número mayor de pruebas estadísticas aplicadas en los diseños más frecuentemente utilizados en la investigación clínica y epidemiológica. Se persigue ofrecer al lector una orientación práctica en un punto crucial de la elaboración de un proyecto de investigación. Con este propósito, el artículo incluye algoritmos para contrastes de hipótesis particulares como, por ejemplo, la comparación de proporciones inferiores al 5% o de curvas de supervivencia.

¿Por qué estimar el tamaño de la muestra? El principio general que justifica trabajar con muestras es que resulta más barato, más rápido, más fácil y es más exacto (en el sentido de que es posible realizar medidas más precisas y exhaustivas) que hacerlo con poblaciones completas[7]. A partir de los datos observados en la muestra, se realizarán pruebas estadísticas que permitan generalizar los resultados a la población de la que proceden, con una mínima probabilidad de error (en general < 5%). Las estimaciones del número mínimo de sujetos necesario para ello se deben basar en la prueba estadística que se prevea utilizar. Por otro lado, el diseño de un proyecto de investigación debe incluir la estimación del tamaño de la muestra a utilizar. De este modo, el investigador se ve obligado a precisar la magnitud y la dirección de sus hipótesis principales de trabajo. La estimación del tamaño muestra! puede considerarse, por lo tanto, un instrumento del que dispone el investigador para evaluar la factibilidad y la necesidad de recursos de su proyecto. En el cálculo del tamaño muestral, la utilización de hipótesis verosímiles deberá prevalecer sobre otros intereses de los investigadores como las posibilidades económicas, la disponibilidad de pacientes u otros. Cabe recordar que no es ético realizar un estudio con un tamaño de muestra que no ofrezca un poder estadístico suficiente, ya que, desde el punto de vista de la metodología científica, el diseño no es adecuado.

¿Qué información se necesita para la estimación del tamaño de la muestra? El resultado de calcular el tamaño de la muestra es una *estimación,* ya que es imposible predecir los hallazgos reales del estudio que se diseña. Se parte, por lo tanto, de datos hipotéticos basados, en el mejor de los casos, en experiencias del pasado. Como toda estimación, está sujeta a error -en realidad a errores, ya que se precisan varios elementos para realizar los cálculos-, cuya magnitud dependerá de la validez de los datos de referencia utilizados. A pesar de que existe un buen número de diseños de investigación clínica y epidemiológica, la estimación del tamaño de la muestra requiere un número de ecuaciones pequeño (véase el apéndice) que se adaptan a cada circunstancia. Los factores que condicionan el número de individuos que pueden participar en un estudio pueden ser cuestiones de *orden logístico* -como las limitaciones financieras, el número de pacientes disponibles en el caso de enfermedades infrecuentes y las propias necesidades de tiempo del estudio-en las que no entraremos. En segundo lugar, existen factores de *orden estadístico* (tabla 1) que condicionan el cálculo del tamaño de la muestra:

1. El error α, o riesgo de primera especie, constituye la pro-babilidad de rechazar en una prueba estadística la hipótesis nula cuando la alternativa es, en realidad, falsa: 1-α es el nivel de confianza de esta prueba. El criterio más corriente es aceptar un riesgo $\alpha \leq 0,05$.

2. El error β, o riesgo de segunda especie, representa la probabilidad de error al rechazar la hipótesis alternativa cuando, en realidad, es cierta. El criterio más corriente es aceptar un riesgo β entre 0,10 y 0,20.

TABLA 1
Tipos de error posibles al tomar una decisión en los contrastes de hipótesis

		REALIDAD	
		H_0	H_1
DECISIÓN	H_1	α	1 - ß
	H_0	1 - α	ß

Hipótesis alternativa (H_1): existe un efecto de la intervención o exposición. Hipótesis nula (H_0): no existe tal diferencia. El error α (error de tipo 1 o riesgo de primera especie) constituye la probabilidad de rechazar en una prueba estadística la hipótesis nula cuando es, en realidad, cierta: (1 - α) es el nivel de confianza de esta prueba. El error β (error de tipo II o riesgo de segunda especie) representa la probabilidad de error al rechazar la hipótesis alternativa cuando, en realidad, es cierta: (1 - β) es el poder estadístico de la prueba

Correspondencia: Dr. J. Marrugat.
Unidad de Lípidos y Epidemiología Cardiovascular.
Instituto Municipal de Investigación Médica (IMIM).
Doctor Aiguader, 80. 08003 Barcelona

Manuscrito aceptado el 1-4-1997

Med Clin (Barc) 1998; 111: 267-276

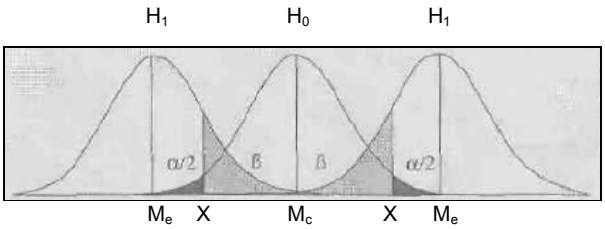

Fig. 1. Distribución de los valores de una variable continua según la hipótesis nula y alternativa en un contraste bilateral H_1: hipótesis alternativa (existe un Efecto de la intervención o exposición): H_0: hipótesis nula (no existe un efecto de la intervención o exposición); α: riesgo de primera especie: constituye la probabilidad de rechazar en una prueba estadística la hipótesis nula cuando la alternativa es, en realidad, falsa; β: riesgo de segunda especie: representa la probabilidad de error al rechazar la hipótesis alternativa cuando, en realidad, es cierta: M_e: media en el grupo expuesto o intervenido; M_c: media en el grupo control (no expuesto o no intervenido); X: punto de decisión para el re-chazo/aceptación de la H_1 correspondiente al valor de la distribución normal asociada a un riesgo α/2.

3. El poder estadístico (1 - β) de la prueba se define como la probabilidad de rechazar la hipótesis nula cuando es cierta la alternativa.

En términos generales, se debería escoger un riesgo α igual al riesgo β si los tratamientos de ambos grupos son nuevos, de coste similar y hay buenas razones para considerar que los dos son relativamente seguros. Es posible escoger un riesgo a mayor que el riesgo β cuando no existe un tratamiento establecido y el que se desea probar es relativamente barato, fácil de aplicar y no tiene serios efectos secundarios conocidos. Finalmente, la situación más habitual es aquella en que se dispone de un tratamiento ya establecido, seguro y eficaz con el que se desea comparar un tratamiento nuevo: en este caso, el riesgo a debería ser inferior al riesgo β, ya que se considera que las eventuales consecuencias de un error de tipo II serían menos importantes que las de un error de tipo I [4,7].

4. La variabilidad de la medida. Cuanto más se agrupen los valores individuales de la variable estudiada alrededor de uno central, se requerirán menos individuos. En el caso de las variables continuas, la de mayor coeficiente de variación (CV = DE/X, donde DE es la desviación estándar, y X es la media) requerirá la muestra más grande. En el caso de las categóricas debe utilizarse la estimación de la proporción que se acerque más al 50%. En caso de existir hipótesis con ambos tipos de variables se utilizará la categórica para el cálculo del tamaño de la muestra, puesto que generalmente obligará a reclutar un mayor numero de sujetos.

5. El tipo de contraste de hipótesis (unilateral o bilateral) que se desea realizar (fig. 1): cabe recordar que el punto Z de la distribución normal correspondiente al riesgo a aceptado (Z_α cambia según el tipo de contraste, siendo $Z_{\alpha/2}$ en el bilateral y Z_α en el unilateral. Es recomendable utilizar siempre el contrasté bilateral ya que ofrece la estimación más conservadora. El contraste unilateral debería reservarse exclusivamente para aquellas circunstancias en que una diferencia en un sentido llevaría a la misma acción que la inexistencia de diferencias[8,9]. El valor correspondiente al riesgo β aceptado en ambos tipos de contraste es Z_β (unilateral).

6. El incumplimiento de las intervenciones, las retiradas del estudio, las pérdidas de seguimiento y los d*rop-in* (administración de un tratamiento en estudio diferente al asignado por aleatorización o de un tratamiento prohibido en el ensayo) han de preverse en el cálculo. De hecho, el tamaño de la muestra estimado se refiere a los pacientes disponibles para el análisis al final del estudio y no a los inicialmente incluidos.

En las enfermedades graves, un incremento de la muestra entre un 10 y un 20% suele ser suficiente para compensar las pérdidas de seguimiento debido a la alta dependencia que presenta el enfermo del sistema sanitario. Sin embargo, en patologías más banales et número de pérdidas puede superar el 50%[7,10,11].

7. La distribución de referencia. La hipótesis de trabajo se centra principalmente en una variable de interés que puede ser categórica o continua. El teorema del límite central nos dice que la distribución de las proporciones o de las medias de todas las muestras posibles que se pueden obtener con esta variable sigue una distribución aproximadamente normal cuando las muestras son de tamaño grande, en general n > 30 para variables continuas y tanto (n*p) como (n*[l -p]) > 5 en variables categóricas. Cuando estas condiciones no se cumplen deberían contemplarse otras distribuciones como la binomial exacta, la de Poisson, la de Student. etc.

8. La magnitud de la diferencia del efecto a detectar entre los grupos en el acontecimiento de interés. La variable que mida dicho acontecimiento puede ser categórica (presencia o ausencia de curación, por ejemplo) o continua (como las concentraciones plasmáticas de colesterol total). Cuando exista más de una, deberá considerarse aquella que requiera el mayor número de individuos. La magnitud de la diferencia del efecto a detectar como estadísticamente significativa será el condicionante más importante de los cálculos, ya que el resto de valores puede fluctuar relativamente poco (como la Z_α, la Z_β y el tipo de contraste) o no depende del investigador (variabilidad). El criterio para establecer la magnitud de esta diferencia es enteramente clínico o epidemiológico. Por ejemplo, es a todas luces claro que una reducción en la mortalidad de 5 unidades porcentuales (supongamos del 20% en el grupo control al 15% con un tratamiento nuevo) representa una mejora importante desde el punto de vista clínico. Para detectar este diferencia como estadísticamente significativa se requerirían unos 1.100 sujetos en cada grupo en condiciones de cálculo estándar (α = 0,05, β = 0,20 en un contraste bilateral). Sin embargo, no está tan claro que la reducción de una unidad porcentual en la mortalidad represente un beneficio terapéutico relevante. En el supuesto de que dicha unidad porcentual representara la diferencia entre el 20 y el 19% serían necesarios unos 26.000 sujetos en cada grupo. De estos ejemplos se comprende la afirmación con que empezábamos este artículo, según la cual cualquier diferencia puede detectarse como estadísticamente significativa si el número de sujetos es suficientemente elevado[12,13]. Tan sólo la capacidad de re-clutamiento (generalmente ampliable mediante estudios multicéntricos de gran envergadura) pone límite a este hecho. Hay que añadir que las estimaciones de! tamaño de la muestra se refieren a cada grupo de estudio (control e intervención o expuesto). En el caso de análisis estratificados, el tamaño estimado se refiere siempre al estrato más pequeño considerado. A veces, sin embargo, se pondera uno de los grupos (generalmente, el control o no expuesto a un factor o intervención) con el fin de precisar menos sujetos expuestos a la intervención o que presenten el acontecimiento de interés, dependiendo del tipo de diseño. Para ello se utiliza un factor de ponderación «λ» (véase apéndice).

Tamaño fijo o tamaño variable

Las muestras de tamaño fijo se refieren a la situación más común en la investigación clinico-epidemiológica, en la que se requiere la estimación del tamaño muestral previa a la ejecución del estudio. En otras situaciones, el tamaño de la muestra se irá incrementando hasta obtener un resultado

predeterminado (diseño secuencial), o se realizará un diseño experimental de un solo caso. Estas dos últimas se deno-minan muestras de tamaño variable. A continuación, se revisan las ecuaciones utilizadas para cada tipo de variable (continua o categórica) en los diseños de investigación más comunes, y se discuten las limitaciones de cada una de ellas y las alternativas más adecuadas en cada caso. Se insiste en las muestras de tamaño fijo y al final se presentan brevemente los diseños alternativos con muestras de tamaño variable.

Muestras de tamaño fijo

Estimación de parámetros poblacionales. Concierne esencialmente a los estudios transversales en los que, a partir de una muestra, se pretende estimar el valor poblacional de una variable continua (medias) o categórica (proporciones). Se trata de establecer *a priori* la precisión (amplitud del intervalo de confianza [IC]) deseada para dicha estimación. Es necesario disponer de una estimación, aunque sea imprecisa, del valor que se espera hallar en las proporciones y de la variancia en el caso de las medias. Esta estimación puede obtenerse de referencias bibliográficas o realizando un estudio piloto. En el caso particular de las proporciones, si no se dispone de una estimación previa, el investigador puede tener presente que la proporción de 0,5 (50%) es el valor que requiere mayor tamaño muestra.

1. Proporciones. Se utiliza la fórmula correspondiente a la aproximación a la curva normal para el cálculo del IC y se despeja n (véase la fórmula 1 del apéndice).
2. Medias. Se utiliza la fórmula del cálculo del IC para me-días que corresponde a la aproximación normal (véase la fórmula 2 del apéndice).
Las fórmulas más sencillas suponen que las poblaciones son infinitas. Si, por el contrario, el tamaño poblacional es limitado (menor de un millón de elementos, por convención) puede Introducirse la corrección por finitud de la muestra (véase la fórmula 3 del apéndice).

Comparación de medias. Nótese, ante todo, que lo que in-fluye en la estimación no son los valores de las medias de cada grupo, sino su diferencia. Distinguimos varias situaciones correspondientes a diseños distintos:

1. Comparación con una media poblacional de referencia. El investigador desea saber si la media observada en una muestra difiere de una media poblacional conocida. El algoritmo para el cálculo se encuentra en la fórmula 4 del apéndice[7].
2. Comparación de dos medias apareadas (medidas repetidas) en un solo grupo. Interesa comparar el cambio medio entre una medida basal y otra posterior, contrastando la H_0. de que la medida de este cambio es igual a cero en la población. Corresponde al diseño cruzado.
3. *Comparación de dos medias apareadas (medidas repeti-das) en 2 grupos.* El cambio entre una medida basal y otra posterior se compara entre 2 grupos distintos de sujetos. Dado que puede existir un grado sustancial de correlación entre la magnitud de la medida basal y la posterior, es ne-cesario corregir la fórmula por el coeficiente de correlación, que deberá estimarse, al igual que la DE, a partir de expe-riencias previas (fórmula 5 del apéndice)[14].
4. Comparación de dos medias independientes. Para la comparación de dos medias en muestras independientes se asume que la variancia es la misma en ambas poblaciones y ésta es la que se utiliza para la estimación. Los cálculos se basan en la distribución normal (véase la fórmula 6 del apéndice)[7]. Como puede verse en el apéndice, la fórmula 5 es semejante a la anterior y puede deducirse fácilmente del razonamiento realizado sobre la figura 1. 5. *Análisis de la variancia.* A pesar de que se pueden utilizar tablas que simplifican la tarea[15], se han propuesto métodos de cálculo específicos[16] basados en la distribución F de Snedecor. Sin embargo, una aproximación más sencilla que proponemos consiste en utilizar el cálculo para la comparación de las medias de 2 grupos corrigiendo el riesgo a para el número de pares de comparaciones posibles y tomando el resultado obtenido como el tamaño de cada grupo del estudio. La dife-rencia a detectar como estadísticamente significativa será la mínima que se considere clínica o epidemiológicamente rele-vante entre al menos un par de grupos estudiados. Este cálculo proporciona resultados muy parecidos a los obtenidos en tablas específicas y mediante algoritmos mucho más complejos.

Comparación de proporciones. Se pueden emplear diferen-tes aproximaciones en el caso de tratarse de datos aparea-dos, en la comparación de una proporción observada con una poblacional de referencia, en el de un diseño cruzado con medidas repetidas o en el de muestras independientes. En todas estas situaciones se precisa una estimación previa de las proporciones que se espera hallar en ambos grupos. Dado que el error estándar (EE) de una proporción P se cal-cula del siguiente modo:

$$EE(P) = \sqrt{\frac{P\_(1 - P)}{n}}$$

donde P es la proporción poblacional que se sustituirá por la proporción observada en la muestra (la mejor estimación) y n se refiere al tamaño de muestra, el valor de la proporción que maximiza el EE es 0,5 (50%). Por lo tanto, a igualdad de dife-rencias a detectar, mayor será la muestra requerida cuanto más se acerquen las proporciones a dicho valor. Además, en los contrastes unilaterales la dirección de la comparación también influirá: es distinto alejarse que acercarse al 50%, siendo mayor e! número de sujetos en el último caso.
1. Comparación de una proporción observada con una pro-porción poblacional de referencia. Se utiliza la fórmula ba-sada en la aproximación normal a la distribución binomial (véase la fórmula 7 del apéndice).
2. Comparación de dos proporciones independientes. Se pueden utilizar varias aproximaciones:
- *Aproximación normal a la distribución binomial. Se* puede utilizar este método cuando las proporciones se encuentran entre el 20 y el 80% y tanto (n x p) como [n x (1 - n)] son superiores a 5, ya que la distribución de las muestras tiende a la normal (véase la fórmula 8 del apéndice)[7].
- *Aproximación del arco-seno.* Proporciona estimaciones precisas basadas en el cálculo de la probabilidad exacta de Fisher en proporciones pequeñas (de forma aproximada, por debajo del 20% o por encima del 80%). En éstas, el muestreo no sigue una distribución normal y la DE depende de la proporción[17]. Mediante dicha transformación se eluden estos inconvenientes (véase la fórmula 9 del apéndice).
- *Aproximación de Poisson.* En proporciones por debajo del 5% o por encima del 95% se prefiere la utilización de la dis-tribución de Poisson. Se basa en el proceso de Poisson para fenómenos de baja frecuencia (véase la fórmula 10 del apéndice)[7].
- *Hipótesis de bioequivalencia entre dos proporciones.* Co-rresponde al caso particular en el que interesa demostrar que la eficacia de una intervención nueva no difiere de la de otra ya existente -por ejemplo, por qué la disponible es más cara o produce más efectos secundarios-. Se trata de establecer cuál sería la diferencia que, desde el punto de vista

clínico o epidemiológico, resultaría relevante, y realizar la estimación con dicha diferencia. Oíros autores menos conservadores han propuesto métodos de estimación alternati-vos para este tipo de hipótesis que requieren, menos sujetos por grupo[10,18] (véase en el apéndice la fórmula 11).

3. *Proporciones apareadas (repetidas en un grupo).* Corres-ponde a las investigaciones que observan la proporción de la presencia de un acontecimiento antes y después de una intervención o exposición en un mismo grupo de individuos. Los cálculos se basan en la prueba de McMemar (véase la fórmula 12 del apéndice).

Contrastes de hipótesis con odds ratio *(OR).* Una *odds* no es otra cosa que la razón entre la probabilidad «p» de un acontecimiento y la probabilidad contraria (1 - p): por con-siguiente, la razón de *odds* indica cuántas veces resulta más probable el acontecimiento en un grupo de expuestos a un factor con respecto a otro grupo de no expuestos. Se trata de una medida de riesgo empleada en los diseños de casos y controles o transversales.
Estos tipos de diseño requieren que el investigador decida *a priori* la OR que se desea detectar como estadísticamente significativa, y proporcione una estimación de la proporción de expuestos en el grupo control. La estimación en el grupo de sujetos con el acontecimiento de interés entre los casos se realiza a partir de este dato y de la propia OR (véase la fórmula 13 del apéndice).
La estimación se basa en la aproximación normal a la distri-bución binomial propuesta por Fleiss[19] para contrastes bila-terales, totalmente equivalente a la comparación de propor-ciones descrita anteriormente mediante la aproximación normal. Otros autores han propuesto aproximaciones alter-nativas que proporcionan estimaciones del mismo orden[11].

Contrastes de hipótesis con riesgos relativos (RR). El RR es el cociente entre la proporción de acontecimientos en los expuestos al factor (o intervención) bajo estudio y la propor-ción en los no expuestos. Se trata de la medida de riesgo empleada en los diseños longitudinales de cohortes y en al-gunos ensayos clínicos.
El cálculo del tamaño muestra! para estos tipos de diseño re-quiere una estimación del RR mínimo que se desea detectar como estadísticamente significativo y de la proporción de su-jetos que presentan el acontecimiento de interés en el grupo control no expuesto al factor o intervención estudiados: la es-timación en el grupo de sujetos con el acontecimiento de in-terés se realiza a partir de este dato y del RR (véase la fór-mula 14 del apéndice).
La estimación se basa, como en el caso anterior, en la aproximación normal a la distribución bi-nomial propuesta por Fleiss[19] para contrastes bilaterales. Debe destacarse que valores iguales de OR y de RR requie-ren tamaños de muestra distintos debido a las diferentes propiedades de ambos estimadores de riesgo.

Comparación de curvas de supervivencia. Otra forma fre-cuentemente utilizada para estudiar el efecto de un factor o una intervención sobre la aparición de un acontecimiento de interés en diseños longitudinales de cohortes es la compara-ción de curvas de supervivencia estimadas mediante el método de Ka-plan-Meier: el contraste de hipótesis generalmente se lleva a cabo mediante el estadígrafo de Mantel-Cox. A partir de ello, y entre otros[20], se ha propuesto un método[21] que precisa una esti-mación del RR (o *hazard ratio)* del grupo expuesto respecto al control para los cálculos. También es necesario estimar el tiem-po que durará el reclutamiento, el tiempo que durará el segui-miento tras finalizar el reclutamiento y la tasa de supervivencia en el grupo control al ini-cio, a la mitad y al final del seguimiento. Si no se prevé seguimiento al final del período de reclutamiento se considerará una supervivencia inicial del 100% y las supervivencias medias estimadas a la mitad y al final del reclutamiento, res-pectivamente (véase la fórmula 15 del apéndice).

Poder estadístico de un contraste de hipótesis. En los con-trastes de hipótesis tanto entre dos proporciones como en-tre dos medias, es posible calcular a posteriori la potencia o poder estadístico de dicha prueba conociendo el tamaño muestra l de cada uno de los grupos comparados. Es útil para evaluar la calidad del diseño de un estudio publicado, o de una investigación ya concluida en los que no se esta-bleció el tamaño de la muestra en la fase de diseño. El cálculo *a posteriori de* la potencia de un contraste es reco-mendable, sobre todo, ante la sospecha de que la probabili-dad de rechazar la hipótesis alternativa, siendo ésta en reali-dad cierta (error de tipo II), es elevada. En este caso, se utiliza corno diferencia a detectar como estadísticamente sig-nificativa aquella que se considere clínicamente relevante, independientemente de la que se haya hallado en el estudio en cuestión[57] (véanse las fórmulas 16 a 20 del apéndice).

Muestras de tamaño variable

Cuando en ensayos clínicos existe la posibilidad de obtener respuestas rápidas (el tiempo necesario para la evaluación del acontecimiento de interés es corto), puede utilizarse un diseño secuencial o un diseño experimental con un solo caso.
El diseño secuencial. En este caso, el ensayo sólo puede in-cluir 2 intervenciones probadas secuencialmente en cada participante (análisis con datos apareados) y su resultado debe poder evaluarse inmediatamente después de realiza-do. Dicha evaluación se efectúa en términos de preferencia por el efecto de una u otra intervención. Con este tipo de diseño se pretende minimizar el número de pacientes a incluir en el estudio, ya que la decisión de continuar reclutando a individuos se basa en el resultado obtenido de los que ya han entrado.
Tal como fueron descritos inicialmente, pueden ser abiertos, si el número de pacientes que pueden entrar en el estudio no tiene un límite preestablecido, o cerrados en caso contrario. El número límite de individuos necesario se calcula considerando los riesgos α y β aceptados, y la diferencia de proporciones que se desea detectar como estadísticamente significativa (hipótesis alternativa), frente a la hipótesis nula según la cual la proporción de preferencias por una u otra intervención es de 0,5 (50%). Se establece un límite superior para la proporción de respuestas por encima del que se rechaza la hipótesis nula a favor de una intervención, y uno inferior por debajo del que se rechaza a favor de la intervención alternativa. Mientras la proporción se mantenga entre ambos límites, se considera que se encuentra en la zona de indecisión; cuando se llega a alcanzar un cierto número de parejas de experiencias sin cruzar ninguno de los límites, esto obliga a aceptar la hipótesis nula. Los límites superior e inferior se determinan en fun-ción de los niveles de riesgo α y β previstos, según diferentes métodos descritos inicialmente por Armitage[22] (fig. 2). Las limitaciones de este tipo de diseño se centran en sus propias condiciones de aplicación. También es conocido el fenómeno según el cual, al realizar múltiples contrastes de hipótesis, la probabilidad «p» final de que alguna alcance el nivel de significación estadística es, aproximadamente, el producto de todos los niveles aceptados en las pruebas individuales. Esto obliga a realizar las oportunas correcciones. Los métodos secuenciales clásicos propuestos por McPherson y Armitage[22] no han sido muy utilizados a causa de estas limitaciones.

Diseño experimental de un solo caso. Sería un caso particular del diseño secuencial en el que a un mismo individuo se le administran 2 intervenciones sucesivamente en diversas ocasiones evaluando el efecto a cada pareja de resultados. Este tipo de diseño ha sido utilizado en ensayos en enfermedades psiquiátricas[23] aunque más recientemente se ha propuesto un uso más generalizado en ensayos clínicos terapéuticos de otros ámbitos[24-26].

Nivel de significación estadística en los análisis intermedios

Los análisis intermedios tienen implicaciones en el nivel de significación, ya que aumentan la probabilidad (a causa de la propia repetición de los contrastes de hipótesis) de un hallazgo fortuito. En otras palabras, el riesgo a es mayor que el previsto para un solo contraste al final del estudio. En los contrastes parciales debe corregirse, por lo tanto, el riesgo a global aceptado. Estos análisis se utilizan con frecuencia en los ensayos clínicos como medida de seguridad ante la eventualidad de que el tratamiento evaluado presente un beneficio inesperadamente mejor o peor que el control. La decisión de interrumpir el ensayo debe estar basada en criterios estadísticos claramente definidos. Se trata, en esencia, de «distribuir» el riesgo de primera especie (α) entre todas las evaluaciones parciales, para compensar el aumento de la probabilidad de hallar una asociación significativa espuria debido a la realización de múltiples contrastes de hipótesis. Estos métodos se aplican a los estudios con muestras de tamaño fijo en los que se ha planeado efectuar un número determinado (generalmente pequeño, entre 2 y 5) de evaluaciones con los datos acumulados. Existe un cierto acuerdo en admitir que es mejor establecer el momento de realizar las evaluaciones según la proporción de sujetos re-clutados respecto del planeado. Por ejemplo, al 20, *40,* 60 y 80% del reclutamiento además del análisis final. Se han des-crito varios métodos para realizar la corrección del riesgo α:

1. Reducción estocástica. Si en un tiempo t del estudio faltan A individuos por evaluar, podemos estimar si se modificaría la conclusión del contraste de hipótesis si todos ellos presentasen el efecto opuesto al que se evalúa (evolución conservadora del ensayo). En caso negativo, si la probabilidad de error a es suficientemente pequeña (pongamos menor que el 5%) está justificado recomendar la interrupción del ensayo clínico. En un ensayo en el que se evaluaba el efecto de la flecainida y de la encainida frente a placebo sobre las arritmias ventriculares después de un infarto de miocardio, se interrumpió antes de alcanzarse el tiempo de seguimiento y el reclutamiento previstos al comprobarse, mediante este método, que ambos fármacos se asociaban a mayor mortalidad que el propio placebo[27].

2. Métodos por secuencias agrupadas. Existen varios procedimientos para establecer criterios de significación estadística para evaluar el resultado obtenido en uno o más análisis intermedios realizados durante el desarrollo de un ensayo clínico.

a) Distribución uniforme. Por su simplicidad es uno de los métodos más utilizados para decidir cómo se distribuye el nivel de significación estadística entre las evaluaciones, y es el descrito por Armitage[22]. Cuando se realizan C comparaciones independientes de una misma variable o grupo de variables, el riesgo de primera especie al final de todas las comparaciones (α_{ef}) es realmente

$$\alpha = 1 - (1 - \alpha_i)^C$$

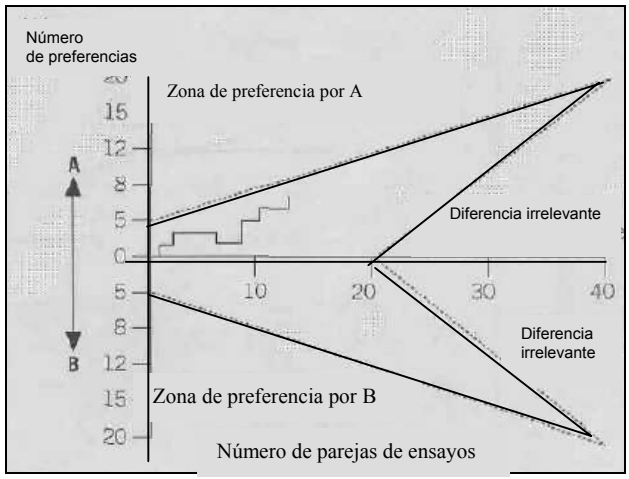

Fig. 2. Limites para la toma de decisiones en un diseño secuencial en que se compara el resultada de dos tratamientos o intervenciones.

si el riesgo α en cada i contraste es constante. Por ejemplo, si realizamos 10 contrastes de hipótesis con α = 0,05:

$$\alpha_{ef} = 1 - (1 - 0{,}05)^{10} = 1 - (0{,}95)^{10} = 1\text{-}0{,}60 = 0{,}40$$

Para calcular la α necesaria para obtener una α_{ef} de 0,05, utilizaremos la siguiente expresión:

$$\alpha_i = 1\text{-}(1\text{-}\alpha_{ef})^{1/C}$$

En la práctica, puede utilizarse la siguiente aproximación:
Por ejemplo, si realizamos 10 contrastes de hipótesis y queremos obtener un riesgo α_{ef} = 0,05, el riesgo aceptado en cada contraste (distribución del riesgo α) será:

$$\alpha_i = 1 - (1 - 0{,}05)^{1/10} = 0{,}005$$

o lo que es lo mismo,

$$\alpha_i = 0{,}05/10 = 0{,}005$$

b) Otras formas de distribuir el riesgo α entre !as diferentes evaluaciones. En la figura 3 se presenta un ejemplo de cuatro análisis intermedios secuenciales (que se denominan *agrupaciones),* según las aproximaciones de Pocock[38], Peto[29]. y O'Brien y Fleming[30], que son las más habitualmente utiliza-das[31], además de las descritas. Pueden obtenerse mayores detalles sobre los cálculos en las referencias citadas.

Comentarios

Cálculos particulares del tamaño muestral

Se han revisado los métodos utilizados para el cálculo del tamaño de la muestra en la mayoría de diseños empleados en los estudios biomédicos.
A lo largo de este artículo se ha asumido la normalidad de la distribución de las variables continuas. En los casos en los que la distribución de una variable se aparte mucho de la normal, pueden utilizarse fórmulas específicas[32,33] o los métodos paramétricos descritos incrementando el tamaño muestral en un 10%[34].

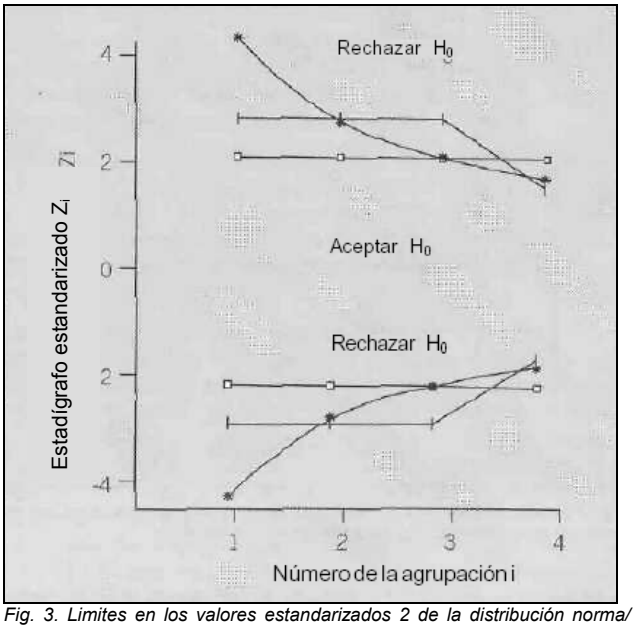

*Fig. 3. Limites en los valores estandarizados 2 de la distribución norma/ para decidir la hipótesis alternativa en cuatro análisis intermedios secuenciales agrupados según tres criterios distintos, Poc ┼┼ Haybiltle-Peto; ──── * ──── Q'Brien-Fleming.*

Los métodos de cálculo del tamaño de la muestra en la regresión logística requieren un comentario aparte. Si bien el tamaño de la muestra para una categoría de una variable cualitativa puede calcularse mediante el procedimiento des-crito para el contraste de hipótesis con OR, también existen tablas[35] y métodos de estimación para variables continuas[36]. Sin embargo, la complejidad de estos cálculos está más allá de las pretensiones del presente trabajo. Existen alternativas a la estimación del tamaño de la muestra centradas en el estudio de los IC de los estimadores[11,37]. Aunque algo más intuitivas, las bases del cálculo son exactamente las mismas que en los métodos presentados.

Recomendaciones para los investigadores

Conviene recordar que la utilización de variables continuas suele conducir a muestras menores que la de variables ca-tegóricas. El uso de contrastes unilaterales también reduce el tamaño de la muestra, aunque su empleo debe reservar-se a situaciones en las que realmente sea aplicable. Por otro lado, el uso de diseños con medidas repetidas (datos apareados) generalmente reduce a la cuarta parte el tamaño de la muestra total.

En los estudios de casos y controles, la finalidad de la ponderación del grupo no expuesto es reducir el tamaño de la muestra en el grupo de sujetos expuestos. Sin embargo, a partir de una ponderación de 3 a 1 el rendimiento, en términos de reducción de sujetos necesarios en el grupo expuesto, se atenúa sustancialmente. De hecho, pasar de 1 a 2 controles por caso implica una reducción de aproximadamente el 25% de expuestos y un incremento del 13% en la muestra total, y si se emplean 3 controles la reducción de expuestos llega al 33%, pero el incremento de la muestra total es del 35%, y al pasar a 4 controles estas cifras son del 37 y del 58%, respectivamente. Este ejemplo se basa en un estudio de casos y controles en el que el porcentaje de controles con el factor de interés fuera del 20% y la OR a detectar como estadísticamente significativa de dos. Los riesgos α y β serían del 5 y del 10%, respectivamente, para un contraste bilateral.

En la realización de un estudio, las normas éticas requieren que el diseno sea correcto y no exponga a seres humanos a riesgos injustificados. Un diseño correcto incluye la estimación realista de los pacientes necesarios para obtener el poder estadístico suficiente en los contrastes de hipótesis y la precisión adecuada en las estimaciones que se vayan a realizar. Considerando que el cálculo del tamaño muestral no es más que una estimación basada en supuestos, adquiere el máxi-mo sentido prever ¡a realización de análisis intermedios de seguridad que permitan interrumpir un ensayo clínico o cualquier otro tipo de estudio de seguimiento si se alcanzan los niveles de significación estadística antes de terminarlo. En cualquier caso, el nivel de significación no debería utilizarse como elemento único de decisión para la interrupción de un ensayo, puesto que tan sólo determina la probabilidad de que las diferencias obtenidas se deban al azar. Supóngase que el nivel de significación α obtenido en un contraste de hipótesis intermedio fuera del 7%, y el acontecimiento de interés fuera la mortalidad: la posibilidad de detener el estudio debido a esta diferencia debería discutirse por un comité de expertos. Dicho comité tendría en cuenta otros aspectos como el número de pacientes reclutados hasta la fecha del análisis, el número de acontecimientos de interés (fallecimientos en este caso) ocurridos y el resultado de oíros estudios concurrentes en el tiempo en su caso. También debería atender a las consecuencias de continuar el ensayo clínico en términos de la actitud de los investigadores a partir de aquel momento y del resto de la comunidad científica.

El software GRANMO *para el cálculo del tamaña de la muestra*

Considerando lo complicado de aplicar los algoritmos de cálculo manualmente, es útil disponer de un programa informático que permita a usuarios no especialistas realizar las estimaciones más habituales fiable y rápidamente. En este sentido, hemos desarrollado un *software* de libre difusión (shareware)[1] denominado GRANMO.

El programa GRANMO está dirigido a profesionales de la medicina, de la epidemiología y de otras ciencias biomédicas interesados en realizar sus propios cálculos de tamaño de la muestra. Es de fácil manejo y permite su uso en castellano, catalán e inglés.

Ciertamente, existen otros programas para realizar los cál-culos para la estimación del tamaño muestral. Algunos in-cluyen más técnicas estadísticas, pero su manejo requiere conocimientos especializados[36,38]. Otros cubren pocos tipos de diseño[39,40]. Excepto el EPIDAT[40], ninguno de ellos está disponible en castellano.

Agradecimiento

Los autores desean expresar su reconocimiento a Cristina Hernández, a Elena López de Briñas y a Joan Sancho de la Unidad de Informática Médica del IMIM por la programación de la versión para MSWindows del programa GRANMO.

Apéndice
Fórmulas para el cálculo del tamaño de la muestra en los diseños de tamaño fijo

Para facilitar la notación se designan los grupos de sujetos como controles (c) que reciben un tratamiento estándar o place-bo en los ensayos clínicos, o que no han estado expuestos a

[1] Las versiones para DOS *y* para Windows de este programa pueden obtenerse a través de las direcciones, postal, electrónica y la siguiente de Internet: HTTP://WWW.IMIM.ES/

factores o intervenciones en otros tipos de diseño, y como expuestos (e) a un factor o a una intervención experimental, Se denominará M_c a la media de referencia o del grupo control, y M_e a la de la hipótesis alternativa que se desea detectar como estadísticamente significativa, X al valor de la distribución normal correspondiente al límite que conduce a aceptar la hipótesis alternativa, Z_α al valor de la distribución normal correspondiente al valor del error α aceptado, Z_β al correspondiente al error β aceptado, y σ la desviación estándar (DE) de la distribución teórica (generalmente obtenida a partir de un estudio piloto, o de otro estudio). Excepto en casos particulares, las ecuaciones se presentan para un contraste unilateral: para un contraste bilateral debe utilizarse $Z_{\alpha/2}$ en lugar de Z_α. Se denominará n al número de individuos en cada grupo.

Todas las fórmulas aplicadas a contrastes de hipótesis en dos grupos pueden ponderarse de modo que el tamaño muestral (n) de los grupos sea desigual ($n_c + n_e$). En este caso el número de c es un múltiplo de los e. Esta ponderación *λ* se obtiene modificando el numerador en las fórmulas en que es aplicable.

$$\lambda = \frac{n_c}{n_e}$$

En los contrastes de hipótesis se requieren siempre los riesgos α y βaceptados por el investigador. En la estimación de parámetros poblacionales sólo es necesario el riesgo a que afecta al intervalo de confianza.

Estimación de parámetros poblacionales

Estimación de una proporción poblacional

El intervalo de confianza (IC) de una proporción es IC - P ± R, donde el recorrido R es:

$$R = Z_{\alpha/2} \times \sqrt{\frac{P \times Q}{n}}$$

donde P es la proporción estimada que esperamos hallar, Q es (1 - P) y $Z_{\alpha/2}$ es el valor de la distribución normal correspondiente al valor del error α aceptado en el cálculo del tamaño de la muestra. Despejando n:

$$n = \frac{Z^2_{\alpha/2} \times (P \times Q)}{R^2} \qquad (1)$$

Estimación de una media poblacional

De forma semejante a las proporciones el IC de una media M poblacional a partir de una muestra es IC = M ± R, donde el recorrido *R* es:

$$R = Z_{\alpha/2} - \sqrt{\frac{\sigma^2}{n}}$$

y donde σ es la desviación estándar, (DE), $Z_{\alpha/2}$ es el valor de la distribución normal correspondiente al valor del error α aceptado y n el tamaño de la muestra. Despejando n:

$$n = \frac{Z_{\alpha/2}{}^2 \times \sigma^2}{R^2} \qquad (2)$$

Si n es pequeña, podría revaluarse sustituyendo $Z_{\alpha/2}$ por $t_{\alpha/2}$ de la distribución de Student-Fisher.

A las fórmulas 1 y 2 puede aplicárseles la corrección por finitud de la población cuando ésta es inferior a un millón de elementos. Para ello basta multiplicarlas por el siguiente factor:

$$\frac{POB}{POB + TM} \qquad (3)$$

donde POB es el tamaño de la población en que se muestrea y TM el tamaño de la muestra obtenido en las fórmulas anteriores.

Contrastes de hipótesis con medias

Base para el cálculo del tamaño muestral

En la figura 1 puede verse una generalización del fundamento de la estimación del tamaño de la muestra en un contraste bilateral. En ella se aprecia el punto M_c correspondiente a la media del grupo de referencia, M_e correspondiente a !a media de un grupo con una intervención alternativa y el punto X representa el valor de la distribución normal correspondiente al riesgo α/2 aceptado en el contraste de hipótesis bilateral en la distribución de media M, y que define el riesgo β en la distribución hipotéticamente distinta de media M_e. La zona en negro es la del riesgo *α* y la sombreada la del riesgo β. Cuando $M_e > M_c$ la distancia

$$X - M_c = Z_{\alpha/2} \times \sqrt{\frac{\sigma^2}{n}}$$

y la distancia

$$M_e - X = Z_\beta \times \sqrt{\frac{\sigma^2}{n}}$$

donde $Z_{\alpha/2}$ es el valor de la distribución normal correspondiente al punto X en la distribución de referencia de media M_C y Z_β en la distribución de media M_e. Obsérvese que al tratarse siempre de un contraste unilateral en la Z_β correspondiente a la hipótesis alternativa, se toma la distancia (M_e - X) para poder tomar el valor Z_β positivo; de otro modo, cuando β es menor que 0,50, Z_β debería tomar valores negativos. Igualando para X obtenemos:

$$Z_{\alpha/2} \times \sqrt{\frac{\sigma^2}{n}} + M_c = - Z_\beta \times \sqrt{\frac{\sigma^2}{n}} + M_e$$

de dónde:

$$n_c = n_e = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{|M_e - M_c|^2} \qquad (4)$$

expresión totalmente equivalente a la anterior en un contraste unilateral.

Comparación de dos medias apareadas (medidas repetidas) en un solo grupo

El algoritmo de cálculo es igual que el anterior, con la salvedad de sustituir $| M_e - M_c |^2$ por d^2, siendo este valor «d», la media de las diferencias individuales entre los valores basales y posteriores.

Comparación de dos inedias apareadas (medidas repetidas) en 2 grupos distintos de sujetos

En este caso se precisa, además de los riesgo α y β una estimación de la DE (σ) de la medida (que se asumen iguales para ambos grupos, y tanto en la medida inicial como en la final), la diferencia entre los valores iniciales y finales en los controles M_{dc} y en los tratados M_{de} (y por lo tanto su diferencia), y del coeficiente de correlación «ρ» entre la medida basal y la final, conjunta para los 2 grupos:

$$n_c = n_e = \frac{2 \times (Z_\alpha + Z_\beta)^2 \times (1 - \rho) \times \sigma^2}{| M_{de} - M_{dc}|^2} \qquad (5)$$

Comparación de dos medias independientes

Sea M, la primera media poblacional y M_e la segunda, X el punto de decisión en una escala de diferencias entre medias, Z_α el valor de la distribución normal correspondiente al valor del error α aceptado y $Z_ß$ el correspondiente al error β aceptado; o la DE de ambas distribuciones (se asumen iguales para simplificar). La especificación del error a correspondiente a la hipótesis alternativa ($M_c - M_e$| 0) lleva a:

$$X - (M_c - M_e) = Z_\alpha \times \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}} = Z_\alpha \times \sqrt{\frac{2\sigma^2}{n}}$$

si $n_c = n_e$,

$$X = Z_{\alpha\_} \times \sqrt{\frac{2\sigma^2}{n}} + (M_c - M_e)$$

De forma similar, la especificación del riesgo β conduce a:

$$(M_c - M_e) - X = Z_\beta \times \sqrt{\frac{2\sigma^2}{n}}$$

de donde:

$$X = - Z_\beta \times \sqrt{\frac{2\sigma^2}{n}} + (M_c - M_e)$$

igualando para X y despejando n, cuando $n_c = n_e$:

$$n_c = n_e = \frac{2 \times (Z_\alpha + Z_\beta)^2 \times \sigma^2}{| M_c - M_e|^2} \qquad (6)$$

Análisis de la variancia simple: comparación de más de dos medias

Los cálculos propuestos se basan en una corrección para comparaciones múltiples. Para ello se utiliza un método parecido a la distribución uniforme descrita en el apartado de análisis intermedios (véase texto). Consiste en dividir el riesgo α global deseado (generalmente 0,05) por el número de pares de comparaciones posibles entre *g* grupos del estudio. Para 3 grupos serían 3 pares de comparaciones, para cuatro serían 6, etc. Los cálculos para la estimación del tamaño de la muestra se basan en la ecuación 6.

Contraste de hipótesis con proporciones

Utilizando un razonamiento semejante al de las medias sobre la figura 1 y aplicando la aproximación normal a la distribución binomial, se puede llegar a las siguientes ecuaciones:

Comparación de una proporción observada con una poblacional de referencia

La ecuación a que se llega es:

$$n_c = n_e = \frac{(Z_\alpha \sqrt{2PQ} + Z_\beta \sqrt{P_e Q_e})^2}{D^2} \qquad (7)$$

donde P es la proporción conocida de referencia; Q = (1 - P); P_e = proporción en el grupo expuesto; $Q_e = (1 - P_e)$; D = (P_e - P). y el resto de notación igual a la de anteriores apartados.

Comparación de dos proporciones independientes

Cuando se cumplen las condiciones de aplicación de la prueba de X^2 *para una* tabla dos por dos, puede utilizarse esta aproximación para la estimación del tamaño de la muestra en la comparación de proporciones independientes. Llamando P a la proporción medía de la proporción de acontecimientos de Interés del grupo control c y del grupo tratado t, Q a 1 - P, P_c a la proporción de acontecimientos de interés en el grupo control, Q_c a 1 - P_c, P_e a la proporción en el grupo expuesto, Q_e a 1 - P_e, D a la diferencia entre P_e y P_c , y utilizando el resto de notaciones igual que en apartados anteriores, de forma similar se puede llegar a la siguiente expresión:

$$n_c = n_e = \frac{(Z_\alpha \times \sqrt{2 \times P \times Q} + Z_\beta \times \sqrt{P_c \times Q_c + P_e \times Q_e})^2}{D^2} \qquad (8)$$

Una alternativa a la ecuación anterior es la aproximación si-nusal inversa al cálculo de la probabilidad exacta de Fisher. Esta aproximación está basada en el hecho de que cuando las proporciones P_c<0,5, el arcoseno (√p) ≈ √p, y el seno (p) ≈ p. El arcoseno se expresa en radianes. La transformación de P:

$$\varphi(P) = 2 \times ARCOSENO(\sqrt{P})$$

cuyo valor oscila entre 0 y p_i, posee la propiedad de que su DE

$$\frac{1}{\sqrt{n}}$$

es independiente de P. Esta transformación permite llegar a la expresión siguiente:

$$n_c = n_e = \frac{(Z_\alpha + Z_\beta)^2}{2 \times (ARCOSENO\sqrt{P_c} - ARCOSENO\sqrt{P_e})^2} \qquad (9)$$

Cuando las proporciones se aproximan al 0 o al 100% (digamos que son inferiores al 5% o superiores al 95%) debería utilizarse la fórmula basada en la distribución de Poisson:

$$n_c = n_e = \frac{(Z_\alpha + Z_\beta)^2 \times (P_c + P_e)}{| P_c - P_e|^2} \qquad (10)$$

Hipótesis de bioequivalencia: se desea demostrar la hipótesis nula en la comparación de dos proporciones

Para contrastes bilaterales ($Z_{\alpha/2}$), la siguiente expresión es menos conservadora que sus homologas anteriores:

$$n_c = n_e = \frac{(Z_{\alpha/2} + Z_\beta)^2 \times P_c \times Q_c}{(P_c - P_e)^2} \qquad (11)$$

donde P_c es la estimación de la proporción de acontecimientos de interés en el grupo control, Q es 1 - P_c , P_e la proporción en el grupo tratado, y utilizando el resto de notaciones igual que en apartados anteriores.

Tamaño muestral en la comparación de proporciones con datos apareados (medidas dicotómicas antes y después)

Se utiliza un cálculo basado en la prueba de McNernar. La fórmula utiliza la proporción de efectivos positivos (+) y negativos (-) observados en la situación inicial y en la situación final (antes y después de una intervención, por ejemplo) en un mismo grupo de individuos.

		Situación final (+)	(-)	
	(+)	p(+ +)	P(+ -)	p(I +)
Situación inicial	(-)	p(- +)	P(- -)	p(I -)
		p(f +)	P(f -)	1

donde, p(i) es la proporción de positivos (p[i+]) o negativos (p[i-]) esperados en la situación inicial; p(f) es la proporción de positivos (p[f+]) o negativos (p[f-]) en la situación final. Esta última es la fijada por el investigador como la proporción mínima en la situación final a detectar como estadísticamente significativa.
Para una casilla cualquiera, se calculan el valor máximo y *el* valor mínimo posibles y las probabilidades condicionales correspondientes. Por ejemplo:

p(++) máximo = valor mínimo (p(f+), p(i+))
p(++) mínimo = p(f+) + p(i+) - 1

A partir de estos valores se calcula ademas el cociente "s":

$$s = \frac{p_{(-+)}}{p_{(-+)} + p_{(+-)}} = \frac{p_{f+} - p_{(++)}}{(p_{(f+)} - p_{(++)}) + (p_{(i+)} - p_{(++)})}$$

Suponiendo p(f+) < p(i+), el valor de p(++) máximo = p(f+), por lo que el numerador de la ecuación anterior se anula y s | (p(++) máximo) = 0. Y s | (p[++] mínimo):

$$s \mid p_{(++)mín} = \frac{p_{(f+)} - p_{(++)mín}}{p_{(f+)} + p_{(i+)} - 2 \times p_{(++)máx}}$$

A continuación se calculan los tamaños de muestra necesarios para las dos probabilidades s|p(++)máx y s|p(++)mín.

$$n_{(s\text{-}máx)} = \frac{0.25 \times (Z_{(\alpha)} + Z_{(\beta)})}{(0.5 - 0)^2 \times (|p_{(f+)} + p_{(i+)} - 2 \times p(++)mín|)}$$

$$n_{(s\text{-}mín)} = \frac{0.25 \times (Z_{(\alpha)} + Z_{(\beta)})}{(0.5 - s_{(mín)})^2 \times (|p_{(f+)} + p_{(i+)} - 2 \times p_{(++)mín}|)}$$

El tamaño de muestra necesario será la media de las dos «n» calculadas:

$$n = \frac{n_{(s\text{-}máx)} + n_{(s\text{-}mín)}}{2} \qquad (12)$$

Contrastes de hipótesis con odds ratios *(OR)*
Se trata en realidad de la fórmula utilizada para la comparación de proporciones independientes en un contraste bilateral. La única característica destacable proviene en realidad de la forma de calcular la proporción en los casos (P_{ca}) a partir dé la OR y de la proporción en los controles (P_{co}):

$$P_{ca} = \frac{P_{co} \times \hat{OR}}{P_{co} \times \hat{OR} + (1 - P_{co})} \qquad (13)$$

donde OR con «^» es la OR estimada que se desea detectar como estadísticamente significativa.

Contraste de hipótesis con riesgos relativos (RR)

El caso de los estudios de cohorte es semejante al de los de casos y controles. Se utiliza la fórmula de la comparación de proporciones independientes en un contraste bilateral. La característica que la distingue del método empleado en los estudios de casos y controles reside en la forma de calcular la proporción en los expuestos (P_e) a partir del riesgo relativo y de la proporción en los controles (P_c):

$$P_e = \frac{\hat{RR}}{P_c} \qquad (14)$$

donde RR con «^» es el RR estimado de los expuestos al factor de interés respecto a los no expuestos que se desea detectar como estadísticamente significativo.

Comparación de curvas de supervivencia

Para la comparación de curvas de supervivencia se contrasta la diferencia entre tasas de mortalidad (o *hazard* rates) de los 2 grupos, esto es su RR (o *hazard ratio).* Las propiedades de la función de supervivencia permiten estimar el número de acontecimientos (d) necesarios que deben ocurrir para detectar como estadísticamente significativo un RR determinado por el investigador:

$$d = \frac{4 \times (Z_{\frac{\alpha}{2}} + Z_\beta)^2}{[\log(\hat{RR}]^2}$$

donde log(RR con "^") es el logaritmo natural del riesgo relativo estimado entre las tasas de mortalidad de los 2 grupos. Para la estimación del número de sujetos necesarios se precisa, además de RR, una estimación de la mortalidad en el grupo control (P_c) en tres puntos de la curva: al final del reclutamiento ($P_{c,r}$), a la mitad del tiempo entre el final del reclutamiento y el final del seguimiento ($P_{c,[r-f]/2}$). y al final del seguimiento ($P_{c,f}$). El número total de sujetos requeridos será:

$$n = \frac{d}{1 - \frac{1}{6} \times (P_{c,r} + 4 \times P_{c,\frac{r+f}{2}} + P_{c,r+f})} \qquad (15)$$

Poder estadístico de un contraste de hipótesis

Contraste de hipótesis entre dos proporciones con un tamaño muestral conocido

Aproximación de X^2. El poder estadístico (1 - β) de un contraste de hipótesis en el caso de medidas dicotómicas sería el área a la izquierda del punto Z_β de la distribución normal:

$$Z_\beta = \frac{Z_\alpha \times \sqrt{\frac{\overline{P}\,\overline{Q}}{n_c} + \frac{\overline{P}\,\overline{Q}}{n_e}} - |P_c - P_e|}{\sqrt{\frac{p_c Q_c}{n_c} - \frac{P_e Q_e}{n_e}}} \qquad (16)$$

Aproximación del seno inverso:

$$Z_\beta = Z_\alpha - \frac{2 \times |ARCOSENO\sqrt{P_c} - ARCOSENO\sqrt{P_e}|}{\sqrt{\frac{1}{n_c} + \frac{1}{n_e}}} \qquad (17)$$

Aproximación de Poisson:

$$Z_\beta = Z_\alpha - \frac{|P_c - P_e|}{\sqrt{\frac{P_c}{n_c} + \frac{P_e}{n_e}}} \qquad (18)$$

En los tres casos debe hallarse el valor de β correspondiente a la Z_β en la distribución normal.

Contraste de hipótesis entre dos medias con un tamaño muestral conocido

De forma semejante al apartado anterior, el poder estadístico (1 - β) en el caso de medidas continuas sería el área a la izquierda del punto Z_β de la distribución normal:

$$Z_\beta = Z_\alpha - \frac{|M_c - M_e|}{\sqrt{\frac{(n_e + n_c) \times \sigma^2}{(n_c \times n_e)}}} \qquad (19)$$

donde M_c y M_e. son las medias del grupo control y tratado, respectivamente. El resto de notación es semejante a la de los apartados anteriores. Si $n_c = n_e$, la expresión se reduce a:

$$Z_\beta = Z_\alpha - \frac{|M_c - M_e|}{\sqrt{\frac{2\sigma^2}{n_c}}} \qquad (20)$$

REFERENCIAS BIBLIOGRÁFICAS

1. GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. N Engl J Med 1993; 329: 673-682.
2. Freiman J, Chalmers TC, Smith H et al. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. N Engl J Med 1978; 299: 690-694.
3. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized clinical trials. J Am Med Assoc 1994; 272: 122-124.
4. Carné X, Moreno V, Porta Serra M, y Velilla E. El cálculo del número de pacientes necesarios en la planificación de un estudio clínico. Med Clin (Barc), 1989; 92: 72-77.
5. Porta M, Moreno V, Sanz F, Carné X, Velilla E. Una cuestión de poder. Med Clin (Barc) 1989; 92: 223-228.
6. Lemeshow S, Hosmer DW, Klar J, Lwanga SK. Adequacy of sample size in Health studies. New York: John Wiley & Sons, 1995.
7. Meinert CL. Clinical trials. Design, conduct and analysis. New York: Oxford University Press, 1986.
8. Bland JM, Altman DG. One and two sided tests of significance. Br Med J 1994; 309: 248
9. Pease KE. The alternative hypothesis: one-sided or two-sided? J Clin Epidemiol 1989; 42: 473-476.
10. Plasencia A, Porta Serra M. La calidad de la información clínica (y II): significación estadística. Med Clin (Barc) 1988; 89: 122-126.
11. Porta M, Plasencia A, Sanz F. La calidad de la información clínica (y III) ¿Estadísticamente significativo o clínicamente importante?. Med Clin (Barc) 1988; 90: 463-468.
12. Lui KJ, Cumberland WG. Sample size requirement for repeated measurements in continuous data. Statist Med 1992; 11: 633-633.
13. Bratcher TL, Moran MA, Zimmer WJ. Tables of sample sizes in the analysis of variance. J Quality Technology 1970; 2: 156-164.
14. Kastenbaum MA, Hoel DG, Bowman KO. Sample size requirements: one-way analysis of variance. Biometrika 1970; 54: 421-430.
15. Eisenhart C. Inverse sine transformation of proportions. In Eisenhart C, Hastay M W, Vallis W A (eds). Selected techniques of statistical analysis. New York: McGraw-Hill, 1947. pp 395-416.
16. Makuch R, Simon R. Sample size requirements for evaluating a conservative therapy. Cancer Treat Rep 1978; 62: 1037-40.
17. Pocock SJ. Clinical Trials: a practical approach. New York: John Wiley & sons, 1983.
18. Fleiss JL. Statistical methods for rates and proportions. New York: Wiley & sons, 1981.
19. Lemeshow S, Hosmer DW, Klar J. Sample size requirements for studies estimating odds ratios or relative risks. Statist Med 1988; 7: 759-764.
20. Cantor AB. Sample size calculations for the log rank test: a Gompertz model approach. J Clin Epidemiol 1992; 45: 1131-1136.
21. Collet D. Modelling survival data in medical research. London: Chapman Hall, 1994.
22. Armitage P, Berry G. Statistical methods in medical research. Oxford: Blackwell Scientific Publications, 1987.
23. Barlow DH, Hersen M. Single case experimental designs. Strategies for studying behavior change. 2nd ed. New York: Pergamon, 1984.
24. Senn S. Cross-over trials in clinical research. New York: John Wiley and sons, 1993.
25. Guyatt G, Sackett DL, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy - randomized trials in individual patients. N Engl J Med 1986; 314: 889-892.
26. Porta M. The search for more clinically meaningful research designs: Single-patient randomized clinical trials (editorial). J Gen Intern Med 1986; 1: 418-419.
27. The Cardiac Arrhythmia Suppression Trial (CAST) Investigators. Increased mortality due to encainide or flecainide in a randomized trial of arrhythmia suppression after myocardial infarction. N Engl J Med 1989; 321: 406 - 412.
28. Pocock SJ. When to stop a clinical trial. Br Med J 1992; 305: 235-240.
29. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I: Introduction and design. Br J Cancer 1977; 34: 585-612.
30. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979; 35: 549-556.
31. Task Force of the Working Group on Arrhythmias of the European Society of Cardiology. The early termination of clinical trial: causes, consequences, and control. Eur Heart J 1994; 15: 721-738.
32. Cantor AB. Power estimation for rank tests using censored data: conditional and unconditional. Contr Clin Trials 1991; 12: 462-473.
33. Campbell MJ, Julius SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and outcomes in two group comparisons. Br Med J 1995; 311: 1145-1148.
34. Al-Sunduqchi M. Determinig the appropriate sample size for inferences based on the Wilcoxon statistics. Laramie: University of Wyoming, Ph.D. dissertation, 1990.
35. Hsieh FY. Sample size tables for logistic regression. Stat Med 1989; 8: 795-802.
36. Hintze JL. SOLO power analysis. Los Angeles: BMDP Statistical Software, 1992.
37. Bristol DR. Samples sizes constructing confidence intervals and testing hypothesis. Statist Med 1989; 8: 803-811.
38. Elashoff JD. N query advisor. Los Angeles CA: Dixon Associates, 1995.
39. Dean AG, Dean JA, Burton AH, Dicker RC. Epi Info, version 5. A word processing, database, and statistics system for Epidemiology on microcomputers. Stone Mountain (Georgia): USD, Incorporated, 1990.
40. Alonso JM, Campillo C, Castillo C, Cotos T, Fernández E, Hervada X, Rodriguez L, Rodriguez MA, Sánchez C, Vázquez E. EPIDAT: análisis epidemiológico de datos tabulados. Santiago de Compostela: Xunta de Galicia, 1994.