

Estimación del tamaño muestral en estudios biomédicos por diferentes programas de computación

Gastón Duffau T.¹

Resumen

El cálculo del tamaño muestral en estudios biomédicos es parte indispensable de la investigación. El uso de programas computacionales facilita la tarea, pero presenta nuevos desafíos al ofrecer estimaciones variables. *Objetivo:* Presentar estimaciones del tamaño muestral empleando 6 programas, en estudios de medias, proporciones, cohortes y caso-control. *Resultados:* Se observaron diferencias a veces muy manifiestas en los cálculos realizados por los distintos programas. En el estudio de medias, EpiDat sugiere un tamaño que corresponde a la mitad de los otros. En proporciones, MedCalc calcula una muestra cuyo tamaño es el 58% del mayor propuesto. En cohortes y caso-control, con grupos de igual tamaño, la variación no es tan manifiesta, aunque no despreciable y depende de la prueba de hipótesis que se empleará. En caso-control con 4 controles por caso, la diferencia puede ser muy notoria entre Epilinfo y EpiDat pero Power incluye explicación en el sentido que depende de la prueba de hipótesis planeada usar. *Conclusión:* Importante ayuda se puede esperar de los programas que permiten estimar el tamaño de la muestra a investigar en diferentes situaciones, pero es indispensable disponer del manejo necesario de los conceptos bioestadísticos que subyacen a tales estimaciones para mejor decidir. Además el investigador debiera elegir y estar familiarizado con las características de un programa específico.

(**Palabras clave:** tamaño muestral, programas estadísticos, investigación biomédica.)

Estimation of sample size using different computer programmes for biomedical investigations

Estimation of sample size is important in biomedical research studies. The use of statistical software requires a thorough study of potential drawbacks of these tools, especially when different programmes may estimate different sample sizes when given the same basic data. Objective: To present sample size estimations with six different softwares in the study of means, proportions, cohorts and case-control designs. Results: Significant differences in sample size estimates were observed using the same basic data. When studying means EpiDat suggests half the sample size estimations performed by other softwares. In proportions Medcalc output is 58% of the larger estimation. In cohort or case-control studies with groups of the same size, estimations are not all equal though similar depending on hypothesis test applied. In a case-control study with 4 controls per case there is a striking difference between Epilinfo and EpiDat softwares, but Power provides an explanation establishing the need to consider the hypothesis test to be used. Conclusions: Valuable support is to be expected when estimating sample size in different scenarios by using appropriate softwares. However, the investigator should be aware of the biostatistical concepts underlying these estimations in order to take better decisions in computer assisted estimations. It is recommended to become familiar with one specific software.

(**Key words:** sample size, statistical software, biomedical research.)

1. Médico. Departamento de Pediatría, Campus Norte, Facultad de Medicina, Universidad de Chile.
Correspondencia: e-mail: xv-gd@ctcinternet.cl

El cálculo del tamaño muestral en estudios biomédicos es complejo y constituye solo una aproximación pero es indispensable efectuarlo por cuanto tiene implicancias muy importantes. Desde luego, hay un aspecto de factibilidad que no se puede soslayar. Es así como en no pocas oportunidades el investigador podría saber por anticipado que con los recursos disponibles o el número de pacientes que cree poder reunir en el plazo que cuenta para el estudio, no podrá detectar una diferencia esperada, si esta existe. Por otra parte, hay un ángulo ético que respetar: no es posible involucrar pacientes que acceden a participar en la investigación sin contar al menos con razonable seguridad que el estudio será terminado y que tendrá el poder suficiente. De las posibles varias medidas de resultado que se escogen para incluirlas en la investigación deberá emplearse la que requiera el tamaño muestral mayor. En ocasiones, en modo casi inadvertido, la investigación propone, entre otras, la hipótesis de igualdad de dos terapias. En tal caso con toda seguridad esta es la que requerirá un tamaño muestral más elevado para ser verificada.

Muy importante será también establecer qué clase de datos serán obtenidos y la forma como serán analizados, puesto que el tamaño de la muestra depende también del método de análisis a emplear.

El uso de programas computacionales facilita la tarea, pero presenta nuevos desafíos al ofrecer estimaciones variables.

Elementos en la estimación del tamaño muestral: 1. El error alfa, que indica la probabilidad que los hallazgos de interés pudieran ser justificados por variaciones explicables por el azar. 2. El error beta, que corresponde a la probabilidad de no detectar un hallazgo como importante y atribuirlo al azar. 3. Diferencia clínicamente significativa. Para cualquier estadígrafo, al realizar una investigación, se deberá definir el valor que el clínico considere importante: diferencia, sea de proporciones o promedios, así como razón de ventaja o riesgo relativo comparados con la unidad, índice kappa o coeficiente de correlación comparados con cero, etc. 4. Hipótesis uni o bilateral. En los estudios clínicos corresponde casi siempre plantear una hipótesis bilateral. 5. Posible pérdida de casos dentro del estudio. Es importante tener presente la necesidad de evitar hasta donde sea posible la pérdida de casos pues-

to que si esta es muy importante el estudio puede quedar invalidado. Como veremos, algunos programas calculan el tamaño muestral incrementando el resultado en un porcentaje arbitrario con el fin de compensar en parte una eventual pérdida de casos. 6. Método de análisis de los resultados. Este es otro importante factor a considerar, como será visto más adelante.

Se deberá tener presente que las investigaciones basadas en muestras insuficientes pueden conducir a: 1. Disminuir la probabilidad de detectar una diferencia supuesta, si existe. 2. Reducir la efectividad de la aleatorización de los integrantes de la muestra. 3. Alejar la posibilidad de análisis estratificado.

Por otra parte, también se debe tomar en cuenta lo siguiente: Cuando un estudio con muestra pequeña arroja resultados "negativos" se suele explicar ello por falta de poder y se considera, como ya señalamos, que se actuó mal por no haber hecho una estimación apropiada del tamaño muestral. Sin embargo, si en ese mismo estudio los resultados son "significativos" frecuentemente se olvida que la falta de poder igualmente significó involucrar recursos humanos, materiales y riesgos de diversa magnitud en los participantes que otorgaron su consentimiento informado en una empresa con baja probabilidad de "éxito"¹⁻⁵.

El objetivo es presentar estimaciones de tamaño muestral y su variación, empleando diferentes programas computacionales en estudios de medias, proporciones, cohortes y caso-control.

MATERIAL Y MÉTODO

Se dispuso de 6 programas estadísticos de empleo relativamente común, que incluyen en sus opciones el cálculo del tamaño muestral. Estos no cubren habitualmente todos los problemas de tamaño de muestra más frecuentemente planteados y fueron los siguientes: EPIINFO 6.04a y b⁶, EPIDAT 2.0⁷, WIN EPISCOPE 1.0⁸, MEDCALC⁹, POWER¹⁰, PS(Power-Sample)¹¹.

A todos ellos se les plantearon problemas de cálculo del tamaño de la muestra en comparación de proporciones, medias, estudio de cohorte, caso-control y se observaron las diferencias generadas ante un mismo problema.

RESULTADOS

En la tabla 1 se observa que hay diferencias a veces muy manifiestas en los cálculos realizados por distintos programas. En el estudio de medias, EpiDat 2.0 sugiere un tamaño que corresponde a la mitad de los otros programas. Esto corresponde directamente a un error de programación no detectado oportunamente. Realmente la cifra que genera corresponde al número de casos por grupo pero está errónea. En proporciones, MedCalc versión 4.16, calcula una muestra cuyo tamaño es el 58% del mayor propuesto. No señala el procedimiento de cálculo empleado pero en una versión posterior (MedCalc 4.20 win 95) alcanza las cifras de los programas que utilizan el procedimiento de Fleiss. Se debe tomar en cuenta que Epi-Info 6.04 incrementa en 10% el tamaño muestral inicialmente estimado por el programa, para dar cuenta de una cierta pérdida de casos. En cohortes y caso-control, con grupos de igual tamaño, la variación no es tan manifiesta, aunque no despreciable y depende de la prueba de hipótesis que se

empleará. EpiDat 2.0 en el estudio de cohortes (que no se muestra en la tabla), para detectar un riesgo relativo preestablecido, no ofrece la opción de grupos de tamaño diferente (razón expuestos/no expuesto solo 1) y por otra parte, no opera como le interesa al clínico, a saber, cuántos casos debe investigar para detectar un RR determinado o mayor. El programa genera tamaño muestral para detección de un cierto RR más menor cierto error. En caso-control con 4 controles por caso, la diferencia puede ser muy notoria entre los subprogramas StatCalc y Epitable de EpiInfo y entre estos y EpiDat. Power muestra tamaños diferentes según la prueba de hipótesis planeada usar, refiriéndose al empleo de Ji cuadrado sin corrección de Yates o a la misma prueba con la corrección, así como Fisher-Irwin. PS, en este punto, opera de modo similar aunque no indica contemplar la opción de Ji cuadrado corregido.

No hemos mencionado el programa Epi-Info 2000 porque la rutina StatCalc que incluye no es nueva, de hecho es la versión de 1993 y, por otro lado, no tiene Epitable.

Tabla 1

Muestras independientes. Dos colas. N por grupo.
Errores: α : 5%, β : 20%

| | EpiInfo 6.04 | EpiDat 2.0 | Win Episcope 1.0 | Medcalc 4.16e | Power | PS 1.0.12 |
|------------------------------|--------------|------------|---------------------|------------------|-----------|-----------|
| Medias (M) | — | 31 | 65 | 63 | 63 | 63 |
| M1 = 50, M2 = 40, DE = 20 | | | | | | |
| Proporciones (p) | 176 | 163 | 162 | 103 | 162* | 162* |
| p1 = 0,30, p2 = 0,45 | | | | | 175** | 175** |
| Cohorte | 159 (S) | — | — | — | 140* | 140* |
| Exp/no exp = 1 | 160 (E) | | | | 159** | 159** |
| frec. en no exp. = 5% | | | | | | |
| RR = 3,0 | | | | | | |
| Caso-control | 199 (S) | 177 | 177 | — | 177* | 177* |
| no pareado | 200 (E) | | | | 199** | 199** |
| Controles/caso = 1, OR = 3,0 | | | | | | |
| Exposic. en controles = 5% | | | | | | |
| Igual, excepto | 444/111 (S) | 388/97 | — | — | 388/97* | 388/97* |
| Controles/caso = 4 | 536/134 (E) | | | | 444/111** | 444/111** |

S = Statcalc

E = Epitable

* = Empleando Ji cuadrado sin corrección.

** = Empleando Ji cuadrado corregido o Fisher-Irwin según Power y Fisher-Irwin según PS.

CONCLUSIONES

1. Existen errores en los programas, como se ha podido observar con EpiDat en estudio de medias o EpilInfo 6.04 ("a" o "b") en los subprogramas StataCalc y Epitable.
2. No todos los programas muestran las fórmulas empleadas en las estimaciones.
3. Los programas no suelen abordar todos los diseños más comunes.
4. EpilInfo incrementa en 10% el tamaño muestral determinado, para compensar posibles pérdidas. Se debe aclarar que no se puede solucionar la pérdida de casos por el solo expediente de aumentar la muestra.
5. El investigador debiera informar siempre al lector acerca del procedimiento empleado para estimar la muestra en su estudio.
6. Entonces, aunque importante ayuda se puede esperar de los programas que permiten estimar el tamaño de la muestra a investigar en diferentes situaciones, resulta indispensable disponer del manejo necesario de los conceptos bioestadísticos que subyacen a tales estimaciones para mejor decidir. Además, el investigador debiera ele-

gir y estar familiarizado con las características de un programa específico y emplear este.

REFERENCIAS

1. Fleiss JL: Statistical methods for rates and proportions. 2nd edition, John Wiley & Sons, 1981: 38-48.
2. Snedecor GW, Cochran L: Statistical methods. 7th edition, The Iowa State University Press, 1986: 89-122.
3. Haseman JK: Exact sample size for use with the Fisher-Irwin test for 2 x 2 tables. *Biometrics* 1978; 34: 106-9.
4. Duffau G: Tamaño muestral en estudios biomédicos. *Rev Chil Pediatr* 1999; 70: 314-24.
5. Herrera P, Duffau G: Estudio Caso-Control. Editorial Mediterráneo, Santiago de Chile, 1997: 61-70.
6. C.D.C., U.S.A. EpilInfo 6.04a. W.H.O., Geneva, Switzerland, 1996.
7. Programa Análisis de Situación de Salud (HDA)-OPS/OMS. España. Epidat versión 2.0 para Windows, 1997. Análisis epidemiológico de datos tabulados. Xunta de Galicia, España.
8. Blas N, Ortega C, Franken K, Noordhuizen J: Win Episcope 1.0a. Veterinary Faculty, Zaragoza and Agricultural University, Wageningen, 1996.
9. Frank Schoojans. Medcalc statistical software, version 4.16a, Windows 95. U.S.A 1997.
10. Gustafson TL: Power. Texas USA, 1984.
11. Dupont WD, Plummer WD. PS: power and sample size calculations, U.S.A. 1997.

AVISO A LOS AUTORES

Los autores que deseen enviar sus trabajos en disquete deben hacerlo en disco flexible 3,5" de alta densidad, PC compatible, escrito en WordPerfect o MS Word, además de 2 copias impresas. Recordamos que el uso de disco permite agilizar el proceso editorial al que es sometida cada publicación.