

Análisis Crítico de la Literatura

Medwave, Año XII, No. 5, Junio 2012. Open Access, Creative Commons.

Rol del azar en los estudios clínicos

Autor: Miguel Araujo Alonso⁽¹⁾

Filiación: ⁽¹⁾Hospital de Los Andes, Chile; Medwave, Santiago, Chile

Correspondencia: Hospital de Los Andes, Chile; Medwave, Santiago, Chile

doi: <http://dx.doi.org/10.5867/medwave.2012.05.5423>

Ficha del Artículo

Citación: Araujo M. Rol del azar en los estudios clínicos. *Medwave* 2012 Jun;12(5) doi: 10.5867/medwave.2012.05.5423

Fecha de envío: 7/5/2012

Fecha de aceptación: 15/5/2012

Fecha de publicación: 1/6/2012

Origen: solicitado

Tipo de revisión: sin revisión por pares

Resumen

En este artículo el autor se refiere al error aleatorio y cómo este afecta también los resultados de los estudios clínicos y se ve reflejado en los intervalos de confianza que calculan los diferentes autores.

Asegura que para contrarrestar el error aleatorio existe un principio muy simple: calcular adecuadamente el tamaño de muestra al momento de diseñar el estudio. Para lograr esto, describe y profundiza en conceptos como "interferencia" y los factores que inciden en la posibilidad de que se produzca un error por azar en los resultados: "variabilidad", el "tamaño muestral" y la "magnitud de las diferencias". También delimita la "significación estadística" y el "valor P".

Abstract

This article refers to random error and how it may affect results of clinical trials, which is reflected in the confidence intervals that different authors calculate. In order to counteract random error a simple principle is applied – correct estimation of the sample size during the study design phase. In order to achieve this, the article describes and expounds on concepts such as interference and the factors that may lead to greater likelihood of random error occurrence in the results: variability, sample size and magnitude of differences. The article also mentions statistical significance and p value.

Abreviaturas:

H₀: Hipótesis Nula

H₁: Hipótesis de Trabajo

H_A: Hipótesis Alternativa

Error aleatorio (muestral)

Más de alguna vez habrá visto publicados en la prensa los resultados de una encuesta de opinión pública, en los que se señala que el error muestral es de un 3%. Eso implica que si, por ejemplo, el resultado en la muestra fue un 40% de aprobación, si se encuestara a la población total el verdadero nivel de aprobación posiblemente sea un valor entre 37% y 43%. Ese "margen" de error aleatorio afecta también los resultados de los estudios clínicos, y se ve reflejado en los intervalos de confianza que calculan los autores.

El potencial de error por azar de un estudio no afecta la "validez" de los resultados, pero sí compromete la precisión de los mismos. Cuando un resultado es muy impreciso (su margen de error muestral es muy amplio),

no podemos saber cuál es la verdadera magnitud de una asociación. E inclusive, puede que no sea posible discriminar si la asociación efectivamente existe.

Para contrarrestar el error aleatorio existe un principio muy simple: calcular adecuadamente el tamaño de muestra al momento de diseñar el estudio.

Por tanto, la primera premisa es que el tamaño muestral de un estudio no es un indicador de su validez interna, es decir, no nos dice qué tan sesgado puede estar un estudio.

Tal como hemos visto, no lo es en el sentido de predecir el potencial de sesgo o error sistemático del estudio.

Vamos a la evidencia empírica: cuando se han estudiado las discrepancias en los resultados entre estudios grandes con los de tamaño pequeño, se ha observado que las diferencias se explican no por su tamaño, sino por variaciones en características vinculadas al control de sesgos y confundentes (randomización adecuada, enmascaramiento, etc.)¹. Aún más, cuando se ha comparado directamente los resultados de estudios grandes y pequeños sobre la misma intervención, la regla general es la concordancia en los resultados².

Concepto de inferencia

El rol del azar en los estudios clínicos tiene que ver con el concepto de inferencia.

La inferencia implica poder generalizar los resultados obtenidos en una muestra de pacientes a un grupo mayor de individuos.

Cuando se obtiene la muestra para realizar el estudio, estamos sujetos a un cierto error de muestreo. Repetidas muestras obtenidas de la misma población serán probablemente distintas entre sí. Esto implica que los resultados en cada muestra no son necesariamente un fiel reflejo de lo que ocurre en la población de la que proviene, sino que tendrán un margen de error aleatorio.

¿Qué factores inciden sobre la posibilidad de que se produzca un error por azar en los resultados (del estudio que estamos analizando o en el que hemos decidido diseñar)? Fundamentalmente tres.

1. Variabilidad

La variabilidad es inherente al fenómeno que estamos midiendo.

Los fenómenos fisiológicos pueden ser más o menos estables en un horizonte dado de tiempo. Piense en la variabilidad temporal del siguiente grupo de parámetros y saque sus propias conclusiones:

- Presión arterial
- Peso corporal
- Sensación de dolor

Por azar –o por un conjunto de factores que no alcanza a entender o controlar de manera suficiente-, al momento de medir la sensación de dolor a través de un cuestionario pudiera encontrar al sujeto relativamente relajado, en un estado tal que su percepción subjetiva del nivel de dolor es favorable. O al revés, pudiera hallarlo en un instante de crisis, en la que el dolor se ha intensificado súbitamente. Ningún dolor es constante, plano.

Por el contrario, si el parámetro es el peso corporal, o más aún, la talla, la variabilidad de su medición probablemente sea mínima.

La variabilidad también se expresa en el comportamiento de la variable ya no en un mismo sujeto, sino *entre* sujetos. En ocasiones una exposición produce, de manera bastante constante y homogénea, el mismo efecto en

todas las personas. En otras palabras, el tipo o la magnitud de la respuesta pueden ser muy distintos.

¿Cómo se puede cuantificar la variabilidad de las mediciones realizadas en un estudio? A través del error estándar, la varianza, la desviación estándar, es decir, mediante los indicadores de variabilidad clásicos que nos aporta la bioestadística.

2. Tamaño muestral

A medida que el tamaño muestral del estudio crece, la posibilidad de que las inferencias respecto de la población total sean incorrectas disminuye.

3. Magnitud de las diferencias

Mientras mayor sea la diferencia –asociación- encontrada, menor la posibilidad de que ella se deba al azar.

Significación Estadística y Valor p

¿Cómo se evalúa el rol del azar en los estudios clínicos?

El azar en los estudios clínicos se evalúa a través del testeo de hipótesis, o aplicación de los test de significación estadística. Los test de significación estadística permiten cuantificar hasta qué punto la variabilidad de la muestra puede explicar los resultados del estudio. Esto implica establecer en forma explícita las hipótesis a contrastar.

La Hipótesis Nula (denominada H_0) representa la aserción de que no existe relación entre la exposición y el resultado, y la Hipótesis de Trabajo o Alternativa (denominada H_1 o H_A), representa la afirmación de que sí existe un vínculo de dependencia entre las dos variables.

Los test de significación estadística evalúan entonces el rol que puede haber jugado el azar como explicación de las diferencias observadas en las muestras. Existen diversos test de significación estadística (por ejemplo X cuadrado, t Test, entre otros), cuya selección depende del tipo de diseño utilizado y el tipo de variable estudiada. Tienen en común el ser una función de la diferencia entre los valores observados en el estudio y aquellos que se esperaría encontrar si la Hipótesis Nula fuera cierta, considerando la variabilidad de la muestra. Conceptualmente, pueden ser entendidos como una fracción en la que el numerador aumenta junto con la diferencia entre los valores observados y los esperados si la hipótesis nula fuera cierta, y un denominador que representa la variabilidad del resultado, el cual disminuye en la medida que el tamaño muestral aumenta.

¿Cómo se expresa finalmente el resultado del test?

El resultado del test se representa a través del valor p, mientras mayor el resultado del test el nivel del valor p es menor. El valor p se define entonces de la siguiente forma:

"Si la hipótesis nula fuera cierta, es la probabilidad de observar una diferencia igual más extrema que la encontrada". O, dicho de otra manera, "la probabilidad de obtener sólo por azar, un resultado igual o más extremo que el encontrado, asumiendo que no existe una verdadera asociación entre la exposición y resultado que está siendo estudiado".

Este concepto necesita un tiempo y razonamiento para asimilarlo. Algunas señales:

- Para entender el valor primero se debe asumir que no se sabe cuál es la verdad. No se sabe cuál de las dos hipótesis es la cierta.
- Luego se supone que H_0 es la hipótesis cierta. Es decir, la verdad es que no existe asociación.
- En vista de lo anterior, si los resultados del estudio muestran una diferencia (asociación entre las variables), ¿cuál es la probabilidad de haber encontrado esa asociación sólo por azar?

En términos más cualitativos, el valor p es una medida de la fuerza de la evidencia contra la hipótesis nula.

¿Es el valor P la probabilidad de que la hipótesis nula sea cierta?

No, la condición de verdad de la hipótesis no se establece a través de los test de significación estadística. Estos sólo permiten estimar el potencial de error por azar.

¿Podemos comparar dos o más estudios en función de su valor p y establecer que la diferencia (asociación) encontrada en uno es más "significativa" que la encontrada en otro?

No. El hecho que un estudio tenga un valor P de 0,05 y otro de 0,01 no indica que la segunda asociación sea de mayor magnitud que la primera.

El valor P depende críticamente del tamaño muestral. Las asociaciones de baja magnitud pueden arrojar un valor p muy pequeño (por ejemplo, 0,00001), si el tamaño muestral es grande. Esa misma asociación, en una muestra chica, puede dar lugar a un valor P mucho más grande (por ejemplo, 0,05).

¿Cuál es un valor p significativo? ¿Cuándo un estudio es estadísticamente significativo? Es decir, ¿cuándo razonablemente sus hallazgos no se explican por azar?

La significación estadística es un umbral convencional. Fisher estableció en 1950 que:

"Si p está entre 0,1 y 0,9, ciertamente no hay razón para sostener la hipótesis testeada. Si está bajo 0,02, ello indica fuertemente que la hipótesis no da cuenta de la totalidad de los hechos... A menudo no nos equivocaremos si establecemos una línea convencional en 0,05..."

Este autor también dijo que la interpretación es una facultad del investigado. Esto implica operar con un margen de error de 1 en 20 (0,05). Consideramos aceptable equivocarnos en uno de cada 20 casos. En ocasiones, el autor de un estudio puede considerar que su margen de error aceptable debe ser menor. Por esta razón, encontraremos con cierta frecuencia estudios en los que el nivel de significación estadística fue fijado en un valor más exigente, por ejemplo 0,01.

En su momento, la posibilidad o conveniencia de someter a interpretación el valor p obtenido en cada estudio generó controversia. Frente a la postura de Fisher, Norman y Pearson señalaron:

"Ningún test basado en una teoría de probabilidades puede por sí mismo proveer ninguna evidencia válida sobre la verdad o falsedad de una hipótesis..."

Según esta postura, se debía fijar un nivel de error aceptable previamente. Luego el resultado de nuestro análisis es simplemente rechazar o no H_0 , aquí no cabe la interpretación del valor p en cada estudio individual.

Como sea, diremos en general que los resultados de un estudio son estadísticamente significativos cuando su valor p es menor o igual que 0,05 y que no lo son cuando ese valor es mayor. Quizás aplicando inconscientemente el razonamiento de Fisher, o en un intento por "salvar" los resultados del estudio, cuando el valor p obtenido por el estudio no alcanza el valor de 0,05 pero se aproxima a él (por ejemplo, 0,06), los autores suelen referirse a ese hallazgo como "una tendencia" o como un resultado "sugerente pero no concluyente" de asociación³.

Referencias

1. Kjaergard LL, Villumsen J, Glud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med. 2001 Dec 4;135(11):982-9. [↑ | Link |](#)
2. Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC, Lau J. Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA. 1996 Oct 23-30;276(16):1332-8. Review. [↑ | Link |](#)
3. Laird AM, Miller PR, Kilgo PD, Meredith JW, Chang MC. Relationship of early hyperglycemia to mortality in trauma patients. J Trauma. 2004 May;56(5):1058-62. [↑ | Link |](#)



Esta obra de Medwave está bajo una licencia Creative Commons Atribución-NoComercial 3.0 Unported. Esta licencia permite el uso, distribución y reproducción del artículo en cualquier medio, siempre y cuando se otorgue el crédito correspondiente al autor del artículo y al medio en que se publica, en este caso, Medwave.