

Rol del azar en los estudios clínicos II

Random error in clinical trials (II)

Autor: Miguel Araujo Alonso⁽¹⁾

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

E-mail: maraujo@medwave.cl

doi: <http://dx.doi.org/10.5867/medwave.2012.06.5437>

Ficha del Artículo

Citación: Araujo M. Rol del azar en los estudios clínicos II. *Medwave* 2012 Jul;12(6):e5437 doi: 10.5867/medwave.2012.06.5437

Fecha de envío: 11/6/2012

Fecha de aceptación: 11/6/2012

Fecha de publicación: 1/7/2012

Origen: solicitado

Tipo de revisión: sin revisión por pares

Correspondencia a:

⁽¹⁾Medwave, Villaseca 21, Oficina 702, Ñuñoa, Santiago, Chile

Resumen

El azar dentro de los estudios clínicos puede inducir varios tipos de incorrecciones. Uno de ellos es el mostrar una posible asociación que en realidad no exista cuando se intenta asociar demasiadas variables (Error Tipo I). Otro tipo de fallos es la situación inversa, en que existiendo la relación el estudio no da cuenta de ella por no tener un tamaño muestral suficiente (Error Tipo II). Otro error frecuente se produce en la unidad de análisis en los ensayos aleatorizados por grupos, al analizar los datos de los sujetos individualmente como si fueran independientes.

Abstract

Random error in clinical trials can induce several types of flaws. One of them is to show a possible association that does not really exist when too many variables are linked (Type I error). Another type of failure is the opposite: the correlation does actually exist, but the study does not account for it because of insufficient sample size (Type II error). Another common mistake occurs in the unit of analysis in cluster randomized trials, when analyzing data from individual subjects as if they were independent.

Los errores vinculados al azar son dos:

Error tipo I

Ocurre cuando, pese a que en realidad no existe una determinada asociación, el estudio muestra que sí.

El nivel al cual un resultado se declara estadísticamente significativo es el límite de error tipo I aceptado, o α [alfa]. Como hemos dicho, habitualmente este nivel de error se fija en 5% (0,05).

¿Cuándo es más factible encontrar un error tipo I?

En términos prácticos, ello ocurre cuando se analiza un número grande de asociaciones simultáneamente. Por ejemplo, cuando se realizan múltiples análisis de subgrupos en un ensayo clínico o cuando en una cohorte se analizan por separado numerosas variables de

resultado para una misma exposición. También, cuando en estudios de casos y controles los autores exploran una gran cantidad de factores de riesgo posibles simultáneamente. Un ejemplo clásico de este tipo son los estudios de casos y controles que se realizan cuando se evalúa un brote de intoxicación alimentaria, cuando se aplican cuestionarios con una larga lista de alimentos potencialmente causantes del brote.

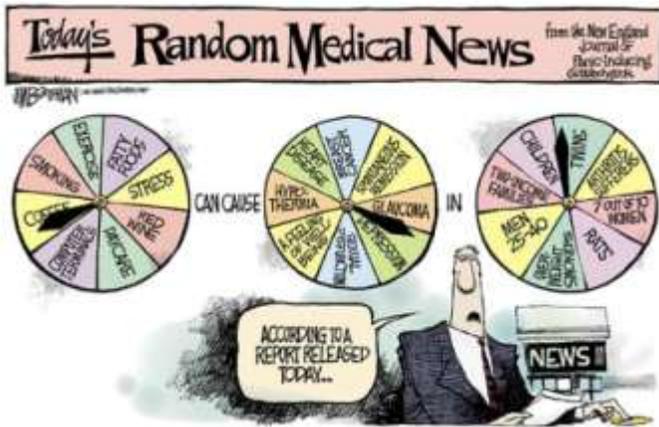
¿Qué ocurre en tales casos? Como nuestro nivel de error aceptable es 1 en 20, cuando intentamos vincular demasiadas variables podemos encontrar 1 o más asociaciones simplemente por azar. A este fenómeno se le ha denominado "castigo por husmear" (*penalty for seeking*).

En el ámbito de los factores de riesgo cardiovasculares, quizás los más explorados de todos, a menudo un estudio encuentra un nuevo factor de riesgo que posteriormente la evidencia acumulada no ratifica como tal. ¿Qué pasó

con el primer estudio en el cual el factor sí parecía asociado? Un error tipo I.

Para compensar el mayor riesgo de un error de este tipo en los estudios que incluyen comparaciones múltiples, se han descrito métodos de ajuste que reducen el umbral de p en función del número de variables estudiadas, como por ejemplo el método de Bonferroni. Sin embargo, su uso es discutible y no debe considerarse un procedimiento habitual en el análisis de resultados de los estudios clínicos¹.

Algunos han tomado esta situación con humor. He aquí un chiste aparecido en un periódico de Estados Unidos, en el que se hace alusión a esta avalancha de factores de riesgo que son detectados y difundidos día a día en la prensa:



Caricatura de Jim Borgman, publicada por primera vez por el Cincinnati Inquirer and King Features Syndicate 1997 Apr 27

Error Tipo II

Existe diferencia (asociación), pero el estudio no la demuestra.

Ocurre en la práctica cuando un estudio detecta la asociación, pero el resultado no alcanza significación estadística (el valor p es mayor de 0,05).

Al nivel de error tipo II se le denomina β [beta] y se encuentra directamente relacionado con el poder de un estudio.

Poder de un estudio

El poder del estudio es la probabilidad de encontrar la diferencia si esta realmente existe (rechazar H_0 cuando es falsa), y corresponde a $1 - \beta$. Normalmente, el poder del estudio se fija en al menos un 80% al determinar el tamaño muestral.

Suponga que diseña un ensayo clínico para evaluar la efectividad de la profilaxis antibiótica en cierto tipo de cirugía. Según su criterio clínico, le parece relevante poder detectar si el tratamiento es capaz de reducir la tasa de infecciones en al menos un 10% respecto de placebo. Si el tratamiento efectivamente reduce en más de un 10% la tasa de infecciones, su estudio -dado el tamaño muestral

que posee- tendrá un 80% de probabilidades de detectar esa diferencia y de que el resultado alcance significación estadística.

Lamentablemente, por desconocimiento de los investigadores o por dificultades prácticas para reclutar el número necesario de pacientes, un porcentaje importante de los estudios que se realizan no tiene poder adecuado^{2,3}.

¿El hecho de que el resultado de un estudio no alcance significación estadística, implica que la asociación evaluada no exista?

No. Las posibilidades incluyen:

- Que efectivamente no existe asociación, o bien
- Que el estudio simplemente no tuvo poder para demostrarla. La frase para ilustrar esto ya es conocida: "La ausencia de evidencia no es evidencia de ausencia"

¿Qué se requiere en tal caso (por ejemplo, si el resultado sugiere la asociación pero el valor p fue mayor de 0,05)? Repetir el estudio o continuarlo hasta alcanzar un tamaño muestral mayor.

La otra medida que nos aporta información sobre el rol del azar en los resultados de un estudio, los intervalos de confianza, la abordaremos en un número posterior de la serie siguiente al referirnos a la precisión de los resultados.

Errores en la unidad de análisis

Este tipo de error ocurre principalmente en los ensayos aleatorizados por grupos [clúster], cuando se analiza los datos de los sujetos individualmente como si fueran independientes. Dado que lo que aleatorizó en estos estudios fueron los grupos -por ejemplo, los consultorios- y no los sujetos uno a uno, lo que corresponde es comparar los resultados de los grupos entre sí.

Usar métodos estadísticos habituales en este tipo de estudio, que no toman en cuenta los "efectos de grupo" y en definitiva la correlación que existe entre los sujetos de un mismo grupo; tienden a producir una sobreestimación de la efectividad de la intervención, dando lugar a valores p extremadamente bajos e intervalos de confianza muy estrechos.

Imagine un estudio en que se aleatorizó 10 consultorios para evaluar un programa de manejo de obesidad en la población. Cinco consultorios aplicaron el programa (la intervención) y cinco no. Cada consultorio tiene 50 pacientes obesos en control. La variable de resultado es la diferencia del Índice de Masa Corporal de los pacientes seis meses después de la intervención. Para evaluar los resultados, deberíamos promediar las mediciones individuales de los 50 pacientes de cada consultorio para obtener desenlaces "por grupo". Entonces, los cinco desenlaces grupales de los consultorios intervenidos se comparan con los correspondientes de los cinco consultorios no intervenidos, usando métodos estadísticos que consideran la correlación o interdependencia de los sujetos de cada grupo entre sí. Esto es distinto a comparar los 250 desenlaces individuales de los pacientes de los consultorios intervenidos, como si fueran una rama de un ensayo tradicional, con los de los 250 pacientes de los consultorios no intervenidos⁴.

Referencias

1. Bland JM, Altman DG. Statistics notes: Multiple significance tests: the Bonferroni method. *BMJ*, Jan 1995; 310: 170. [↑ | Link |](#)
2. Bailey CS, Fisher CG, Dvorak MF. Type II error in the spine surgical literature. *Spine*. 2004 May 15;29(10):1146-9. [↑ | Link |](#)
3. Williams HC, Seed P. Inadequate size of negative clinical trials in dermatology. *Br J Dermatol* 1993;128:317-26. [↑ | Link |](#)
4. Altman DG, Bland JM. Statistics Notes: Units of analysis *BMJ* 1997;314:1874. [↑ | 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.