

# Riesgos competitivos de muerte y equilibrio de Hardy-Weinberg en estudios de casos y controles sobre asociación entre genes y enfermedades

Javier Llorca<sup>a</sup> / Dolores Prieto-Salceda<sup>a</sup> / Onofre Combarros<sup>b</sup> / Trinidad Dierssen-Sotos<sup>a</sup> / José Berciano<sup>b</sup>

<sup>a</sup>Medicina Preventiva y Salud Pública, Facultad de Medicina, Universidad de Cantabria, Santander, España;

<sup>b</sup>Servicio de Neurología, Hospital Universitario Marqués de Valdecilla, Cantabria, Santander, España.

(Competing risks of death and Hardy-Weinberg equilibrium in case-control studies of gene-disease association)

## Resumen

**Objetivo:** Estudiar el impacto de los riesgos competitivos sobre el equilibrio de Hardy-Weinberg y sus consecuencias en los estudios de casos y controles sobre la relación gen-enfermedades de comienzo tardío.

**Métodos:** A partir de una población que, respecto de un determinado gen, se encuentre en equilibrio de Hardy-Weinberg al nacer, se deduce la composición genética cuando el gen está asociado con una enfermedad letal de aparición precoz, y sus consecuencias sobre una enfermedad de comienzo tardío. Cuando se realiza un estudio de casos y controles con muestreo por densidad se obtiene una estimación no sesgada de la *odds ratio*, a pesar de que los controles no estén en equilibrio de Hardy-Weinberg.

**Resultados:** Se presenta un ejemplo en el que el gen mutante se asocia con una mayor mortalidad precoz al romper el equilibrio de Hardy-Weinberg y da lugar a que en un estudio realizado en edades más avanzadas los controles se encuentren en desequilibrio, con una *odds ratio* de 1,61.

**Conclusión:** Aunque las principales causas de desequilibrio de Hardy-Weinberg en los controles son un sesgo de selección o un error de genotipificación, un riesgo competitivo de muerte asociado con el gen mutante puede ocasionar que la muestra de controles esté en desequilibrio de Hardy-Weinberg.

**Palabras clave:** Asociación gen-enfermedad. Equilibrio de Hardy-Weinberg. Estudios de casos y controles. Genética. Ley de Hardy-Weinberg. Riesgos competitivos. Sesgo. Sesgo de selección.

## Abstract

**Objective:** To study the impact of competing risks on Hardy-Weinberg equilibrium and their consequences in case-control studies of gene-late onset disease association.

**Methods:** Based on a population born in Hardy-Weinberg equilibrium for a particular gene, the genetic composition when the gene is associated with a lethal early-onset disease and its consequences on a late-onset disease can be deduced. Odds ratios estimates are unbiased in case-control studies when controls are sampled by density, even if the controls are in Hardy-Weinberg disequilibrium.

**Results:** An example in which a mutant gene is associated with early mortality is presented, producing a departure from Hardy-Weinberg equilibrium; as a result, controls in later ages are in disequilibrium, producing an odds ratio equal to 1.61.

**Conclusion:** Although the main causes of Hardy-Weinberg disequilibrium in controls are selection bias or genotyping error, a competing risk of death associated with the mutant gene would also result in Hardy-Weinberg disequilibrium among controls.

**Key words:** Gene-disease association. Hardy-Weinberg equilibrium. Case-control studies. Genetics. Hardy-Weinberg law. Competing risks. Bias. Selection bias.

Correspondencia: Javier Llorca.

Medicina Preventiva y Salud Pública. Facultad de Medicina. Avda. Cardenal Herrera Oria, s/n. 39011 Santander. Cantabria. España.

Correo electrónico: llorcaj@unican.es

Recibido: 24 de noviembre de 2004. Aceptado: 4 de marzo de 2005.

## Introducción

El equilibrio de Hardy-Weinberg (HW) es una pieza clave cuando se estudian asociaciones entre mutaciones genéticas y enfermedades. En estudios de casos y controles, los controles tienen como función representar a la población general; por

lo tanto, la muestra de controles debe estar en equilibrio de HW, mientras que el desequilibrio de HW en los casos sugiere una asociación entre la mutación y la enfermedad<sup>1</sup>.

Xu et al<sup>2</sup>, en un trabajo reciente, mostraron que una alta proporción de los estudios de casos y controles publicados sobre factores genéticos son inconsistentes con el equilibrio de HW en los controles; además, encontraron que es más frecuente informar de valores  $d$  e  $p < 0,05$  cuando los controles están en desequilibrio. Los autores sugirieron que el desequilibrio HW en controles puede deberse a errores en la genotipificación y mostraron varios procedimientos para evitarlo.

Aunque compartimos en términos generales las conclusiones de Xu et al, el objetivo de este trabajo es mostrar que, en presencia de riesgos competitivos de muerte, una muestra no sesgada de controles puede no encontrarse en equilibrio de HW cuando se estudian enfermedades que ocurren en edades avanzadas.

## Métodos

Considérese una enfermedad (D) que aparece típicamente en estadios tardíos de la vida (podría ser la enfermedad de Alzheimer o muchos cánceres). Las personas con riesgo de desarrollar D no son representativas de la población general, sino de la población que consigue sobrevivir hasta edades avanzadas. A continuación, supóngase que se desea estudiar la relación entre un polimorfismo de un solo nucleótido (G) y D. Denotando como «1» al alelo salvaje con frecuencia  $p$  y como «2» al alelo mutado con frecuencia  $1-p$ , en el momento del nacimiento y de acuerdo con la ley de HW, los genotipos «11», «12» y «22» tendrían frecuencias  $p^2$ ,  $2p(1-p)$  y  $(1-p)^2$ , respectivamente.

Para estudiar esta relación entre G y D se pretende realizar un estudio de casos y controles en el que para cada caso incidente se seleccionan por densidad  $m - 1$  controles entre los individuos que se encuentran en riesgo.

### Equilibrio de Hardy-Weinberg en los controles

Supongamos que el polimorfismo estudiado influye en la mortalidad precoz por otras enfermedades. Por ejemplo, la mortalidad acumulada a los 70 años podría ser  $m_{11}$ ,  $m_{12}$ , y  $m_{22}$  para cada uno de los genotipos. Entonces, las frecuencias genotípicas a los 70 años serían:

$$\frac{p^2 (1-m_{11})}{1-p^2 m_{11} - 2p(1-p) m_{12} - (1-p)^2 m_{22}}$$

$$\frac{2p(1-p) (1-m_{12})}{1-p^2 m_{11} - 2p(1-p) m_{12} - (1-p)^2 m_{22}}$$

$$\frac{(1-p)^2 (1-m_{22})}{1-p^2 m_{11} - 2p(1-p) m_{12} - (1-p)^2 m_{22}}$$

para «11», «12» y «22», respectivamente.

Sólo en algunas situaciones especiales (p. ej., cuando

$$\frac{m_{12}}{m_{11}} = \frac{m_{22}}{m_{12}}$$

esta población se encontrará en equilibrio de HW; lo mismo ocurrirá si se selecciona una muestra de controles representativa de esta población.

### Estimación de la odds ratio

Si en el estudio de casos y controles se toma el genotipo «11» como referencia se puede estimar la *odds ratio* (OR) como razón de tasas (*hazard ratio*) por análisis de supervivencia mediante<sup>3</sup>:

$$h_n(t_i) = h_0(t_i) e^{\beta_{12} X_{12} + \beta_{22} X_{22}}$$

donde  $h_0(t_i)$  es la tasa basal instantánea en el tiempo  $t_i$ :

$$h_0(t_i) = \frac{1}{w_{11} + w_{12} e^{\beta_{12}} + w_{22} e^{\beta_{22}}}$$

$X_{12}$  y  $X_{22}$  son variables indicadoras de los genotipos «12» y «22»,

$$w_k = \frac{n_k(t_i)}{n_{ck} + 1},$$

donde  $n_k(t_i)$  es el número de supervivientes con el genotipo  $k$  en el tiempo  $t_i$  (con la terminología de la sección anterior, a los 70 años se tendría  $n_{11}(70) = p^2 \times m_{11} \times N$ ,  $n_{12}(70) = 2p \times (1-p) \times m_{12} \times N$  y  $n_{22}(70) = (1-p)^2 \times m_{22} \times N$ , siendo  $N$  el total de población al nacimiento), y  $n_{ck}$  el número de controles con el genotipo  $k$ .

Las tasas instantáneas para cada genotipo son:

$$h_{11}(t_i) = h_0(t_i)$$

$$h_{12}(t_i) = \frac{e^{\beta_{12}}}{w_{11} + w_{12} e^{\beta_{12}} + w_{22} e^{\beta_{22}}}$$

$$h_{22}(t_i) = \frac{e^{\beta_{22}}}{w_{11} + w_{12} e^{\beta_{12}} + w_{22} e^{\beta_{22}}}$$

Por último, bajo la asunción de riesgos proporcionales, las OR para los genotipos «12» y «22» son:

$$OR = \frac{H_{12}(t_i)}{H_{11}(t_i)} = e^{\beta_{12}} \quad (1)$$

$$OR = \frac{H_{22}(t_i)}{H_{11}(t_i)} = e^{\beta_{22}} \quad (2)$$

que resultan ser insesgadas e independientes de la mortalidad diferencial por genotipos en la población general.

En consecuencia, si  $\beta_{12} = 0$  o  $\beta_{22} = 0$  y  $m_{11} = m_{12}$  o  $m_{11} = m_{22}$  se encontrará una situación que cumple simultáneamente: a) hay una asociación entre gen y enfermedad; b) los controles están en desequilibrio de HW, y c) la estimación de la OR es insesgada.

Se ilustra el impacto de esta situación con un ejemplo numérico en el que el genotipo salvaje está asociado con un aumento de la mortalidad antes de los 70 años.

## Resultados

El siguiente ejemplo se desarrolla en la tabla 1. Se parte de una población de 1.000.000 de recién nacidos con frecuencias alélicas 0,7 para el alelo 1 y 0,3 para el alelo 2 y con completo equilibrio de HW en el momento del nacimiento. El gen está asociado con una mortalidad antes de los 70 años del 20% en los genotipos «11» y «12» y del 50% en el genotipo salvaje (es decir:  $m_{11} = m_{12} = 0,2$ ;  $m_{22} = 0,5$ ). Una muestra de 1.000 controles representativa de esta población a los 70 años debe tener la misma distribución genotípica y se encuentra en desequilibrio de HW ( $p = 0,005$  para el equilibrio de HW en controles).

Al comparar este grupo de controles con 1.000 casos que tengan la distribución genotípica inicial se encuentra una asociación clara entre la enfermedad y el genotipo «22» ( $OR_{22} = 1,61$ ; intervalo de confianza [IC] del 95%, 1,13-2,28;  $p = 0,008$ ). Debe observarse que esta

distribución de casos se utiliza como ejemplo y no como hipótesis (posteriormente se discutirá este punto).

En la figura 1 se muestra el cambio de la OR en función de la mortalidad acumulada en los individuos con genotipo «22» para 3 mortalidades en el grupo «11»: 0,2 (línea continua), 0,4 (línea discontinua) y 0,6 (línea de puntos).

## Discusión

El impacto de riesgos competitivos de muerte sobre la asociación gen-enfermedad en procesos patológicos de comienzo tardío ha recibido poca atención en la bibliografía médica. Cuando un gen se asocia de manera causal con una enfermedad que se inicia en edades avanzadas, si está simultáneamente asociado con otra enfermedad de inicio más precoz y letal pueden producirse situaciones paradójicas; por ejemplo, se ha publicado que la OR entre el alelo  $\epsilon 4$  del gen de la apolipoproteína E y la enfermedad de Alzheimer disminuye con la edad<sup>4</sup>. Este resultado podría ser consecuencia de la también conocida asociación entre este alelo y el riesgo de muerte por infarto de miocardio. Igualmente, se ha descrito un sesgo similar al de confusión cuando se investiga la interacción gen-ambiente, si el factor ambiental está asociado con una causa de muerte precoz<sup>5</sup>.

Los controles seleccionados ortodoxamente en edades avanzadas podrían no estar en equilibrio de HW; sin embargo, como muestran las fórmulas (1) y (2), las OR que se obtienen son insesgadas a pesar de la mortalidad diferencial debida a cada genotipo. Este resultado es independiente del modelo que se utilice para la mortalidad (paramétrico o no).

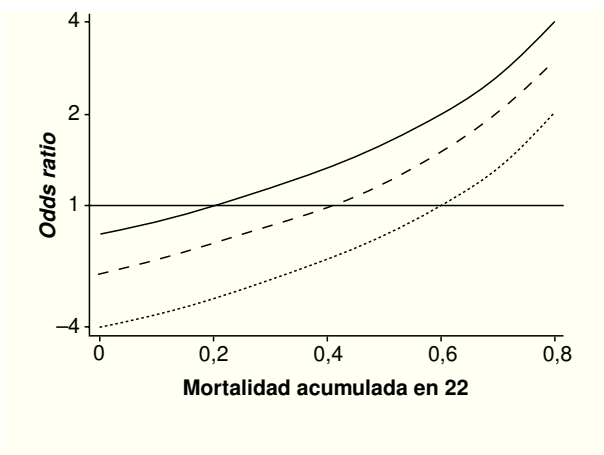
Otras alternativas de muestreo (distintas del estudio de casos y controles sin emparejamiento) pueden ser problemáticas. Un estudio de cohorte y casos, en el que la subcohorte que actúa de control se seleccione en el momento de nacimiento, debe preservar el equilibrio de HW en el grupo de comparación, pero requiere el seguimiento de la cohorte durante décadas para

Tabla 1. Ejemplo hipotético del impacto de un riesgo competitivo de muerte sobre el equilibrio de Hardy-Weinberg (HW)

	Genotipo 11	Genotipo 12	Genotipo 22
Población en el nacimiento (n = 1.000.000)	490.000	420.000	90.000
Mortalidad acumulada hasta los 70 años	0,2	0,2	0,5
Población a los 70 años	392.000	336.000	45.000
Controles representativos de la población de 70 años (n = 1.000)	507	435	58
Casos en el equilibrio de HW original (n = 1.000)	490	420	90
Odds ratio (p)	1 (referencia)	1,00 (0,992)	1,61 (0,008)

La población en el nacimiento y los casos están en equilibrio de HW. Los controles están en desequilibrio de HW ( $p = 0,005$ ).

**Figura 1. Cambio en la odds ratio (OR) en función de la mortalidad acumulada en la población con el genotipo «22».** Las 3 líneas representan la OR cuando la mortalidad acumulada en el genotipo «11» es 0,2 (línea continua), 0,4 (línea discontinua) y 0,6 (línea de puntos).



identificar los casos en edad avanzada. Por último, un estudio de casos y controles emparejado por edad introduce, como es sabido, un sesgo de selección en los controles que ya no representan a la población general, por lo que no tiene sentido plantearse si se encuentran o no en equilibrio de HW.

El número de artículos destinados a explorar una asociación entre gen y enfermedad es creciente y, con frecuencia, los resultados positivos iniciales no pueden ser reproducidos en estudios posteriores<sup>6</sup>; igualmente se ha señalado que es más frecuente informar sobre resultados significativos cuando la muestra de controles no se encuentra en equilibrio de HW. Un metaanálisis ha mostrado que la principal fuente de heterogeneidad entre los estudios de casos y controles que analizan un mismo gen es el desequilibrio de HW en los controles<sup>7</sup>.

En este artículo se ha utilizado un ejemplo *ad hoc* en el que la influencia de los genotipos en la enfermedad «reequilibra» los casos y hace que éstos se encuentren en el equilibrio de HW original. Ésta no es una hipótesis del estudio: la ley de HW se refiere a la

población general y, por tanto, afecta únicamente a los controles; las fórmulas 1 y 2 son válidas con independencia de la distribución genotípica en los casos.

Cuando se presenta el desequilibrio de HW en los controles, antes de considerar otras posibilidades, los investigadores deberían sospechar que su muestra está sesgada (no es representativa) o que ha ocurrido un error de genotipificación<sup>2</sup>. Cuando estas 2 opciones se puedan descartar, nuestro ejemplo muestra que, en enfermedades que ocurren en personas de edad avanzada (p. ej., enfermedad de Alzheimer), si un genotipo está fuertemente asociado con la mortalidad general, la OR puede estar insesgada incluso si la población de interés y la muestra de controles no se encuentran en equilibrio de HW.

#### Bibliografía

1. Clayton D. Population association. En: Balding DJ, Bishop M, Cannings C, editors. Handbook of statistical genetics. West Sussex: Wiley 2001.
2. Xu J, Turner A, Little J, Bleeker ER, Meyers DA. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? Hum Genet. 2002;111:573-4.
3. Borgan O, Goldstein L, Langholz B. Methods for the analysis of sampled cohort data in the Cox proportional hazards model. Ann Statist. 1995;23:1749-78.
4. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayenso R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA. 1997;278: 1349-56.
5. Elbaz A, Alperovitch A. Bias in association studies resulting from gene-environment interactions and competing risks. Am J Epidemiol. 2002;155:265-73.
6. Llorca J, Dierssen-Sotos T, Combarros O, Berciano J. Consistency in gen-Alzheimer's Disease studies. J Epidemiol Commun Health. 2005;59:83-5.
7. Combarros O, Infante J, Llorca J, Peña N, Fernández-Viadero C, Berciano J. Interaction between interleukin-6 and intercellular adhesion molecule-1 genes and Alzheimer's disease risk. J Neurol. 2005;252:485-7.