

Diseños de investigación en epidemiología genética

Eugenia Flores-Alfaro,¹
Ana I. Burguete-García²
y Eduardo Salazar-Martínez³

Forma de citar: Flores-Alfaro E, Burguete-García AI, Salazar-Martínez E. Diseños de investigación en epidemiología genética. Rev Panam Salud Publica. 2012;31(1):88-94.

SINOPSIS

Se describen los distintivos de los diseños de investigación epidemiológica que son utilizados con mayor frecuencia en los estudios de asociación genética. Los estudios de casos y controles proporcionan un método eficiente para evaluar asociaciones entre genes candidatos y enfermedad. Los estudios de cohorte, por su parte, si bien aportan un mayor grado de causalidad, no son eficientes para la exploración inicial en la identificación de las asociaciones gen-enfermedad. Los estudios transversales son menos costosos, requieren períodos de tiempo más cortos y son de utilidad para estimar la prevalencia de enfermedades, de factores de riesgo o de variantes genéticas. Los estudios basados en familias han sido exitosos para encontrar alelos que confieren mayor riesgo para el desarrollo de enfermedades de transmisión mendeliana.

Palabras clave: investigación genética; diseño de investigaciones epidemiológicas; estudios de asociación genética; riesgo relativo.

¹ Universidad Autónoma de Guerrero, Unidad Académica de Ciencias Químico Biológicas, Laboratorio de Enfermedades Crónicas Degenerativas. Chilpancingo, Guerrero, México. La correspondencia se debe dirigir a Eduardo Salazar-Martínez. Correo electrónico: efloresa@espm.insp.mx

² Instituto Nacional de Salud Pública, Centro de Investigación sobre Enfermedades Infecciosas. Cuernavaca, Morelos, México.

³ Instituto Nacional de Salud Pública, Centro de Investigación en Salud Poblacional; Instituto Mexicano del Seguro Social. Cuernavaca, Morelos, México.

Los estudios de asociación genética (EAG) tienen como propósito determinar el efecto de los factores genéticos sobre las enfermedades y sus posibles interacciones con los elementos del medio ambiente. Morton define a la epidemiología genética como la ciencia que estudia la etiología, la distribución y el control de las enfermedades en familias y las causas hereditarias de enfermedad en las poblaciones (1). Con tal fin utiliza principios y métodos propios para estudiar factores o cambios genéticos, y su relación con el desarrollo de enfermedades, a partir de leyes que rigen la transmisión de genes de padres a hijos (2, 3).

En la actualidad se han venido realizando esfuerzos por identificar y analizar los polimorfismos que confieren susceptibilidad a enfermedades complejas, así como las interacciones gen-gen y gen-ambiente (4). En las Américas, 68% de los EAG se realizan en cuatro países: Estados Unidos (30%), Brasil (20%), México (10%) y Canadá (8%). El presente trabajo tiene como objetivo describir los distintivos de los diseños de investigación epidemiológica que son utilizados con mayor frecuencia en los estudios de asociación genética.

DISEÑOS DE CASOS Y CONTROLES

Los diseños de casos y controles se utilizan preferiblemente para evaluar asociaciones entre la variación en genes candidatos y enfermedades complejas porque arrojan información sólida acerca de factores de riesgo ambientales, bioquímicos o genéticos después de que ha ocurrido la enfermedad de interés. En estudios de base poblacional permiten analizar i) factores que originan mutaciones, ii) indicadores genéticos inespecíficos (como la consanguinidad y la mezcla de razas) en la etiología de las enfermedades y iii) el papel etiológico de los rasgos genéticos (5, 6).

Fortalezas

Los EAG que utilizan el modelo de casos y controles ofrecen mayor eficiencia estadística que otros modelos de investigación porque hacen posible seleccionar controles no relacionados genéticamente y con base poblacional, son menos costosos, se realizan en menor tiempo, permiten examinar asociaciones de múltiples genes con la enfermedad de interés y conllevan un análisis de datos más sencillo (7, 8). Más todavía, según refiere Khoury, el modelo de casos y controles se adapta bien a la epidemiología genética por la estabilidad de los marcadores genéticos relacionados al riesgo de enfermedades, en comparación con otros factores (p. ej. nutricionales u ocupacionales); su idoneidad para evaluar interacciones gen-gen, gen-ambiente y sus

efectos en el desarrollo de la enfermedad, y su utilidad para investigar enfermedades poco comunes, como por ejemplo defectos de nacimiento, cáncer, asma, diabetes, esclerosis múltiple, entre otras (6).

Debilidades

El diseño de casos y controles utilizado en EAG asimismo conllevan probabilidades de introducir distintos tipos de sesgos.

Sesgo de selección. Se puede presentar cuando la selección de los controles se realiza en relación a la enfermedad o a la exposición, y es más frecuente cuando se utilizan controles hospitalarios o familiares (9).

Sesgo de supervivencia. Este sesgo se introduce cuando los casos estudiados difieren de aquellos que no lo fueron, tales como los de una forma particular de enfermedad letal y que murieron antes de que pudieran ser involucrados en el estudio. Podría darse, por ejemplo, al tratar de estimar una asociación genética con el síndrome de Down, ignorando que una buena parte de los fetos con este síndrome son abortados espontáneamente y que este síndrome esté relacionado con la variante genética en estudio (9).

Sesgo de recuerdo. Es común que los casos en comparación con los controles recuerden mejor su historia de exposición ambiental después del diagnóstico del evento (9).

Sesgo de información. Se puede dar por una medición inadecuada de las exposiciones ambientales, de los marcadores biológicos o en la identificación de la variante genética en estudio (9). El error de medición de los genotipos se puede verificar determinando el equilibrio de Hardy-Weinberg (EHW) en los controles, suponiendo que los alelos se distribuyen aleatoriamente en la población (10, 11).

Para determinar el EHW para un polimorfismo bialélico (A/G) se utiliza una prueba de Ji cuadrada ($X^2_{HW} = \sum_{i=1}^3 \frac{(O_i - E_i)^2}{E_i}$), donde los valores observados (O) corresponden a las frecuencias absolutas obtenidas para cada genotipo (p. ej. $n_{AA} = 135$, $n_{AG} = 118$ y $n_{GG} = 31$). Las frecuencias esperadas (E) se estiman como se indica en el cuadro 1, para lo cual primero se calculan las frecuencias alélicas (\hat{P}) sumando la contribución de cada alelo en los genotipos y dividiendo entre N por duplicado; por ejemplo, para la frecuencia del alelo A se tendría: $(135 + 135 + 118)/(284 \times 2) = 0,683$ y para el alelo G: $(31 + 31 + 118)/(284 \times 2) = 0,317$. Luego se calculan las frecuencias E para cada genotipo (AA: $0,683^2 \times 284 = 132,5$; AG: $2 \times 0,683 \times 0,317 \times 284 = 123$ y GG: $0,317^2 \times 284 = 28,5$) y se sustituyen:

$$X^2_{H-W} = \left[\frac{(135 - 132,5)^2}{132,5} \right] + \left[\frac{(118 - 123)^2}{123} \right] + \left[\frac{(31 - 28,5)^2}{28,5} \right] = 0,469$$

CUADRO 1. Cálculo del equilibrio génico de Hardy-Weinberg para un polimorfismo bialélico

Valor	Genotipo			Total
	AA	AG	GG	
Observado (O)	n_{AA}	n_{AG}	n_{GG}	N
Esperado (E)	$\hat{P}^2 N$	$2\hat{P}(1 - \hat{P})N$	$(1 - \hat{P})^2 N$	N_E

Fuente: elaboración de los autores.
n: representa las frecuencias absolutas; *N*: el número total de individuos incluidos; \hat{P} : frecuencias alélicas relativas.

Finalmente, se obtiene un valor P asociado a una distribución de X^2 ($P = 0,497$), concluyendo que el polimorfismo en cuestión se encuentra en EHW.

Sesgo de publicación. Se reconoce que suele haber una publicación selectiva de investigaciones con asociaciones positivas. Un análisis realizado en 19 estudios de casos y controles del polimorfismo DD del gen de la enzima convertidora de angiotensina y enfermedad coronaria, demuestra que hubo sesgo de publicación a favor de los estudios con resultados positivos (12). Asimismo, Ioannidis y colaboradores reportan sesgo de publicación después de haber analizado los efectos de cinco genes implicados en asociaciones gen-enfermedad hallados en 50 metanálisis (13).

Sesgo de confusión. Los estudios en epidemiología genética son susceptibles a sesgos por estratificación poblacional, que tienen lugar cuando i) los controles difieren étnicamente de los casos, ii) las frecuencias alélicas varían entre las etnias/razas o iii) las tasas de ocurrencia de la enfermedad son diferentes por grupos étnicos —aunque dos trabajos reportan que la estratificación poblacional conduce a un sesgo mínimo (14–16). Para minimizar este sesgo se recomienda aparear a los controles por origen étnico/raza o medir marcadores genéticos de ancestralidad o controlar —en el análisis de los datos— por autorreporte del origen étnico. Por otra parte, se puede presentar confusión por la presencia de desequilibrio de ligamiento entre el marcador genético en estudio y la variante genética de riesgo verdadera, cuando esta última está cercana al marcador en estudio y se asocia con el evento, aunque en realidad se asocia con la variante verdadera (17).

Si se han de reducir los sesgos al mínimo, se requiere de un buen diseño que comprenda la selección adecuada de casos y controles, la capacitación y el cegamiento del personal que realice las mediciones clínicas y de laboratorio, métodos idóneos para la medición de los factores ambientales y biológicos (5, 9, 18).

DISEÑOS DE COHORTE

Los diseños de cohorte se utilizan en estudios de seguimiento y suponen la realización de al menos dos mediciones en diferentes tiempos. En este tipo de trabajos, generalmente se almacenan muestras biológicas que servirán para medir marcadores bioquímicos o identificar variantes genéticas que permitan evaluar

asociaciones entre marcadores genéticos y enfermedad o rasgos fenotípicos, utilizando métodos como el análisis de supervivencia. También son el punto de partida para diseñar estudios de casos y controles anidados en una cohorte, o estudios de caso-cohorte (19). Existen opiniones divergentes respecto a la utilidad de los diseños de cohorte en el ámbito de la epidemiología genética. Hattersley menciona que a menos que la enfermedad sea muy común, el tamaño de muestra para los estudios de casos y controles anidados a menudo será pequeño, además de que la naturaleza de no selección de los casos podría comprometer la eficiencia estadística (5).

Fortalezas

Los estudios de cohorte permiten obtener información precisa sobre la ocurrencia de la enfermedad (casos incidentes) y sobre diversos factores de exposición medidos en el tiempo; evaluar posibles interacciones entre uno o varios genes —o gen-ambiente— y el desarrollo de una enfermedad, y realizar estudios de agregación familiar y de estratificación poblacional. Mediante el seguimiento y la recolección retrospectiva de datos sobre exposiciones ambientales se pueden lograr resultados libres de sesgos de selección, de supervivencia y de recuerdo, así como realizar estudios de múltiples eventos (19, 20).

Debilidades

Los integrantes de una cohorte a menudo no son representativos de la población en riesgo porque pertenecen a grupos de alta exposición, como por ejemplo el Estudio Multicéntrico de Sida que se emprendió en 1983 con voluntarios homosexuales de cuatro estados de los Estados Unidos (21). Este tipo de investigaciones suelen ser costosas y muy extendidas en el tiempo, no contar con suficientes muestras biológicas, y experimentar cambios en los métodos de medición o de diagnóstico. Sus resultados dependen en gran medida de la calidad de los datos y de una férrea disciplina en el seguimiento, o de si las pérdidas se asocian con una mayor letalidad debido a la variante genética en estudio (20).

También se pueden realizar estudios de cohorte retrospectivos, como por ejemplo en la población de un área geográfica determinada o en trabajadores de una institución, en cuyos casos se identifican las variantes genéticas y se las relaciona con uno o varios eventos. En los estudios de cohorte retrospectivos se debe contar con registros médicos de buena calidad y con pacientes que sean capaces de reconstruir adecuadamente su historia de exposición. Estos diseños son menos costosos, pero también más susceptibles a sesgos (19).

DISEÑOS TRANSVERSALES

En los EAG de diseño transversal se obtiene una muestra poblacional al azar y se determina una sola

vez la prevalencia de enfermedad, de tal forma que las personas con el factor de exposición ambiental o genético de interés se comparan con la prevalencia del evento.

Fortalezas

Los estudios transversales tienen la ventaja de ser menos costosos ya que requieren períodos de tiempo más cortos. Resultan particularmente útiles para estimar la prevalencia de enfermedades, factores de riesgo o variantes genéticas, así como para proponer hipótesis etiológicas y aportar datos relevantes a la gestión y planificación en salud pública (9, 18).

Debilidades

Aun cuando los polimorfismos genéticos no cambian a lo largo de la vida y es posible establecer asociaciones gen-enfermedad a partir de diseños transversales, estos estudios tienen varias limitaciones, entre ellas la probabilidad de ser afectados por sesgos. Por ejemplo, al estudiar casos prevalentes se pueden originar sesgos de supervivencia, de información por medición retrospectiva de las exposiciones ambientales y de selección al brindar una muestra representativa de la población objetivo (9, 18).

DISEÑOS BASADOS EN FAMILIAS

Los estudios de asociación genética basados en familias (EAGBF) han sido exitosos para encontrar alelos que confieren un riesgo alto para el desarrollo de enfermedades monogénicas o de transmisión mendeliana, aunque no lo han sido tanto para el estudio de genes asociados a enfermedades de rasgos complejos (22, 23). En general, los objetivos de los EAGBF son evaluar el grado de segregación familiar a través del número de integrantes que presentan la misma enfermedad; determinar si tal segregación se relaciona con exposiciones ambientales, susceptibilidad heredada o herencia cultural, y establecer la forma de herencia de la susceptibilidad genética (4). Estos diseños utilizan diversos enfoques: i) basados en tríos (selección de un caso afectado —caso índice o probando— y compararlo con los dos padres), ii) parejas de hermanos (hermano afectado comparado con un hermano no afectado), iii) parejas de padre e hijo y iv) familias extendidas (familiar afectado, padres y abuelos) (22–24).

Fortalezas

Al seleccionar a las familias por tener integrantes afectados, los EAGBF aportan mayor información sobre enfermedades hereditarias y sobre la prevalencia de la enfermedad en un subgrupo de familiares. Además, al utilizarse controles de la misma familia, no hay riesgo de confusión por estratificación poblacional.

Debilidades

En los EAGBF se puede dar una participación incompleta de los integrantes de la familia, dado que no siempre hay disponibilidad de padres controles o abuelos cuando la enfermedad es de aparición tardía, suelen ser menos eficientes para el estudio de polimorfismos de baja penetrancia o pueden presentar sobrepareamiento por utilizar controles familiares. Adicionalmente, debido a que los datos y mediciones obtenidos de los integrantes de una familia no son independientes, requieren la aplicación de métodos estadísticos especiales para el análisis, pueden no ser efectivos en la detección de interacciones ambientales cuando los polimorfismos o factores de riesgo ambientales son de baja frecuencia (dependen del tamaño de muestra), no representan a la población en riesgo y los resultados solo son extrapolables a las familias estudiadas (25).

DESEQUILIBRIO DE LIGAMIENTO

La asociación de una enfermedad con la variación de un alelo puede correlacionarse con alelos cercanos a esta variante, dando lugar al denominado “desequilibrio de ligamiento”, es decir la combinación de alelos a lo largo de un cromosoma denominado haplotipo. Un haplotipo se forma de alelos que se transmiten juntos de un solo *locus* o de múltiples *loci* en un cromosoma (26). Por otro lado, puede hacer referencia a un conjunto de polimorfismos de un solo nucleótido que se encuentran asociados (27). En un estudio de simulación llevado a cabo con datos de familias (tríos o dos hijos afectados), Becker y colaboradores observaron que el análisis de asociación entre enfermedades y haplotipos fue más eficiente, y que las familias con varios hermanos afectados tienen mayor carga genética relacionada con la enfermedad y, por lo tanto, mayor eficiencia estadística (28).

MÉTODOS DE TIPIFICACIÓN Y SECUENCIACIÓN

Para la identificación de variantes genéticas (polimorfismos de un solo nucleótido, secuencias repetidas o inserciones-delecciones) se han diseñado diferentes técnicas, entre ellas la reacción en cadena de la polimerasa/polimorfismos de longitud del fragmento de restricción (RCP/PLFR), la RCP en tiempo real y la hibridación. Dado que la secuenciación del ADN —que es el patrón de oro para identificar nuevos polimorfismos— es muy costosa, actualmente se han desarrollado otras opciones como los marcadores microsatelitales usados en el rastreo genómico, técnica idónea para identificar genes de susceptibilidad a enfermedades mediante diseños de casos y controles, o de cohortes, útil también para el análisis de ligamiento en familias. También se aplican las técnicas de resecuenciación para detectar cambios en el ADN (mutaciones) relacionados con enfermedades, los cuales son

identificados por comparación con una secuencia de referencia (29). En el cuadro 2 se resumen las principales fortalezas y debilidades de los diseños utilizados en epidemiología genética.

Por otra parte, a partir de diferentes EAG se realizan metanálisis con el propósito de hacer una síntesis sistemática y estadística (resumen cuantitativo) de la información obtenida por estudios similares. Los metanálisis presentan mayor eficiencia estadística para la estimación del efecto y controlan mejor la varianza interestudios, cuya combinación ofrece estimaciones más confiables. Si bien los metanálisis no son estudios científicos en sí mismos ni controlan las fuentes de sesgos en cada uno de los estudios, aportan información importante para la toma de decisiones clínicas y científicas (23, 30).

ESTIMACIÓN DEL EFECTO

Cuando se evalúa el efecto de la variación en un gen (p. ej. un polimorfismo bialélico A y G) sobre un evento, o cuando no se conoce el genotipo de riesgo de enfermedad en una población específica, se puede realizar bajo la estructura de cinco posibles modelos de herencia: i) dominante, ii) recesivo, iii) codominante, iv) sobredominante y v) aditivo (10, 31, 32). En el ejemplo presentado en el cuadro 3, se parte del supuesto de que el alelo menor (G) modifica el riesgo de la enfermedad de interés, resumiéndose la información de estos modelos y las ecuaciones que estiman el riesgo relativo en razones de momios (RM) o en coeficientes de regresión (β).

Modelo dominante. Una sola copia del alelo G es suficiente para modificar el riesgo y el ser portador de dos copias lo modifica en igual magnitud, es decir, que tanto los portadores del genotipo AG como los del GG tienen el mismo riesgo. Por tanto, la combinación AG + GG se compara con el genotipo AA.

Modelo recesivo. Hacen falta dos copias de la variante G para modificar el riesgo: los heterocigotos (AG) y los homocigotos del alelo de mayor frecuencia (AA) tienen el mismo riesgo, de tal forma que el genotipo GG se compara con AG + AA.

Modelo codominante. Cada genotipo proporciona un riesgo de enfermedad diferente y no aditivo.

Modelo sobredominante. Los individuos homocigotos para el genotipo más frecuente (AA) y el genotipo variante (GG) tienen el mismo riesgo.

Modelo aditivo. Cada copia de G modifica el riesgo en una cantidad aditiva, por lo tanto los homocigotos GG tienen el doble de riesgo que los heterocigotos AG.

En la estimación del riesgo de la variante genética en los estudios de asociación basados en familias, los métodos estadísticos parten de datos correlaciona-

CUADRO 2. Principales fortalezas y debilidades de los diseños utilizados en los estudios de asociación genética

Diseño	Fortalezas	Debilidades
Cohortes	<ul style="list-style-type: none"> • Cuando se recolectan y almacenan muestras biológicas es posible realizar estudios de casos y controles anidados o de caso-cohorte, incluso de eventos raros cuando la cohorte es grande. • Al tener casos incidentes, se estiman con precisión las tasas de incidencia y el riesgo relativo. • Libres de sesgos de temporalidad, de recuerdo y de supervivencia. • Menor probabilidad de sesgo por estratificación poblacional. • Se pueden estudiar múltiples eventos, como también interacciones entre genes y factores de riesgo ambientales en el desarrollo de la enfermedad. 	<ul style="list-style-type: none"> • Costosos. • De larga duración. • Pérdidas en el seguimiento que pueden generar sesgo de selección. • Cuando las cohortes son de grupos de alta exposición, no son representativas de la población en riesgo.
Casos y controles no relacionados	<ul style="list-style-type: none"> • De gran utilidad para el estudio de enfermedades raras. • Mayor eficiencia que los estudios con controles relacionados. • De menor costo. • Relativamente rápidos de realizar. • Se pueden estudiar asociaciones de múltiples genes con el evento. • Buena eficiencia para evaluar polimorfismos de baja penetrancia o polimorfismos comunes. 	<ul style="list-style-type: none"> • Susceptibles a sesgos (p. ej. por estratificación poblacional, de supervivencia y de recuerdo). • Falta de potencia para detectar interacciones gen-ambiente cuando el polimorfismo o el factor ambiental es de baja frecuencia. • No permiten estimar los parámetros poblacionales (riesgo relativo y riesgo atribuible poblacional).
Transversales	<ul style="list-style-type: none"> • Fáciles de conducir. • Menos costosos. • Se realizan en menos tiempo. • Se pueden evaluar varias exposiciones y enfermedades en el mismo estudio. 	<ul style="list-style-type: none"> • Al utilizar casos prevalentes se puede generar sesgo de supervivencia (si las variantes genéticas se asocian con una pobre supervivencia). • No se puede determinar causalidad. • Susceptibles a sesgos: de selección, de recuerdo, de información y por estratificación poblacional.
Casos y controles basados en familias	<ul style="list-style-type: none"> • Permite evaluar el grado de agregación familiar de la enfermedad y sus potenciales asociaciones con exposiciones ambientales. • Se evalúa el ligamiento en familias. • Libre de sesgo por estratificación poblacional. 	<ul style="list-style-type: none"> • Difícil reclutamiento de los miembros de una familia. • No disponibilidad de padres controles por inicio tardío de la enfermedad. • En ocasiones son menos eficientes que los estudios de casos y controles. • Se puede presentar sobrepareamiento. • No representan a la población en riesgo y la validez externa se limita a las familias estudiadas.

Fuente: elaboración de los autores.

dos, estimando los coeficientes de regresión (RM o β) a través de modelos lineales generalizados, como regresión logística condicional o estimación de ecuaciones generalizadas (EEG) (9, 33, 34). Una forma simple de estimar el efecto de la variante de interés para un rasgo dicotómico es el cálculo de la RM para datos pareados, utilizando la prueba de significancia de McNemar (18). Estos estimadores se pueden obtener utilizando algún software estadístico (p. ej. SPSS, STATA o SAS). Benyamin y colaboradores (35) realizan una amplia revisión de los diferentes programas estadísticos utilizados en el análisis de asociación, o en el cálculo de la prueba de desequilibrio de transmisión de datos basados en familias (p. ej. HAPLOVIEW, FBAT, GHOST y MERLIN).

Interacción gen-ambiente y ge-gen

La interacción gen-ambiente y ge-gen tiene lugar cuando el riesgo de un evento Y , entre los individuos expuestos y no expuestos a un factor de riesgo A , difiere (es heterogéneo) en presencia de una tercera variable Z . En este contexto, la definición de interacción es equivalente a la definición de modificación de efecto, mientras que desde la perspectiva estadística

es el producto de los dos factores de exposición (gen-ambiente) y se conoce como interacción multiplicativa (9, 22, 36).

La confusión y la interacción son fenómenos distintos. La confusión no es deseable porque sus efectos dificultan la evaluación de una asociación causal, en tanto que la interacción —si es verdadera— forma parte de la red causal de la enfermedad y tiene aplicaciones importantes en la prevención de las enfermedades (9, 22). Para estimar el efecto de la interacción gen-ambiente se pueden elaborar tablas de contingencia como la que se muestra en el cuadro 4.

CONCLUSIONES

Los estudios de casos y controles presentan la mejor oferta metodológica en el descubrimiento inicial de la relación gen-enfermedad y en enfermedades de baja frecuencia, además de que requieren menores períodos de tiempo y son menos costosos. Los estudios de cohorte, por su lado, son más efectivos para confirmar estas relaciones causales, mientras que los basados en familias permiten ampliar el conocimiento de la herencia en la transmisión de enfermedades de rasgos complejos.

CUADRO 3. Modelos de herencia y estadísticos para la evaluación del efecto de un polimorfismo bialélico de un gen sobre un evento

Modelo de herencia	Genotipo	Casos (C)	Controles (C̄)	RM	Modelo estadístico para un evento dicotómico
Dominante	AA	C ₀	C̄ ₀	1,0	$\log\left[\frac{P}{(1-P)}\right] = \beta_0 + \beta_1(AG + GG) + \beta_2X_2 + \dots + \beta_iX_i$
	AG + GG	C ₁	C̄ ₁	$RM = \frac{C_1 \bar{C}_0}{C_0 \bar{C}_1}$	
Recesivo	AA + AG	C ₀	C̄ ₀	1,0	$\log\left[\frac{P}{(1-P)}\right] = \beta_0 + \beta_1(GG) + \beta_2X_2 + \dots + \beta_iX_i$
	GG	C ₁	C̄ ₁	$RM = \frac{C_1 \bar{C}_0}{C_0 \bar{C}_1}$	
Codominante	AA	C ₀	C̄ ₀	1,0	$\log\left[\frac{P}{(1-P)}\right] = \beta_0 + \beta_1(AG) + \beta_2(GG) + \beta_3X_3 + \dots + \beta_iX_i$
	AG	C ₁	C̄ ₁	$RM = \frac{C_1 \bar{C}_0}{C_0 \bar{C}_1}$	
	GG	C ₂	C̄ ₂	$RM = \frac{C_2 \bar{C}_0}{C_0 \bar{C}_2}$	
Sobredominante	AA+GG	C ₀	C̄ ₀	1,0	$\log\left[\frac{P}{(1-P)}\right] = \beta_0 + \beta_1(AG) + \beta_2X_2 + \dots + \beta_iX_i$
	AG	C ₁	C̄ ₁	$RM = \frac{C_1 \bar{C}_0}{C_0 \bar{C}_1}$	
Aditivo					$\log\left[\frac{P}{(1-P)}\right] = \beta_0 + \beta_1(G) + \beta_2X_2 + \dots + \beta_iX_i$

Fuente: elaboración de los autores.

Notas: Dado que en el modelo aditivo los portadores del homocigoto GG tienen el doble de riesgo que los heterocigotos AG, no pueden ser representados y las celdas correspondiente aparecen en blanco. $\log[P/(1-P)]$ = logaritmo del momio, donde P representa la probabilidad de tener el evento. En eventos cuantitativos, el logaritmo del momio es el promedio estimado de este evento (p. ej. índice de masa corporal); β: coeficientes de regresión; β₀: ordenada al origen. Para eventos categóricos el exponencial del coeficiente de regresión (e^β) es igual al estimador del riesgo relativo; RM: (e^{β₁} = RM₁); X: representa a otros factores de exposición. Abreviaturas. C₀: casos con el genotipo de no riesgo (AA), en los modelos estadísticos representa la categoría de referencia; C₁: casos con el alelo de riesgo (AG+GG); C₂: casos con el genotipo de riesgo (GG) (similar representación para los controles); RM: razón de momios.

CUADRO 4. Estimación del efecto de la interacción gen-ambiente sobre un evento

Genotipo	Exposición	Casos (C)	Controles (C̄)	RM
AA = 0	No = 0	C ₀₀	C̄ ₀₀	RM = 1,0
	Sí = 1	C ₀₁	C̄ ₀₁	$RM = \frac{C_{01} \bar{C}_{00}}{C_{00} \bar{C}_{01}}$
AG+GG = 1	No = 0	C ₁₀	C̄ ₁₀	$RM = \frac{C_{10} \bar{C}_{00}}{C_{00} \bar{C}_{10}}$
	Sí = 1	C ₁₁	C̄ ₁₁	$RM = \frac{C_{11} \bar{C}_{00}}{C_{00} \bar{C}_{11}}$

Fuente: elaboración de los autores.

Nota: Abreviaturas: C₀₀: casos sin el genotipo de riesgo y sin la exposición; C₀₁: casos sin el genotipo de riesgo y con la exposición; C₁₀: casos con el genotipo de riesgo (AG o GG) y sin la exposición; C₁₁: casos con el genotipo de riesgo y con la exposición (similar representación para los controles).

La solidez de los resultados en los estudios de asociación genética depende en gran medida del diseño de investigación elegido. De ahí la importancia de planificar cuidadosamente estrategias, métodos y técnicas que ayuden a minimizar los sesgos en el proceso de selección de la muestra, en la recolección de información por encuestas o entrevistas, en la iden-

tificación de las variantes genéticas y en el análisis estadístico de los datos para estimar el efecto —utilizando modelos genéticos y estadísticos adecuados— y vigilar la calibración y la calidad de los instrumentos de medición.

ABSTRACT

Genetic epidemiology research designs

This article describes the features of the epidemiologic research designs most commonly used in genetic association studies. Case-control studies are efficient in evaluating associations between candidate genes and disease. Cohort studies, in contrast, yield a greater degree of causality but are not efficient for the initial exploration to identify gene-disease associations. Cross-sectional studies are less expensive, require less time, and are useful for estimating the prevalence of diseases, risk factors, and genetic variants. Family-based studies have been successful in finding alleles that confer greater risk for developing Mendelian inheritance disorders.

Keywords: genetic research; epidemiologic research design; genetic association studies; relative risk (public health).

REFERENCIAS

1. Morton NE. Genetic epidemiology, genetic maps and positional cloning. *Philos Trans R Soc Lond B Biol Sci.* 2003; 358(1438):1701–8.
2. Sham P. Genetic epidemiology. *Br Med Bull.* 1996;52(3):408–33.
3. Kaprio J. Science, medicine, and the future. *Genetic epidemiology.* *BMJ.* 2000; 320(7244):1257–9.
4. Wyszynski DF. La epidemiología genética: disciplina científica en expansión. *Rev Panam Salud Publica.* 1998;3(1): 26–34.
5. Hattersley AT, McCarthy MI. What makes a good genetic association study? *Lancet.* 2005;366(9493):1315–23.
6. Khoury MJ, Beaty TH. Aplicaciones del método de casos y controles en epidemiología genética. *Bol Of Sanit Panam.* 1996;121(5):421–40.
7. Slager SL, Schaid DJ. Evaluation of candidate genes in case-control studies: a statistical method to account for related subjects. *Am J Hum Genet.* 2001; 68(6):1457–62.
8. Teng J, Risch N. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases II. Individual genotyping. *Genome Res.* 1999;9(3): 234–41.
9. Szklo M, Nieto FJ. *Epidemiología intermedia, conceptos y aplicaciones.* Madrid, España: Ediciones Díaz de Santos; 2003.
10. Iniesta R, Guinó E, Moreno V. Statistical analysis of genetic polymorphisms in epidemiological studies. *Gac Sanit.* 2005;19(4):333–41.
11. Salanti G, Amountza G, Ntzani EE, Ioannidis JP. Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet.* 2005;13(7):840–8.
12. Bracken MB. Genomic epidemiology of complex disease: the need for an electronic evidence-based approach to research synthesis. *Am J Epidemiol.* 2005; 162(4):297–301.
13. Ioannidis JP, Trikalinos TA, Khoury MJ. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am J Epidemiol.* 2006;164(7):609–14.
14. Wang LY, Lee WC. Population stratification bias in the case-only study for gene-environment interactions. *Am J Epidemiol.* 2008;168(2):197–201.
15. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst.* 2000;92(14):1151–8.
16. Millikan RC. Re: Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst.* 2001; 93(2):156–8.
17. Campbell H, Rudan I. Interpretation of genetic association studies in complex disease. *Pharmacogenomics.* 2002;2(6): 349–60.
18. Hernández-Ávila M. *Epidemiología: diseño y análisis de estudios.* México, DF: Editorial Médica Panamericana; 2007.
19. Szklo M. Population-based Cohort Studies. *Epidemiol Rev.* 1998;20(1):81–90.
20. Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. *J Natl Cancer Inst Monogr.* 1999;26(3):39–42.
21. Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR, Jr. The Multi-center AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. *Am J Epidemiol.* 1987;126(2):310–8.
22. Sevilla SD. Metodología de los estudios de asociación genética. *Rev Insuf Cardíaca* 2007;2(3):111–4.
23. Rothman KJ, Greenland S. *Modern Epidemiology.* United States: Lippincott Williams & Wilkins 1998.
24. Cordell HJ, Clayton DG. Genetic association studies. *Lancet.* 2005;366(9491): 1121–31.
25. Whittemore AS, Nelson LM. Study design in genetic epidemiology: theoretical and practical considerations. *J Natl Cancer Inst Monogr.* 1999;26(6):61–9.
26. Gao G, Allison DB, Hoeschele I. Haplotyping methods for pedigrees. *Hum Hered.* 2009;67(4):248–66.
27. International HapMap Consortium. A haplotype map of the human genome. *Nature.* 2005;437(7063):1299–320.
28. Becker T, Knapp M. A powerful strategy to account for multiple testing in the context of haplotype analysis. *Am J Hum Genet.* 2004;75(4):561–70.
29. Checa-Caratachea MA. Polimorfismos genéticos: Importancia y aplicaciones. *Rev Inst Nal Enf Resp Mex.* 2007;20(3): 213–221.
30. O'Rourke K. An historical perspective on meta-analysis: dealing quantitatively with varying study results. *J R Soc Med.* 2007;100(12):579–82.
31. Zintzaras E, Lau J. Synthesis of genetic association studies for pertinent gene-disease associations requires appropriate methodological and statistical approaches. *J Clin Epidemiol.* 2008;61(7): 634–45.
32. Solé X, Guinó E, Valls J, Iniesta R, Moreno V. SNPStats: a web tool for the analysis of association studies. *Bioinformatics.* 2006;22(15):1928–9.
33. Zhang L, Bonham AJ, Li J, Pei YF, Chen J, Papasian CJ, et al. Family-based bivariate association tests for quantitative traits. *PLoS One.* 2009;4(12):e8133.
34. Zorn CJW. Generalized Estimating Equation Models for Correlated Data: A Review with Applications. *Am J Pol Sci.* 2001;45(2):470–90.
35. Benyamin B, Visscher PM, McRae AF. Family-based genome-wide association studies. *Pharmacogenomics.* 2009;10(2): 181–90.
36. Yang Q, Khoury MJ. Evolving methods in genetic epidemiology. III. Gene-environment interaction in epidemiologic research. *Epidemiol Rev.* 1997; 19(1):33–43.

Manuscrito recibido el 19 de septiembre de 2010.
Aceptado para publicación el 12 de julio de 2011.