

Editorial

Aleatorización mendeliana: presente y futuro de los estudios epidemiológicos en cardiología



Mendelian Randomization: Present and Future of Epidemiological Studies in Cardiology

Ignacio Iturrieta-Zuazo^{a,b} y Stefan Walter^{c,d,*}^aDepartamento de Ciencias Médicas Básicas, Facultad de Medicina, Universidad San Pablo CEU, Madrid, España^bDepartamento de Medicina y Cirugía, Psicología, Medicina Preventiva y Salud Pública e Inmunología y Microbiología Médica, Facultad de Ciencias de la Salud, Universidad Rey Juan Carlos, Madrid, España^cDepartment of Epidemiology and Biostatistics, University of California, San Francisco, California, Estados Unidos^dDepartment of Social and Behavioral Sciences, Harvard School of Public Health, Boston, Massachusetts, Estados Unidos

Historia del artículo:

On-line el 20 de noviembre de 2014

SUPERANDO LAS LIMITACIONES DE LOS ESTUDIOS OBSERVACIONALES

No cabe duda de que los estudios epidemiológicos observacionales han resultado de enorme utilidad en la búsqueda de los posibles factores determinantes de la enfermedad y en la elaboración de hipótesis etiológicas. Sin embargo, históricamente se los ha criticado por tener importantes limitaciones a la hora de establecer algo más que una simple asociación entre el fenotipo o la exposición estudiada y la enfermedad de interés¹.

En la literatura científica se encuentran trabajos de investigación de tipo observacional en los que se señala la existencia de algún grado de asociación entre factores de diversa índole (sociales, conductuales, ambientales, nutricionales, genéticos, epigenéticos, etc.) y diferentes enfermedades cardiovasculares. Muchos de estos estudios pasan inadvertidos a la comunidad científica, principalmente debido a que sus hallazgos terminan siendo resultados aislados que otros estudios no han podido reproducir y corroborar posteriormente. Sin embargo, en otras ocasiones, varios estudios independientes pueden llegar a la misma conclusión en sus investigaciones, de manera que el grado de evidencia científica obtenido resulta suficiente para que la comunidad científica se plantee la experimentación en humanos mediante la realización de un ensayo clínico aleatorizado (ECA) que permita comprobar la validez del hallazgo observacional e inferir causalidad.

Desgraciadamente, a menudo se comprueba que las asociaciones planteadas por algunos de estos estudios observacionales no se confirman por los resultados de los ECA. En líneas generales, esta circunstancia se debe a la incapacidad para descartar, por un lado, la presencia de factores o variables de confusión inesperadas o no medidas y, por otro, la existencia de una relación de causalidad inversa entre la exposición y el resultado¹. De manera que, a pesar del uso sistemático de métodos de ajuste (p. ej., análisis

multivariable) con los que se trata de controlar los factores de confusión del estudio, la más que probable existencia de alguna clase de sesgo (que impide controlar la confusión residual) no permite inferir causalidad con seguridad. Otra explicación para la aparente falta de concordancia entre los resultados consiste en que, a menudo, las preguntas de investigación planteadas en el estudio observacional y el ECA son diferentes. Esto ha generado que se desarrolle un enfoque epidemiológico en el que los estudios observacionales se analizan como si fueran un ECA².

Por su parte, se considera a los ECA el método de referencia (*gold standard*), gracias a que aportan el mayor nivel de evidencia estadística posible y representan el tipo de estudio ideal para responder a los interrogantes que pudieran surgir en la investigación clínica. Sin embargo, a pesar de todas sus ventajas, las circunstancias técnicas, éticas y económicas particulares de cada investigación y sus intervenciones no siempre permiten llevarlos a cabo.

APLICACIÓN DEL ANÁLISIS DE ALEATORIZACIÓN MENDELIANA: BASES CONCEPTUALES Y METODOLÓGICAS

A los estudios de aleatorización mendeliana (AM) se los considera un caso particular dentro de un tipo más amplio de métodos de estadística aplicada, denominados en conjunto variables instrumentales (VI). El origen de esta clase de análisis se encuentra en las ciencias sociales (particularmente la econometría), en las que se lo utiliza habitualmente para estimar el impacto de determinadas políticas o medidas sociales cuando no es posible realizar un diseño experimental para su estudio³.

Estos análisis se basan en la identificación de algún tipo de fenómeno natural variable (denominado de modo genérico VI o simplemente instrumento), que se utiliza en el análisis estadístico para el ajuste de los posibles factores de confusión que pudieran existir en la investigación.

Para que se pueda considerar a dichos fenómenos (variables o instrumentos) VI adecuadas, deben cumplir una serie de premisas. Una manera habitual de definir y explicar dichas condiciones es utilizando los llamados gráficos acíclicos dirigidos, comúnmente denominados DAG (del inglés *directed acyclic graphs*)⁴.

* Autor para correspondencia: China Basin Landing, 185 Berry Street, Lobby 5, Suite 5700, Campus Mail Box 0560, San Francisco, CA 94107, Estados Unidos.

Correo electrónico: swalter@hsph.harvard.edu (S. Walter).

Full English text available from: www.revespcardiol.org/en

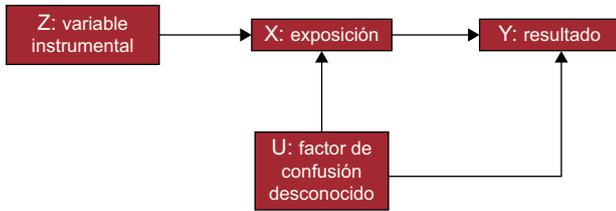


Figura 1. Representación gráfica de las características de una variable instrumental.

Estos gráficos permiten representar el fundamento de estos estudios mostrando las posibles relaciones causales de las variables consideradas en el estudio (figura 1).

En primer lugar, la variable debe tener una relación cuantificable con el factor de exposición de interés. Por otro lado, la variable no puede estar relacionada directamente con el resultado o desenlace de la intervención, excepto a través de su relación con la propia exposición. Y, por último, la VI debe ser independiente de los posibles factores de confusión desconocidos o no controlables que pueda haber entre la intervención y el resultado⁵. Si se cumplen estos requisitos, el análisis de VI permite comprobar si existe una relación de causalidad entre una exposición (tratamiento o intervención) (p. ej., un contaminante ambiental o un fármaco) y un resultado (desenlace) (p. ej., un aumento de la presión arterial o un infarto de miocardio).

Como ya se ha mencionado, los estudios de AM son un tipo particular de análisis dentro de las VI. Lo que caracteriza estos estudios es la utilización de variantes génicas como VI para el análisis, es decir, en este caso, el fenómeno natural variable elegido corresponde a un determinado polimorfismo génico que esté relacionado con algún factor de exposición, candidato de ser causa de alguna enfermedad o desenlace que se quiera analizar. Esta metodología se inspira y se fundamenta en la segunda ley de

Mendel (ley de segregación de los caracteres), que indica que, salvo excepciones (p. ej., presencia de ligamiento génico o estratificación en la población), el reparto de los genes parentales entre los gametos sucede de manera aleatoria gracias a la división meiótica que sufren las células germinales durante la gametogénesis¹. Es decir, la transferencia génica de padres a hijos es un suceso al azar que puede asemejarse a la asignación aleatoria en los diferentes grupos experimentales que se hace en un ECA. Esta circunstancia ha permitido que los estudios de AM se equiparen a los ECA¹, con la diferencia de que a los individuos participantes en el estudio de AM se los asigna al azar a diferentes genotipos, en lugar de a los distintos grupos del ensayo clínico (figura 2).

Como muestra la figura 2, los estudios de AM representan una alternativa no experimental a los ECA. En este caso, la asignación aleatoria grupal de los individuos no depende del investigador, sino que son los propios genes heredados de manera independiente los que sirven para realizar una aproximación estadística, es decir, como se muestra en la figura 1, el instrumento (Z), representa la asignación al tratamiento, X corresponde al tratamiento e Y, al resultado del ensayo (figuras 1 y 2).

Tal y como ya se mencionó, estos análisis resultan de especial utilidad en los estudios epidemiológicos observacionales, donde permiten inferir causalidad a pesar de la existencia de problemas con factores de confusión desconocidos o no medibles. Por otro lado, también se ha comprobado su utilidad al analizar los datos provenientes de un ECA. Estos no presentan las limitaciones propias de los estudios observacionales, gracias a que el proceso de aleatorización hace que dichos factores de confusión se repartan por igual entre los grupos experimentales. Sin embargo, estos estudios no están exentos de complicaciones, ya que a menudo los voluntarios del ensayo presentan problemas de adherencia o incumplimiento del tratamiento estudiado. En estas circunstancias, se recomienda el análisis de los datos por intención de tratar (*intention to treat* [ITT]) que, pese a ser la opción ideal, puede

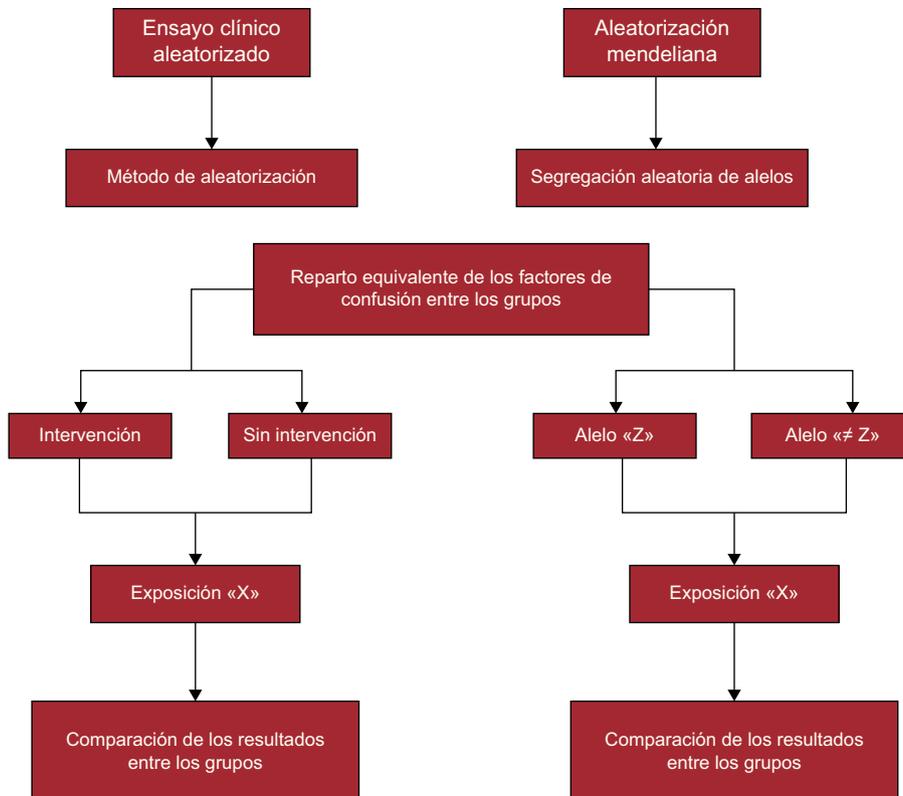


Figura 2. Esquema comparativo entre un ensayo clínico aleatorizado y un estudio de aleatorización mendeliana.

ofrecer resultados sesgados debido a la presencia de las ya mencionadas faltas de adherencia al tratamiento⁶.

En un análisis por ITT, se comparan los resultados obtenidos por los dos grupos del ensayo sin considerar si realmente todos los participantes incluidos en el análisis habían recibido el tratamiento. En esta situación, el análisis por ITT no estima el efecto que recibir el tratamiento (intervención) produce en el resultado, sino que realmente está estimando qué efecto en el resultado causa ser asignado a recibir el tratamiento. Ante esta situación, el análisis con VI ofrece la posibilidad de estimar verdaderamente de qué manera se ven afectadas por el tratamiento las personas que sí lo recibieron⁶.

OBTENCIÓN DEL ESTIMADOR DE VARIABLES INSTRUMENTALES

A lo largo de estos años se han desarrollado diversos métodos estadísticos con los que calcular los estimadores de las VI según las circunstancias particulares de cada análisis. Cuando se dispone de un instrumento binario (p. ej., un gen con dos únicas variantes alélicas), habitualmente se utiliza el llamado estimador de Wald⁶. Para obtenerlo, lo primero que se debe hacer es calcular la relación entre la VI (Z) y la variable resultado (Y) (figura 1), lo que en un ECA equivaldría a la estimación del análisis por ITT (si se está analizando un ECA, Z equivale, como se ha visto anteriormente, a la asignación aleatoria, por lo que la relación entre Z e Y corresponde a la estimación del análisis por ITT). Dicha relación permite comprobar si realmente existe una relación causal entre la variable exposición (X) e Y. Si la exposición (X) no tiene efecto alguno en el resultado o desenlace (Y), Z e Y han de ser independientes, y viceversa, si X afecta a Y, Z e Y no son independientes. No obstante, la asociación entre Z e Y no será tan fuerte como la existente entre X y Z.

A la hora de estimar el efecto de X en Y, se debe tener en cuenta que Z no determina perfectamente a X; por ello el segundo paso de este método consiste en redimensionar el efecto de Z en Y, con base en el efecto de Z en X. De esta forma se obtiene el estimador de la VI, que corresponde al cociente entre la diferencia de Y según los distintos valores de Z y la diferencia de X según los valores de Z (ecuación 1):

$$\frac{E(Y = 1|Z = 1) - E(Y = 1|Z = 0)}{E(X = 1|Z = 1) - E(X = 1|Z = 0)}$$

Teniendo en cuenta que $E(Y = 1|Z = 1)$ se define como el valor medio de Y entre los sujetos del estudio asignados al tratamiento $Z = 1$, el estimador calculado será sensible a los abandonos o retiradas de pacientes del ensayo. Si todos los voluntarios cumplieran con su tratamiento asignado, la estimación por VI equivaldría a la estimación que se obtiene por el análisis por ITT; sin embargo, según aumenta el número de abandonos, dicha estimación queda sobredimensionada proporcionalmente a la gravedad de las pérdidas en el ensayo.

Por otro lado, hay situaciones donde se dispone de un instrumento con más de dos valores (p. ej., polimorfismos múltiples) o se quiere ajustar simultáneamente por otras covariables. En estas circunstancias se suele utilizar el estimador de mínimos cuadrados en dos etapas (*two stage least squares estimator*)⁶. Brevemente, este método se aplica de la siguiente manera. En primer lugar, se obtienen los valores ajustados de la variable exposición (X) realizando una regresión de dicha variable (X) sobre la VI (Z) y las variables de confusión (U) que se conozcan. De esta forma, se estiman los valores predichos de X en función del ajuste realizado con la regresión (ecuación 2 a). En segundo lugar, se lleva a cabo otra regresión, en esta ocasión de la variable desenlace (Y) sobre los valores predichos de X, que funcionan como variables independientes en el modelo de regresión (ecuación 2 b). En este modelo, el coeficiente del valor predicho de X (β) (ecuación 2 b) se interpreta

como la estimación (por medio de la VI) del efecto de la exposición en el resultado:

$$\begin{aligned} a) X &= \alpha_{10} + \alpha_{1k}K + \varepsilon \\ b) Y &= \beta_{10} + \beta_{11}X + \beta_{1k}K + \varepsilon \end{aligned}$$

Dado el carácter introductorio de este artículo, se recomienda la lectura de otros trabajos más específicos publicados, como fuente de información y ayuda para profundizar en los estudios de AM y la estimación de la razón de probabilidades (*odds ratio*) y el riesgo relativo, para el cálculo del poder estadístico, la validación de las premisas y el diseño de estudios de dos etapas⁵⁻⁹.

APORTACIONES DE LA ALEATORIZACIÓN MENDELIANA EN CARDIOLOGÍA: EJEMPLOS Y HALLAZGOS DESTACADOS

La investigación cardiovascular se ha beneficiado durante los últimos años de la aplicación de los métodos estadísticos de AM¹⁰. Este hecho ha permitido confirmar la existencia de relaciones causales entre diversos factores de riesgo y ciertas enfermedades cardíacas. En la mayoría de los casos, la asociación entre dichos factores y las enfermedades era bien conocida de antemano; sin embargo, la ya mencionada presencia de posibles sesgos (propios de los estudios observacionales) impedía la inferencia de mecanismos causales con seguridad.

Un buen ejemplo que ilustra de manera idónea los fundamentos de estos estudios se encuentra en un trabajo, recientemente publicado, sobre la relación entre el índice de masa corporal y ciertas cardiopatías¹¹.

Los investigadores de ese trabajo se preguntaron si existía una relación causal entre exceso de grasa corporal y aparición de diferentes enfermedades cardiovasculares. Siguiendo la estrategia de los estudios de AM, lo primero que hicieron fue identificar un polimorfismo génico (instrumento) relacionado con un factor de riesgo (exposición) que estuviera asociado con una enfermedad cardíaca. En este caso se optó por un polimorfismo de un solo nucleótido del gen *FTO* (gen relacionado con la obesidad y la materia grasa), el cual, como su propio nombre indica, había sido relacionado en numerosos estudios previos con la cantidad de grasa corporal y, por lo tanto, con el índice de masa corporal.

El siguiente paso consistió en obtener los datos necesarios para realizar el análisis estadístico. En concreto, en ese trabajo se obtuvo información sobre el genotipo y el índice de masa corporal de > 190.000 participantes, a través de diferentes estudios internacionales realizados por los grupos de investigación colaboradores. A continuación, se los dividió en dos grupos, según la variante del polimorfismo de un solo nucleótido de interés que presentara cada voluntario. Como se señala en la figura 2, este paso equivaldría a la aleatorización de los ECA.

Por último, se analizaron los datos obtenidos en el estudio comparando los resultados de cada grupo según presencia o ausencia de distintas enfermedades cardíacas (enfermedad cardíaca coronaria, insuficiencia cardíaca, accidente cerebrovascular hemorrágico, hipertensión, etc.), metabólicas (dislipemia, diabetes mellitus tipo 2 y síndrome metabólico) o según los valores de diversos factores cardiometabólicos (presión arterial, colesterol, triglicéridos, proteína C reactiva, interleucina 6, etc.).

Los resultados mostraron la existencia de una relación causal entre la obesidad y la insuficiencia cardíaca y confirmaron hallazgos previos obtenidos de estudios observacionales. El razonamiento en que se basa este estudio es el siguiente: si el efecto del polimorfismo elegido (instrumento) es únicamente aumentar la cantidad de grasa corporal (factor de riesgo), se puede deducir que la causa del aumento observado en la incidencia de insuficiencia cardíaca es el índice de masa corporal.

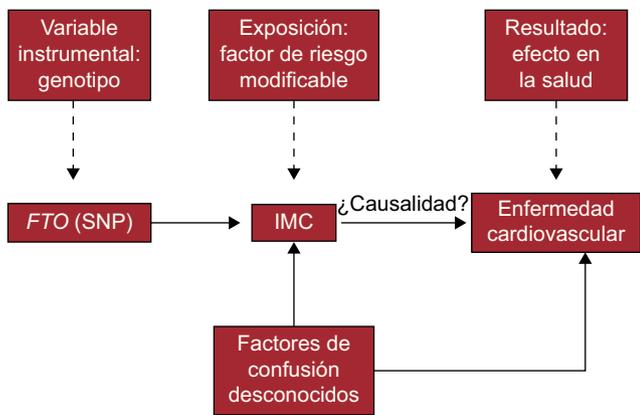


Figura 3. Resumen de los elementos particulares de un estudio de aleatorización mendeliana en cardiología. IMC: índice de masa corporal; FTO: *Fat mass and obesity-associated* SNP: polimorfismo de un solo nucleótido.

A modo de recapitulación, y basándose en el esquema general del análisis por VI que se ha mostrado anteriormente mediante un DAG (figura 1), se puede resumir las particularidades de este estudio en la figura 3.

Otros estudios de estas características han obtenido resultados igualmente relevantes. Cabe destacar el trabajo de Cohen et al¹², quienes demostraron los beneficios de la reducción persistente de la concentración de lipoproteínas de baja densidad en la disminución de infartos de miocardio. Otro ejemplo destacable se encuentra en el estudio de Chen et al¹³, que halló en varones relación entre el consumo elevado de alcohol y la hipertensión. En cambio, otros trabajos de investigación de este tipo muestran resultados opuestos a los obtenidos con métodos observacionales clásicos. Un estudio reciente¹⁴ que investigó la asociación entre la enzima sPLA₂ y la enfermedad cardiovascular no logró encontrar la supuesta relación, contrariamente a lo que habían indicado varios estudios tradicionales.

Merece la pena terminar este apartado mencionando que, como es lógico, dicho tipo de estudios también tiene ciertas limitaciones. Por un lado, se encuentran los problemas habituales que pueden aparecer en cualquier estudio genético de asociación, como la existencia de desviaciones en las frecuencias alélicas esperadas debido a diversas causas (estratificación de la población, existencia de fenómenos de desequilibrio de ligamiento, actividad pleotrópica de algunos genes, etc.) o problemas con el tamaño de la muestra; dado que este es inversamente proporcional al cuadrado de la correlación entre el instrumento genético y la exposición y este raramente excede del 5% para instrumentos genéticos, no siempre es sencillo obtener el número adecuado para estos estudios, que en el escenario más favorable se situaría al menos entre 5.000 y 100.000 pacientes¹⁵. Por otro lado, existen algunas limitaciones específicas del análisis por VI, tales como la dificultad para encontrar un instrumento adecuado (polimorfismos simples, múltiples o puntuaciones de riesgo genético [*genetic risk scores*]) para el estudio de interés, el incumplimiento de las condiciones que debe cumplir la VI o la falta de poder estadístico de esta¹⁶.

CONCLUSIONES

Los estudios de AM se presentan como una aproximación experimental esencial con que desentrañar los fundamentos etiológicos de numerosas enfermedades cardiovasculares. Gracias a esta metodología, actualmente se están realizando importantes avances en el estudio de las relaciones existentes entre diversas exposiciones modificables y enfermedades

cardiacas determinadas. La posibilidad de establecer verdaderas relaciones causales en estudios epidemiológicos observacionales donde se sospecha la existencia de factores de confusión no controlables ha generado una ola de optimismo dentro de la comunidad científica. Dicho entusiasmo se corrobora al observar el aumento exponencial del número de trabajos científicos basados en esta metodología publicados en los últimos años. Por otro lado, el desarrollo y avance continuos de las tecnologías asociadas a las «ómicas», junto con el perfeccionamiento de los métodos estadísticos actuales, sin duda facilitarán la superación de las limitaciones experimentales existentes y confirmarán a los estudios de AM como herramientas fundamentales para la investigación clínica y el desarrollo terapéutico. Estas circunstancias convertirán estos estudios en elementos esenciales de apoyo a los ECA a la hora de elaborar recomendaciones para la práctica clínica y en la implementación de políticas y medidas de salud pública.

Por último, merece la pena comentar que en los últimos años se viene asistiendo a la creación de importantes coaliciones internacionales, responsables del desarrollo y la coordinación de grandes centros de recursos biológicos. Estos organismos, a menudo llamados *biobancos*, representan el sueño de todo investigador, ya que se encargan de recoger, almacenar y procesar todo tipo de muestras biológicas humanas y sirven de repositorio de cuantiosa información relacionada con ellas¹⁷. Esta información incluye resultados de estudios de investigación basados en dichas muestras aplicables directamente al análisis por AM. Esto significa que no es necesario medir la exposición de interés en la misma población en que se observa el resultado. Es decir, siempre y cuando existan fuentes de información válidas sobre el efecto del gen en la exposición, se podrá vincular esta información directamente con el genotipo de la población estudiada (y con el resultado medido) para hacer inferencias causales utilizando el denominado diseño *split sample* de los estudios de AM, lo cual es especialmente relevante, puesto que supone un ahorro económico importantísimo porque se evita la necesidad de medir de nuevo el fenotipo estudiado de toda la muestra y solo se requiere obtener el genotipo⁹.

Estas circunstancias, junto con el inexorable progreso de la tecnología biomédica, sin duda garantizan un futuro prometedor a los estudios de AM.

CONFLICTO DE INTERESES

Ninguno.

BIBLIOGRAFÍA

- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32:1-22.
- Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19:766-79.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10:37-48.
- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Amer Statist Assoc.* 1996;91:444-55.
- Glymour MM, Tchetgen Tchetgen EJ, Robins JM. Credible mendelian randomization studies: approaches for evaluating the instrumental variable assumptions. *Am J Epidemiol.* 2012;175:332-9.
- Glymour MM. Natural experiments and instrumental variable analyses in social epidemiology. En: Oakes JM, Kaufman JS, editores. *Methods in social epidemiology.* San Francisco: John Wiley & Sons; 2006. p. 429-60.
- Tchetgen Tchetgen EJ, Walter S, Glymour MM. Building an evidence base for mendelian randomization studies: assessing the validity and strength of proposed genetic instrumental variables. *Int J Epidemiol.* 2013;42:328-31.
- Pierce BL, Burgess S. Efficient design for mendelian randomization studies: Subsample and 2-sample instrumental variable estimators. *Am J Epidemiol.* 2013;178:1177-84.

9. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in mendelian randomization analyses. *Am J Epidemiol.* 2011;173:1392-403.
10. Galve E, Alegría E, Cordero A, Fácila L, Fernández de Bobadilla J, Lluís-Ganella C, et al. Temas de actualidad en cardiología: riesgo vascular y rehabilitación cardiaca. *Rev Esp Cardiol.* 2014;67:203-10.
11. Holmes MV, Lange LA, Palmer T, Lanktree MB, North KE, Almoguera B, et al. Causal effects of body mass index on cardiometabolic traits and events: a Mendelian randomization analysis. *Am J Hum Genet.* 2014;94:198-208.
12. Cohen JC, Boerwinkle E, Mosley Jr TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med.* 2006;354:1264-72.
13. Chen L, Smith GD, Harbord RM, Lewis SJ. Alcohol intake and blood pressure: a systematic review implementing a Mendelian randomization approach. *PLoS Med.* 2008;5:e52.
14. Holmes MV, Simon T, Exeter HJ, Folkersen L, Asselbergs FW, Guardiola M, et al. Secretory phospholipase A₂-IIA and cardiovascular disease: a Mendelian randomization study. *J Am Coll Cardiol.* 2013;62:1966-76.
15. Freeman G, Cowling BJ, Schooling CM. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *Int J Epidemiol.* 2013;42:1157-63.
16. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* 2008;27:1133-63.
17. Zielhuis G. Biobanking for epidemiology. *Public Health.* 2012;126:214-6.