

## Nota metodológica

## Alcances de la aleatorización mendeliana para el control de confusores no observables en epidemiología

Mónica Ancira-Moreno<sup>a</sup>, Natalia Smith<sup>a</sup> y Héctor Lamadrid-Figueroa<sup>b,\*</sup><sup>a</sup> Departamento de Salud, Universidad Iberoamericana, Ciudad de México, México<sup>b</sup> Dirección de Salud Reproductiva, Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública, Cuernavaca, Morelos, México

## INFORMACIÓN DEL ARTÍCULO

## Historia del artículo:

Recibido el 7 de diciembre de 2020

Aceptado el 29 de enero de 2021

On-line el 26 de marzo de 2021

## Palabras clave:

Aleatorización mendeliana

Métodos epidemiológicos

Sesgo

Confusión

Genética

Variables instrumentales

## RESUMEN

La aleatorización mendeliana es un método epidemiológico propuesto para controlar las asociaciones espurias en los estudios observacionales. Estas asociaciones suelen estar causadas por la confusión derivada de factores sociales, ambientales y de comportamiento, que pueden ser difíciles de medir. La aleatorización mendeliana se basa en la selección de variantes genéticas que se utilizan como variables instrumentales, y que influyen en los patrones de exposición o están asociadas con un fenotipo intermedio de la enfermedad. El presente trabajo pretende discutir cómo seleccionar las variantes genéticas adecuadas como variables instrumentales y presentar las potenciales herramientas metodológicas para lidiar con las limitaciones propias de este método epidemiológico. El uso de variables instrumentales para exposiciones modificables tiene el potencial de atenuar los efectos de limitaciones comunes, como la confusión, cuando se eligen variantes genéticas robustas como variables instrumentales.

© 2021 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Scope of Mendelian randomization to control for unobserved confounding in epidemiology

## ABSTRACT

The Mendelian randomization is an epidemiologic method proposed to control for spurious associations in observational studies. These associations are commonly caused by confusion derived from social, environmental, and behavioral factors, which can be difficult to measure. Mendelian randomization is based on the selection of genetic variants that are used as instrumental variables that influence exposure patterns or are associated with an intermediate phenotype of the disease. The present work aims to discuss how to select the appropriate genetic variants as instrumental variables and to present methodological tools to deal with the limitations of this epidemiological method. The use of instrumental variables for modifiable exposures has the potential to mitigate the effects of common limitations, such as confusion, when robust genetic variants are chosen as instrumental variables.

© 2021 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Keywords:

Mendelian randomization

Epidemiologic methods

Bias

Confounding

Genetics

Instrumental variables

## Introducción

Es conocido que un ensayo clínico aleatorizado puede proveer evidencia rigurosa de causalidad debido a su aleatorización y ocultamiento de la asignación; sin embargo, para muchos tipos de exposición es muy costoso, poco práctico o no ético. Cuando un ensayo clínico aleatorizado no es plausible, la aleatorización mendeliana puede ser una alternativa para controlar por confusores en los estudios observacionales<sup>1</sup>; por ello, el objetivo de esta nota metodológica es discutir la fundamentación, los supuestos, los alcances y las limitaciones relacionadas con este abordaje metodológico.

## Fundamentos y supuestos de la aleatorización mendeliana

La aleatorización mendeliana se basa en la segunda y tercera leyes de Mendel (segregación y asignación independiente), las cuales establecen que los dos alelos de cada gen se separan entre sí y de forma independiente durante la gametogénesis, de modo que cada padre solo puede transmitir un alelo, además de que los alelos se distribuyen independientemente en la siguiente generación<sup>2</sup>.

Las variantes genéticas pueden utilizarse como variables instrumentales (factores que están correlacionados con la exposición, pero son externos a la relación exposición-desenlace) debido a que, al ser asignadas aleatoriamente durante la gametogénesis, aseguran la dirección de causalidad, y al ser heredadas de manera independiente pueden simular la asignación aleatoria realizada en los ensayos clínicos aleatorizados<sup>3</sup>.

La aleatorización mendeliana puede permitir evaluar las asociaciones causales a partir de análisis monogénicos y poligénicos; en los primeros, es posible evaluar la naturaleza causal de los factores

\* Autor para correspondencia.

Correo electrónico: [hlamadrid@insp.mx](mailto:hlamadrid@insp.mx) (H. Lamadrid-Figueroa).

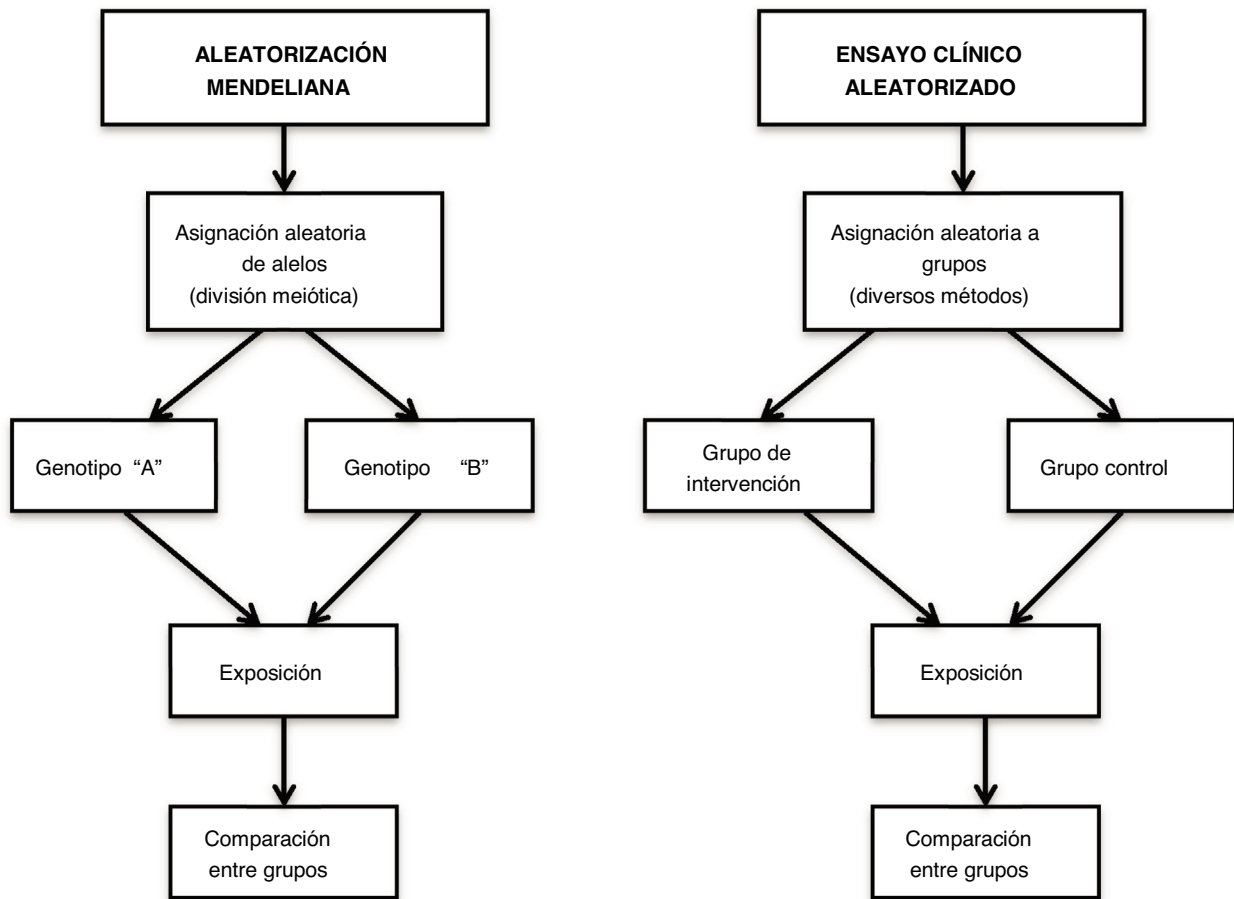


Figura 1. Comparación entre un estudio de aleatorización mendeliana y un ensayo clínico aleatorizado.

de riesgo utilizando variantes de una sola región genética, mientras que en los segundos se incluyen variantes genéticas de múltiples regiones<sup>1</sup>.

Una vez comprobado qué variante genética puede utilizarse como variable instrumental genética, la población a estudiar se divide en subgrupos según la presencia o la ausencia de la variante genética de riesgo o de protección. Estos subgrupos definidos genéticamente pueden considerarse como análogos a las ramas de un ensayo clínico aleatorizado: los individuos con genotipos de riesgo/protección son análogos a los incluidos en un grupo de intervención, y los individuos sin este genotipo son análogos a los de un grupo de control<sup>4,5</sup> (fig. 1).

### Bases de datos y análisis

La publicación de bases de datos de libre acceso por parte de los grandes grupos de investigación, como la *Genome Aggregation Database* (gnomAD)<sup>6</sup>, y el desarrollo de paquetes específicos de análisis de aleatorización mendeliana (<http://www.mrbase.org>) han facilitado la implementación de estos estudios.

Los tres supuestos sobre la variable instrumental genética que se utilizará para obtener una estimación no sesgada del efecto causal son: 1) debe ser reproducible y estar fuertemente asociada con la exposición, 2) debe ser independiente de los factores de confusión que sesgan la asociación exposición-desenlace, y 3) solo debe asociarse con el desenlace a través de la exposición estudiada<sup>3,4</sup>.

Para probar el primer supuesto debe existir evidencia científica previa que haya reportado la asociación en estudio; para el segundo, es necesario analizar las asociaciones entre la variable instrumental

genética y los factores de confusión observados cuando estén disponibles; y para el tercero, debe evaluarse la probabilidad de que la variante afecte el desenlace de otra manera que no sea a través de la exposición<sup>4</sup>.

### Amenazas y limitaciones

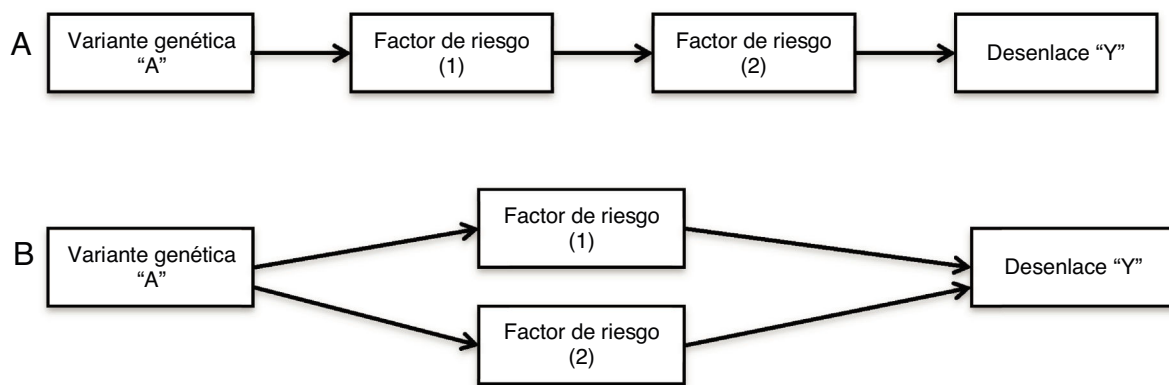
Existen amenazas específicas que pueden conducir a la violación de los tres supuestos básicos<sup>4</sup>; sin embargo, cada amenaza debe considerarse en el contexto del campo de investigación específico para encontrar el método más adecuado que debe aplicarse.

#### Variable instrumental genética débil

Se considera como una «variable instrumental genética débil»<sup>4</sup> si en la regresión de la exposición en la variable genética se identifica un valor del estadístico  $F < 10$ . Una asociación débil entre la variable instrumental genética y la exposición puede ser engañosa, por lo que la solución propuesta para este escenario es utilizar múltiples variables instrumentales genéticas o diseñar un instrumento de puntuación combinando múltiples variables instrumentales genéticas, como el puntaje alélico o puntaje de riesgo genético (ponderado o no ponderado)<sup>3,4</sup>.

#### Estratificación de la población

La estratificación de la población puede resultar en confusión de la asociación en virtud de diferencias étnicas, pues la frecuencia con que ocurren el alelo y la enfermedad, o las tasas de exposición,



**Figura 2.** Pleiotropía vertical y pleiotropía funcional. **A:** En la pleiotropía vertical, la variante genética “A” se asocia con un factor de riesgo (1), la variante genética “A” junto con el factor de riesgo (1) se asocia a otro factor de riesgo (2), y en conjunto llevan al desenlace “Y”. **B:** En la pleiotropía funcional, la variante genética “A” se asocia con dos factores de riesgo (1 y 2) simultáneamente, y en conjunto la variante genética “A”, el factor de riesgo (1) y el factor de riesgo (2) llevan al desenlace “Y”.

pueden variar entre diferentes subgrupos poblacionales<sup>4</sup>. Por ejemplo, en un estudio de casi 5000 nativos americanos Pima y de la tribu Papago se encontró una fuerte asociación inversa entre el haplotipo HLA Gm Gm3 y la diabetes tipo 2 (*odds ratio*: 0.27); no obstante, cuando se llevaron a cabo los análisis por estratos étnicos no se encontró asociación<sup>7</sup>. Por esta razón, es recomendable realizar análisis estratificados en poblaciones heterogéneas, incluso en poblaciones con ascendencia similar<sup>4</sup>.

#### Desequilibrio de ligamiento

El desequilibrio de ligamiento ocurre cuando existe una asociación no aleatoria entre diferentes variantes genéticas que están ubicadas cerca en el mismo cromosoma, y que generalmente se heredan juntas. El desequilibrio de ligamiento suele afectar solo a variantes genéticas cercanas y, por lo tanto, rara vez es motivo de preocupación cuando solo se utiliza una variante genética por gen. Una solución puede ser seleccionar únicamente variantes independientes como instrumentos o ajustar las otras variantes genéticas en el análisis de aleatorización mendeliana<sup>4</sup>.

#### Canalización

Es un proceso mediante el cual la asociación genotipo-enfermedad esperada es amortiguada por variantes genéticas adaptativas. Este proceso es difícil de estudiar en la práctica por el hecho de que el organismo siempre trata de mantener la homeostasis. Se podría asumir que existen muchos genes y vías metabólicas que, a su vez, podrían estar activados en diferente magnitud, con el fin de contrarrestar los efectos negativos de la variable instrumental genética empleada en la aleatorización mendeliana<sup>4</sup>.

#### Pleiotropía

En la práctica, muchas variantes genéticas son pleiotrópicas, es decir, se asocian con múltiples factores de riesgo<sup>4</sup>. Los posibles mecanismos por los cuales una variante genética puede estar asociada con múltiples factores de riesgo se dividen en pleiotropía vertical y pleiotropía funcional (fig. 2). Estos dos casos no se excluyen mutuamente, pues es posible que los dos estén presentes en la misma variante<sup>8</sup>. Con el fin de interpretar las estimaciones del efecto de causalidad, se asume el efecto de cada uno de los factores de riesgo como «causalmente independientes»<sup>8</sup>.

En la actualidad existen métodos de análisis robustos para detectar y ajustar la pleiotropía cuando se utilizan múltiples variables instrumentales genéticas<sup>9</sup>. El análisis de dos muestras se ha vuelto atractivo debido a que aumenta el tamaño muestral, gana

poder estadístico, estima asociaciones causales basadas en diferentes variantes genéticas y evalúa los supuestos básicos de la aleatorización mendeliana<sup>10</sup>. Los métodos más utilizados son el MR-Egger, para el control de sesgos introducidos por las variantes genéticas, y el MR-PRESSO, que elimina las múltiples variantes genéticas del análisis cuya estimación causal específica de la variante difiere sustancialmente de las de otras variantes<sup>9</sup>.

#### Conclusiones

La aleatorización mendeliana es un método de análisis que permite probar hipótesis de asociaciones causales en estudios observacionales mediante el uso de variantes genéticas como variables instrumentales; no obstante, para que el abordaje sea válido, las variables instrumentales genéticas deben cumplir con los supuestos básicos, además de considerar las posibles amenazas a la violación de estos.

Esta nota metodológica pretende permitir al lector identificar que la aleatorización mendeliana es una alternativa metodológica viable que facilita lidiar con la confusión cuando se eligen variantes genéticas robustas como variables instrumentales.

#### Editor responsable del artículo

Carlos Álvarez Dardet.

#### Contribuciones de autoría

Todas las personas firmantes participaron en la elaboración de la nota metodológica. M. Ancira y H. Lamadrid-Figueroa escribieron la primera versión, que fue revisada y editada por N. Smith. Todas las personas firmantes han leído y aprobado la versión final antes de ser enviada para su publicación.

#### Financiación

Ninguna.

#### Conflictos de intereses

Ninguno.

#### Bibliografía

- Burgess S, Foley CN, Zuber V. Inferring causal relationships between risk factors and outcomes from genome-wide association study data. *Annu Rev Genomics Hum Genet.* 2018;19:303–27.

2. Lewis RG, Simpson B. Genetics, autosomal dominant. StatPearls Treasure Island. 2020.
3. Quentin Langdon RJ, Hazel Wade K. Application of mendelian randomization: can we establish causal risk factors for type 2 diabetes in low-to-middle income countries. *Rev Cuid.* 2017;**8**:1391–406.
4. Sekula P, Del Greco MF, Pattaro C, et al. Mendelian randomization as an approach to assess causality using observational data. *JASN.* 2016;**27**:3253–65.
5. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol.* 2013;**42**:1134–44.
6. Koch L. Exploring human genomic diversity with gnomAD. *Nat Rev Genet.* 2020;**21**:448.
7. Lawlor DA, Harbord RM, Sterne JAC, et al. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* 2008;**27**:1133–63.
8. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* 2015;**181**:251–60.
9. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Research.* 2020;**4**:186.
10. García-Santisteban I, Cilleros-Portet A, Moyua-Ormazabal E, et al. A two-sample Mendelian randomization analysis investigates associations between gut microbiota and celiac disease. *Nutrients.* 2020;**12**:1420.