

Conocimiento experto, confusión y métodos causales

M.A. Hernán

Harvard School of Public Health. Boston. EE.UU.

Correspondencia: Dr. M.A. Hernán. Harvard School of Public Health. Program on Causal Inference in Epidemiology and Allied Sciences. 677 Huntington Avenue. Boston, MA 02115, EE.UU. Correo electrónico: miguel_hernan@post.harvard.edu

Este trabajo está parcialmente basado en un comentario escrito por el autor en *Statistics in Epidemiology Report* (2001).

(Expert knowledge, confounding and causal methods)

Resumen

La ausencia de confusión residual es la condición fundamental para la inferencia causal a partir de datos observacionales. Incluso cuando esta condición se cumple y los modelos están correctamente especificados, las estimaciones de los métodos estadísticos tradicionales pueden no tener una interpretación causal si existen confusores cambiantes que están afectados por la exposición. Los métodos «causales», como los modelos estructurales marginales y los modelos estructurales anidados, evitan este problema.

Palabras clave: Confusión. Inferencia causal. Modelos estadísticos.

Abstract

The absence of unmeasured confounding is the fundamental condition for causal inference from observational data. Even when this condition holds and the models are correctly specified, estimates from standard methods may not have a causal interpretation if there are time-dependent confounders that are affected by prior exposure. «Causal» methods, such as marginal structural models and structural nested models, avoid this problem.

Key words: Confounding. Causal inference. Statistical models.

Introducción

Un objetivo frecuente de la investigación epidemiológica es realizar inferencias causales acerca del efecto de una exposición o tratamiento sobre el riesgo de una enfermedad. Para conseguir este objetivo, los epidemiólogos diseñan estudios, recogen datos, los analizan usando métodos estadísticos y estiman una medida de asociación entre exposición y enfermedad. En principio, cualquier medida de asociación (p. ej., un coeficiente de regresión) puede interpretarse de forma causal. Pero la justificación para interpretar causalmente los resultados de un método estadístico depende de consideraciones extraestadísticas. Se trata de un asunto de conocimiento experto *a priori*, es decir, del conjunto de creencias, fundamentadas en la evidencia disponible, que posee el investigador antes de realizar el estudio.

Un ejemplo simplificado

Consideremos dos estudios para estudiar el efecto de la terapia antirretroviral sobre la incidencia de sida

entre pacientes infectados por el VIH. El primer estudio es un ensayo clínico en el que el investigador asigna a los pacientes, de forma aleatoria, antirretrovirales o placebo durante 5 años y determina cuándo/si reciben un diagnóstico de sida durante este tiempo. Supongamos que se trata de un ensayo clínico ideal: gran número de pacientes en cada grupo de tratamiento, adhesión perfecta al tratamiento asignado y seguimiento completo de todos los pacientes hasta el final del experimento. En el segundo estudio, el investigador se limita a observar el tratamiento que reciben los pacientes durante 5 años. Los médicos suelen asignar tratamiento antirretroviral a los pacientes más inmunodeprimidos, es decir, los de valores de CD4 más bajos y carga viral más alta. Como en el estudio previo, supongamos que se tiene gran número de pacientes, adhesión perfecta y seguimiento completo. Los datos recogidos en ambos estudios para cada paciente son la fecha de incorporación al estudio, la posible fecha de inicio de tratamiento antirretroviral y la posible fecha de diagnóstico de sida. Los datos de cada estudio se analizan usando un modelo de supervivencia de Cox correctamente especificado con dosis acumulada de tratamiento como la única covariable. La medida de asociación que se obtiene es la incidencia relativa (*hazard*

ratio) de sida por unidad de tratamiento. Seguidamente se proporcionan los datos recogidos en cada estudio y las correspondientes incidencias relativas a un grupo de epidemiólogos.

Confusión

La mayoría de los epidemiólogos estarían de acuerdo en que la incidencia relativa estimada por este modelo puede interpretarse como el efecto causal —medido en la escala de incidencia relativa— del tratamiento en el primer estudio, pero no en el segundo. La justificación de estas conclusiones causales no se basa en argumentos estadísticos. Decimos que la primera incidencia relativa puede interpretarse causalmente porque el tratamiento fue asignado de manera aleatoria y porque sabemos que la aleatorización en muestras grandes garantiza que ambos grupos de tratamiento son comparables o intercambiables¹ (*exchangeable*). En otras palabras, no existe confusión. Decimos que la segunda incidencia relativa no puede interpretarse causalmente porque nuestro conocimiento *a priori* nos hace sospechar que existen factores pronósticos que también influyen en la decisión de asignar la nueva terapia (p. ej., el valor de CD4 y la carga viral) y que no se han considerado en el análisis. Estos factores son marcadores del valor de inmunodepresión, una causa común de tratamiento y enfermedad que produciría una asociación —una incidencia relativa distinta de uno— incluso si el tratamiento antirretroviral no tuviera efecto causal alguno sobre la incidencia de sida. Si en el análisis estadístico no se ajusta por estas causas comunes o por sus marcadores, los pacientes en cada grupo de tratamiento no son intercambiables: los tratados con antirretrovirales están, como media, más enfermos que los tratados con placebo. En otras palabras, existe confusión.

En ambos estudios se usa el mismo método estadístico, pero las conclusiones respecto a la interpretación causal de la incidencia relativa son distintas. Esto ocurre porque la información no estadística utilizada por el usuario del método es distinta. En general, el parámetro estimado por un determinado método estadístico puede recibir una interpretación causal cuando todos los confusores (causas comunes o sus marcadores) han sido adecuadamente medidos (es decir, no existe confusión residual) y ajustados en el análisis. Es importante enfatizar que la existencia de confusión residual no puede detectarse mediante el análisis de los datos recogidos. Por definición, la confusión residual se debe a factores no medidos y, por tanto, su identificación se debe enteramente al conocimiento experto *a priori* del investigador. En el estudio observacional que nos sirve de ejemplo, los epidemiólogos utilizan su conocimiento experto sobre sida y el diseño del estudio, pero no

la información en la base de datos que se les proporcionó, para sospechar que importantes factores de confusión no se han medido. Precisamente debido a que la condición de no confusión residual no puede comprobarse, la inferencia causal a partir de datos observacionales —o de experimentos aleatorizados con tamaños muestrales pequeños, adhesión imperfecta o seguimiento incompleto— es siempre arriesgada. Los epidemiólogos deben ser expertos en el área que estudian o trabajar con ellos para asegurarse la identificación y medición de los principales confusores conocidos. Podrán así acercarse al cumplimiento aproximado de la condición de no confusión residual.

Métodos tradicionales frente a métodos causales

Llamamos «tradicionales» a los métodos basados en estimar el parámetro de interés en cada estrato definido por los factores de confusión. Por ejemplo, el análisis estratificado de Mantel-Haenszel, los modelos de regresión (lineal, logística, de Poisson, de Cox, etc.), el emparejamiento por *propensity scores* y otros.

Si la validez causal de las estimaciones estadísticas requiere el cumplimiento, al menos aproximado, de la condición de no confusión residual, ¿por qué se usa entonces el calificativo de «causales» para ciertos métodos estadísticos? Podría pensarse que, una vez que se acepta que la información recogida es suficiente para eliminar la confusión, se podría aplicar cualquier método estadístico para analizar los datos. Precisamente éste es el problema: las estimaciones de los métodos estadísticos tradicionales pueden no tener una interpretación causal incluso si se cumple la condición de no confusión residual. En concreto, esto ocurre en estudios longitudinales cuando el investigador está interesado en el efecto causal de una exposición cambiante (p. ej., tratamiento antirretroviral) y existen confusores cambiantes (p. ej., valor de CD4) que están afectados por la exposición. Los métodos «causales» eliminan este problema y garantizan que todos los confusores, identificados y medidos con tanto esmero por el epidemiólogo, serán ajustados apropiadamente.

Efecto causal y variables contrafactuales

Los métodos causales y la definición de efecto causal que emplean se basan en la teoría de contrafactuales. Veamos primero la definición de efecto causal en situaciones simples como un ensayo clínico como el descrito anteriormente, en el que la exposición no es cambiante (cada paciente está siempre tratado con antirretrovirales o con placebo). Decimos que el grupo

de tratamiento A (1 = antirretroviral, 0 = placebo) tiene un efecto causal sobre la carga viral Y_i del paciente i medida al final del seguimiento si su carga viral en caso de que se le hubiese sometido a tratamiento antirretroviral ($Y_{a=1,i}$) fuera diferente de su carga viral en caso de que se hubiese sometido a tratamiento con placebo ($Y_{a=0,i}$). Es decir, el grupo de tratamiento afecta causalmente a la carga viral del paciente i si $Y_{a=1,i} \neq Y_{a=0,i}$. Por supuesto, cada paciente sigue sólo un curso de tratamiento (antirretrovirales o placebo) y, por tanto, sólo podemos observar $Y_{a=1,i}$ o $Y_{a=0,i}$. Si, por ejemplo, el paciente i recibió tratamiento antirretroviral, entonces el valor de su carga viral bajo tratamiento es precisamente su valor observado de carga viral, $Y_{a=1,i} = Y_i$ mientras que el valor de su carga viral bajo placebo ($Y_{a=0,i}$) permanece desconocido (puesto que no recibió placebo). En general, las variables $Y_{a=1,i}$ e $Y_{a=0,i}$ representan situaciones contrarias a los hechos (p. ej., el valor de carga viral bajo tratamiento antirretroviral cuando en realidad el paciente recibió placebo, o viceversa) y por eso se conocen como variables contrafactuales.

El desconocimiento del valor que toma parte de las variables contrafactuales es el problema fundamental de la inferencia causal e imposibilita la determinación de efectos causales para un individuo. Sin embargo, bajo el supuesto de no confusión residual, es posible realizar inferencias sobre el efecto causal medio entre los individuos de una población. Es decir, se puede determinar si la carga viral media en la población cuando se somete a todo el mundo a tratamiento antirretroviral, $E[Y_{a=1,i}]$ es diferente de la carga antirretroviral media cuando se somete a todo el mundo a tratamiento con placebo, $E[Y_{a=0,i}]$. Decimos que existe un efecto causal medio cuando $E[Y_{a=1,i}] \neq E[Y_{a=0,i}]$. Para generalizar la definición de efecto causal a exposiciones cambiantes se emplean variables contrafactuales bajo distintas historias de exposición (p. ej., siempre expuesto, nunca expuesto, etc.).

Las variables contrafactuales o potenciales fueron introducidas en métodos estadísticos por Neyman² para experimentos aleatorizados con tratamientos no cambiantes. Rubin^{3,4} extendió la teoría contrafactual a estudios observacionales con exposiciones no cambiantes. Robins^{5,6} desarrolló una teoría contrafactual formal que es una generalización de las anteriores y que puede aplicarse a estudios longitudinales con exposiciones cambiantes. Este avance es clave porque la mayor parte de las exposiciones de interés en epidemiología son cambiantes. De hecho, los métodos causales actualmente disponibles se basan en la teoría de Robins.

Métodos de Robins

El primer producto de la teoría de Robins fue la fórmula g, un método causal no paramétrico para calcu-

lar efectos causales de exposiciones cambiantes bajo la condición de no confusión residual, incluso en presencia de confusores cambiantes afectados por la exposición. La fórmula g es una generalización del ajuste por estandarización, en contraste con los métodos tradicionales que ajustan mediante estratificación de la medida de asociación. Una posible aplicación de la fórmula g es el cálculo de las medias contrafactuales $E[Y_{a=1,i}]$ y $E[Y_{a=0,i}]$. Si la variable Y se codifica como 0 cuando el paciente sobrevive y como 1 cuando muere, entonces $E[Y_{a=1,i}]$ es el riesgo de mortalidad si toda la población fuera expuesta y $E[Y_{a=0,i}]$ el riesgo de muerte si toda la población permaneciera no expuesta. Una vez calculados estos riesgos usando la fórmula g sólo resta compararlos para decidir si existe un efecto causal medio de la exposición A sobre la respuesta Y . Esta comparación puede llevarse a cabo usando razones, diferencias u *odds ratio*. Por ejemplo, el riesgo relativo causal es $E[Y_{a=1,i}] / E[Y_{a=0,i}]$. Recientemente, investigadores que trabajaban en inteligencia artificial y ciencias de la computación han redescubierto independientemente la fórmula g⁷.

A pesar de su interés teórico, la fórmula g no suele utilizarse en la práctica porque es un método no paramétrico, es decir, no impone restricciones paramétricas sobre la forma de la relación entre las variables. Así, una aplicación directa de la fórmula g en estudios longitudinales, incluso con un número moderado de medidas repetidas y variables, requeriría cantidades enormes de datos y tiempo de cálculo. Además, la carencia de parámetros hace que sea laborioso realizar un test de la hipótesis nula. Por otro lado, la solución no puede consistir en usar métodos tradicionales totalmente paramétricos que, como se ha afirmado anteriormente, pueden llevar a estimadores sesgados del efecto causal.

Para resolver estos problemas, Robins desarrolló dos clases de modelos semiparamétricos que incorporan sólo algunas restricciones paramétricas y que permiten una fácil representación de la hipótesis nula: los modelos estructurales marginales^{8,9} y los estructurales anidados^{10,11}. La palabra *estructural* en ciencias sociales se emplea a menudo como sinónimo de *causal*. Los parámetros de estos modelos se interpretan como el efecto causal de la exposición (en distintas escalas) siempre bajo las condiciones de no confusión residual y correcta especificación del modelo.

Modelos estructurales marginales

Un ejemplo de modelo tradicional de Cox es el modelo descrito en la segunda sección de este artículo. Como todos los modelos tradicionales, éste modela un aspecto de la distribución condicional de la variable dependiente observada: la incidencia observada de muer-

te dada la dosis acumulada de tratamiento. El problema de este modelo es que la incidencia relativa estimada puede no tener una interpretación causal válida incluso si se incluyeran los confusores como covariables. La razón es que los confusores (CD4, carga viral, etc.) están afectados por el tratamiento y, por tanto, su inclusión en el modelo sesga los resultados.

Un ejemplo de modelo estructural marginal de Cox es un modelo para la incidencia contrafactual de muerte si todos los individuos en el estudio hubieran seguido el mismo régimen de tratamiento. Por ejemplo, la incidencia relativa que compara la incidencia si todo el mundo hubiera sido tratado con la incidencia si nadie lo hubiera sido tiene, por definición, una interpretación causal válida. Existen modelos marginales estructurales lineales, logísticos, de medidas repetidas, de Cox, etc. La variable dependiente de los modelos estructurales marginales es siempre algún aspecto de la distribución marginal de una variable contrafactual bajo historias de exposición prefijadas.

Los parámetros de los modelos estructurales marginales se estiman usando una técnica conocida como «ponderación por el inverso de la probabilidad de tratamiento». Por ejemplo, si el objetivo es estimar la incidencia relativa causal descrita en el párrafo anterior mediante un modelo estructural marginal de Cox, el análisis comienza con una fase preliminar en la que se estima la probabilidad de que cada sujeto haya recibido el régimen de tratamiento que realmente recibió. Estas probabilidades pueden estimarse, por ejemplo, usando un modelo de regresión logística con los confusores como covariables. El inverso de esta probabilidad es la ponderación que cada individuo recibe. El análisis final consiste en ajustar un modelo de Cox tradicional, como el descrito en la segunda sección de este artículo, en el que cada sujeto recibe su peso correspondiente. Casi sorprendentemente este método proporciona un estimador consistente (válido) de la incidencia relativa causal.

Los modelos marginales estructurales no controlan la confusión mediante estratificación sino mediante ponderación, un concepto equivalente a estandarización. En otras palabras, los confusores se incluyen en el modelo preliminar que se usa para estimar las probabilidades cuyo inverso son los pesos, pero no en el modelo principal (de Cox en este caso), lo que evita la aparición de sesgos.

Por ejemplo, un modelo marginal estructural de Cox se ha aplicado para estimar el efecto causal sobre la mortalidad de tratamiento antirretroviral y profilaxis para enfermedades oportunistas en el Multicenter AIDS Cohort Study¹². En este caso existían confusores del efecto de los tratamientos sobre la mortalidad: CD4, CD8, parámetros hematológicos, síntomas (pérdida de peso, diarrea, fiebre, etc.), enfermedades oportunistas (neumonía por *Pneumocystis carinii*, etc.) e historia previa

de tratamiento. Muchos de estos confusores están a su vez afectados por el tratamiento previo (p. ej., el valor de CD4 es mayor en pacientes que reciben tratamiento antirretroviral y la incidencia de enfermedades oportunistas es menor en pacientes que reciben cualquiera de los tratamientos considerados). Después de ajustar adecuadamente por estos confusores, la mortalidad relativa (todos los pacientes tratados de forma continua frente a ningún paciente tratado) estimada fue de 0,67 (intervalo de confianza [IC] del 95%: 0,46-0,98) para tratamiento antirretroviral y de 1,14 (IC del 95%: 0,79-1,64) para profilaxis.

La principal ventaja de los modelos marginales estructurales (comparados con los modelos estructurales anidados descritos a continuación) es su facilidad de uso ya que su aplicación simplemente requiere la ponderación de modelos tradicionales. Otra ventaja importante es que pueden usarse para cualquier tipo de variable respuesta (continua, dicotómica, etc.). Entre sus inconvenientes destaca que no pueden usarse para estimar el efecto de la interacción de confusores cambiantes con el tratamiento de interés.

Modelos estructurales anidados

Los modelos estructurales anidados son modelos para variables contrafactuales bajo historias dinámicas de exposición, es decir, aquellas en que la exposición en un momento determinado puede depender tanto de la historia personal de covariables como de la historia de exposición previa. Existen modelos estructurales anidados para análisis de supervivencia y para respuestas continuas, pero no para respuestas dicotómicas. Los modelos estructurales anidados de supervivencia son una generalización del modelo de tiempo acelerado de fracaso (*accelerated failure time model*). Estos modelos no estiman la mortalidad relativa, sino el factor por el que se multiplica el tiempo de supervivencia de cada individuo cuando recibe el tratamiento para obtener su tiempo de supervivencia cuando no recibe el tratamiento.

Los parámetros de los modelos estructurales anidados se estiman mediante el método de estimación g. Este método utiliza explícitamente la información contenida en la condición de no confusión residual para estimar los parámetros de interés. Por ejemplo, un modelo estructural anidado se ha aplicado para estimar el efecto causal de la hipertensión sistólica aislada sobre la mortalidad cardiovascular en el Framingham Heart Study¹³. En este caso, la rigidez arterial (definida por la detección radiológica de placas calcificadas en la aorta) es un confusor para el efecto de la hipertensión sistólica aislada, pero a su vez la presencia de hipertensión sistólica puede contribuir a producir rigidez de la pared arterial. Los autores de este estudio encon-

traron que, tras ajustar de forma adecuada por rigidez arterial, la hipertensión sistólica aislada disminuyó el tiempo de supervivencia cardiovascular en un 45% (IC del 95%: 3-69).

En resumen, todos los métodos estadísticos que se usan para inferencia causal requieren que se cumpla, al menos aproximadamente, la condición de no confu-

sión residual, pero sólo los métodos causales, como los modelos estructurales marginales o anidados, permiten obtener resultados causalmente válidos cuando existen confusores cambiantes afectados por la exposición. La elección del método de análisis depende del conjunto de presupuestos que el investigador esté dispuesto a aceptar.

Bibliografía

1. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* 1986;15:413-9.
 2. Neyman J. On the application of probability theory to agricultural experiments: essay on principles, section 9 (1923). Traducido al inglés en *Stat Sci* 1990;5:465-80.
 3. Rubin DB. Bayesian inference for causal effects: the role of randomization. *Annals of Statistics* 1978;6:34-58.
 4. Holland PW. Statistics and causal inference [with discussion]. *J Am Statistical Assoc* 1986;81:945-70.
 5. Robins JM. A new approach to causal inference with a sustained exposure period. Application to the healthy worker survivor effect [erratas en *Math Modelling* 1987;14:917-21] *Math Modelling* 1986;7:1393-512.
 6. Robins JM. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chron Dis* 1987;40 (Suppl):139-61.
 7. Pearl J. Causal diagrams for empirical research. *Biometrika* 1995;82:669-710.
 8. Robins JM. *Marginal structural models*. Alexandria: 1997 Proceedings of the Section on Bayesian Statistical Science. American Statistical Association, 1998; p. 1-10.
 9. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550-60.
 10. Robins JM. Causal inference from complex longitudinal data. En: Berkane M, editor. *Latent variable modeling and applications to causality*. Lecture Notes in Statistics 120. New York: Springer-Verlag, 1997; p. 69-117.
 11. Robins JM. Structural nested failure time models. En: Andersen PK, Keiding N, editors. *Survival Analysis*. The Encyclopedia of Biostatistics. Chichester: John Wiley and Sons, 1998; p. 4372-89.
 12. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the joint effect of non randomized treatments. *J Am Stat Assoc* 2001;96:440-8.
 13. Wittelman JC, D'Agostino RB, Stijnen T, Kannel WR, Cobb JC, De Ridder MA, et al. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Am J Epidemiol* 1998;148:390-401.
-