

Causalidad en epidemiología (II): La causalidad contrafáctica. Planteamiento teórico

CAUSALITY IN EPIDEMIOLOGY (II): COUNTERFACTUAL CAUSALITY. THEORETICAL APPROACH

Mario RIVERA IZQUIERDO, Luis Miguel MARTÍN de los REYES, Pablo LARDELLI CLARET

Departamento de Medicina Preventiva y Salud Pública. Facultad de Medicina. Universidad de Granada. Avenida de la Investigación nº 11. CP 18071, Granada (España)

Correspondencia: Pablo Lardelli Claret. Correo-e: lardelli@ugr.es

1. CONCEPTOS GENERALES

En el artículo anterior se presentaron las dos formas clásicas de entender la causalidad: determinista y probabilística. Ahora vamos a introducirnos en una tercera alternativa, complementaria a las anteriores, de creciente expansión en epidemiología: el modelo de causalidad contrafáctica, aplicable a la identificación de asociaciones causales no sólo en el ámbito de los estudios experimentales con asignación aleatoria, sino de cualquier estudio analítico observacional, siempre que se den una serie de premisas. Este modelo de causalidad, también llamado de los *desenlaces potenciales*, es aplicable tanto en el marco del pensamiento determinista como del probabilístico. Para entender este modelo hay que deslindar dos formas de estimar asociaciones causales: las que se realizan a nivel individual y las que se realizan a nivel poblacional (estimaciones causales promedio).

1.1. El modelo causal contrafáctico a nivel individual

El planteamiento contrafáctico de la causalidad a nivel individual es muy sencillo, porque está basado en el que aplicamos constantemente en todos los razonamientos de causalidad que realizamos en nuestra vida cotidiana. Imaginemos que encima de una mesa hay una libreta y que, en un momento dado, empujamos horizontalmente la libreta hasta que rebasa el borde de la mesa y cae al suelo. Todos diremos, sin género de duda, que la causa de que la libreta haya caído al suelo es que nosotros la hemos empujado (le

hemos transmitido una energía cinética suficientemente intensa como para desplazarla hasta hacer que rebase el límite de la mesa). Si preguntáramos la razón de que sepamos que ésta y no otra ha sido la causa de que la libreta haya caído, la respuesta es simple: "porque si no la hubiéramos empujado, la libreta no se habría caído". Este es el razonamiento contrafáctico de causalidad individual: comparar los dos desenlaces de dos situaciones que sólo difieren en un único aspecto: en una se produce una intervención (empujar la libreta) que en la otra está ausente (no se empuja la libreta). Así pues, el modelo de causalidad contrafáctico se basa en la comparación de dos situaciones:

1. La libreta está encima de la mesa y nosotros la empujamos.
2. La libreta está encima de la mesa y nosotros no la empujamos.

Es importante tener en cuenta que, de las dos situaciones que se comparan, en realidad solo ha ocurrido una de ellas (la primera). La otra no ha ocurrido (es contrafáctica), pero eso no nos supone problema alguno, porque nos la imaginamos con suma facilidad ("si no la hubiéramos empujado, la libreta no se habría caído"). Esta facilidad con la que imaginamos los desenlaces de situaciones contrafácticas nos impide darnos cuenta de un hecho trascendental: toda asunción de causalidad requiere la estimación del desenlace de un suceso contrafáctico, que nunca, nunca, es directamente observable (por eso lo llamamos también desenlace potencial). Por lo tanto, la verificación de una hipótesis causal pasa ineludiblemente por un proceso de estimación de un dato faltante (el desenlace de un suceso contrafáctico).

Existe una notación matemática (sencilla) para describir los fenómenos implicados en el ejemplo anterior (y en cualquier otra hipótesis de causalidad que se nos pueda ocurrir), así como para definir y cuantificar la existencia de una asociación causal entre ellos:

- La libreta: Unidad Experimental.
- Empujar la libreta con suficiente fuerza como para que rebase el borde de la mesa: Intervención, Exposición, Tratamiento (T). En el caso más sencillo, T es una variable aleatoria con dos valores (0=no; 1=sí).
- Caída de la libreta al suelo: Desenlace potencial (Y). En el caso más sencillo, Y es una variable aleatoria con dos valores (0: no se cae; 1: se cae).

$Y(0)$: desenlace potencial cuando $T=0$ (lo que le pasa a la libreta cuando no se le empuja).

$Y(1)$: desenlace potencial cuando $T=1$ (lo que le pasa a la libreta cuando se le empuja).

Diremos que "empujar la libreta con suficiente fuerza como para que rebase el borde de la mesa" es causa de que la libreta caiga al suelo cuando $Y(1) > Y(0)$. En nuestro ejemplo:

$Y(0) = 0$ (la libreta no se cae si no la empujo con suficiente fuerza).

$Y(1) = 1$ (la libreta sí se cae cuando la empujo con suficiente fuerza).

Por lo tanto, en nuestro ejemplo podemos decir que empujar la libreta con suficiente fuerza es causa de que la libreta caiga al suelo porque se cumple que $Y(1) > Y(0)$. Además, podemos medir la magnitud de la asociación causal entre T e Y, por ejemplo, de la siguiente forma:

Magnitud de la asociación causal entre T e Y = $Y(1) - Y(0)$. En nuestro caso: $1 - 0 = 1$.

Evidentemente, en nuestra vida cotidiana no hay que recurrir a este armamento epidemiológico-matemático para saber que empujar la libreta ha sido la causa de que se caiga al suelo. Ahora bien, ¿qué pasa cuando tratamos de aplicar un razonamiento similar a otras situaciones, por ejemplo, fenómenos causales en el ámbito de la salud? Veamos otro ejemplo:

Unidad experimental: Juan (un varón al que le duele la cabeza).

T: tomarse un comprimido de 600 mg de ibuprofeno (0: Juan no se lo toma; 1: Juan sí se lo toma).

Y: Desaparición del dolor de cabeza media hora después (0: no desaparece; 1: sí desaparece).

Imaginemos que Juan realmente se toma el comprimido y a la media hora se le ha quitado el dolor. Eso quiere decir que $Y(1) = 1$.

¿El ibuprofeno ha sido la causa de que a Juan se le haya quitado el dolor de cabeza? No lo sabemos. ¿Por qué no lo sabemos? Porque, a diferencia de lo que ocurría con el ejemplo de la libreta y su caída, no sabemos, ni nos podemos imaginar con absoluta certeza, cuál es el desenlace de la situación contrafáctica con la que hemos de comparar la situación observada: ¿habría desaparecido el dolor de cabeza de Juan si éste no se hubiera tomado el comprimido de ibuprofeno? No se sabe:

$$Y(0) = ??????$$

Esta situación es la habitual en las hipótesis de causalidad que implican a la mayoría de los fenómenos relacionados con la salud: es muy difícil imaginar el desenlace de la situación contrafáctica, generalmente por dos razones, muy relacionadas entre sí:

- Entre la supuesta causa y el supuesto desenlace transcurre un tiempo; no hay una sucesión cuasi-instantánea de eventos: a diferencia de lo que ocurre entre empujar la libreta y que ésta se caiga, sabemos que el ibuprofeno tarda un tiempo en actuar (al menos media hora).

- Sabemos que, además del ibuprofeno, hay otros muchos factores que pueden influir en el dolor de cabeza de Juan, muchos de los cuales pueden haber actuado en el lapso de tiempo transcurrido (media hora). Esos posibles factores (eventos que han ocurrido de forma concurrente con el consumo del comprimido y que también afectan al dolor de cabeza) son los que, en epidemiología, acostumbramos a llamar "*factores de confusión*".

El razonamiento anterior explica que el modelo causal contrafáctico, que a nivel individual funciona muy bien para explicar la mayoría de asociaciones causales propias de la vida cotidiana, no sirve para, también a nivel individual, verificar hipótesis causales en el ámbito de la mayoría de los fenómenos relacionados con la salud. Así pues, no parece que el modelo contrafáctico nos vaya a ser de mucha utilidad en epidemiología. Sin embargo, como veremos más adelante, se puede demostrar que este mismo modelo sí que sirve cuando lo que se pretende no es estimar asociaciones causales a nivel individual, sino asociaciones causales promedio (poblacionales). Esto es así porque, aunque no es posible estimar el desenlace contrafáctico de un único suceso individual, sí que es posible (si se cumplen algunos requisitos), estimar el desenlace contrafáctico promedio de un conjunto de sucesos individuales.

1.2. El modelo causal contrafáctico a nivel poblacional (causalidad promedio)

Imaginémonos por un momento que viviéramos en el universo contrafáctico (aquél en el que los sucesos contrafácticos también pueden ser observados). En ese universo podríamos saber sin ningún género de dudas si el ibuprofeno que se tomó Juan fue o no el causante

de que se le quitara el dolor de cabeza. Imaginemos que en ese universo hay más personas en la misma situación que Juan (les duele la cabeza) y que para cada una de ellas podemos observar los dos sucesos a comparar (tomar o no tomar el ibuprofeno). Podemos suponer que a algunas personas el ibuprofeno les quita el dolor, a otras el ibuprofeno no les hace ningún efecto, e incluso que a otras el ibuprofeno les provoca dolor de cabeza. Es decir, tendremos cuatro posibles situaciones:

- 1) $Y(0) = 0$ e $Y(1) = 0$ El ibuprofeno no hace nada, el sujeto sigue con el mismo dolor lo tome o no.
- 2) $Y(0) = 1$ e $Y(1) = 1$ El ibuprofeno no hace nada, al sujeto se le quita el dolor lo tome o no.
- 3) $Y(0) = 0$ e $Y(1) = 1$ El ibuprofeno quita el dolor, pues este permanece sólo cuando el sujeto no se lo toma.
- 4) $Y(0) = 1$ e $Y(1) = 0$ El ibuprofeno causa dolor, pues este se quita sólo cuando el sujeto no se lo toma.

Sigamos en el universo contrafáctico. Supongamos que, dada una muestra de población con dolor de cabeza, la situación más frecuente en ellos es la número 3 (aunque hay sujetos que presentan las otras 3 situaciones). ¿Es posible calcular en esta muestra el efecto causal promedio del ibuprofeno? Sí; no hay ningún problema, siempre que tengamos en cuenta que es un concepto promedio, que podríamos expresar informalmente como....:

"En general, el ibuprofeno hace más bien que mal a estos sujetos".

"A la mayoría de la gente con dolor de cabeza el ibuprofeno les alivia".

"La probabilidad de que el ibuprofeno quite el dolor de cabeza es bastante mayor que 0".

"Si te duele la cabeza, toma ibuprofeno, porque es más probable que te alivie que no lo haga".

"En la muestra estudiada hay bastantes personas que han aliviado su dolor de cabeza gracias a que han tomado ibuprofeno".

Ninguna de estas afirmaciones predice lo que le va a pasar a cada persona de la muestra, pero todas son expresiones causales sobre el efecto de ibuprofeno en las personas.

¿Cómo se estima, en el universo contrafáctico, la existencia de un efecto causal promedio del ibuprofeno?:

$$E[Y(1)] \neq E[Y(0)], \text{ que se lee así:}$$

"El promedio de los desenlaces potenciales de Y cuando las personas reciben la intervención es diferente del promedio de los desenlaces potenciales de Y cuando no la reciben".

¿Y cómo se mide la magnitud del efecto causal? Pues, por ejemplo, así:

$$E[Y(1)] - E[Y(0)]$$

La cantidad anterior, en el lenguaje epidemiológico, sería un *Riesgo Atribuible en Expuestos* (RAE).

Para que las dos expresiones anteriores puedan expresarse en términos de causalidad promedio, la intervención ha de reunir dos características:

1. Los dos sucesos a comparar tienen que ser los mismos para todas las unidades experimentales: si unos sujetos toman un comprimido de 600 mg de ibuprofeno y otros lo toman de 250 mg, ya no tiene sentido estimar un efecto causal promedio. Esto es lo que se llama "intervención constante".
2. La intervención sólo ejerce su efecto en la unidad experimental que la recibe, no en el resto de las unidades de la muestra. El ibuprofeno que se toma Juan sólo influye en el dolor de cabeza de Juan, y no en el dolor de cabeza de Antonio o de cualquier otro sujeto de la muestra.

La suma de los dos requisitos anteriores es lo que se conoce como la hipótesis SUTVA (*stable unit treatment level assumption*).

Hasta aquí todo va bien. Pero seguimos estando en el universo contrafáctico, donde todos los pares de sucesos contrafácticos son observables. El problema surge al retornar al universo real. En él, para cada par de sucesos potenciales, uno es observado y el otro no. ¿Cómo estimamos $E[Y(1)]$ y $E[Y(0)]$ en el universo real? Veámoslo en el siguiente apartado.

2. PLANTEAMIENTO TEÓRICO DE LAS ESTIMACIONES CAUSALES PROMEDIO

Ya estamos en el universo real. Imaginemos 2000 sujetos con dolor de cabeza, la mitad toman ibuprofeno ($T=1$) y la otra mitad no ($T=0$). Necesitamos estimar dos cantidades:

a) $E[Y(1)]$: el valor promedio de Y si los 2000 sujetos hubieran tomado ibuprofeno. Para conocer esta cantidad necesitaría tener 2000 valores. Por desgracia, sólo tengo 1000 valores: los 1000 valores de Y cuando $T=1$. Me faltan los otros 1000 valores (los valores de Y(1) en los 1000 sujetos no tratados ($T=0$). En el universo real, lo único que puedo calcular es la siguiente cantidad: $E(Y|T=1)$, es decir, la media de los valores de Y cuando los sujetos son realmente tratados. ¿Es posible afirmar que $E[Y(1)] = E(Y|T=1)$? Sería posible siempre que se cumpliera una condición: que los 1000 sujetos que nos faltan (los realmente no tratados) fueran "intercambiables" con respecto a los 1000 sujetos que tenemos (los realmente tratados). Entonces se cumpliría la siguiente igualdad:

$E(Y|T=1) = E[Y(1)|T=0] = E[Y(1)]$, que se lee así: "el valor promedio de Y observado en los sujetos tratados es igual al valor promedio de Y que se habría obtenido si todos los sujetos que realmente no han sido tratados lo hubieran sido. Ambas expresiones son, por tanto, iguales al promedio de los desenlaces potenciales de Y cuando $T=1$."

b) $E[Y(0)]$: el valor promedio de Y si los 2000 sujetos no hubieran tomado ibuprofeno. El planteamiento es análogo al anterior: para conocer esta cantidad necesitaría tener 2000 valores. Por desgracia, solo tengo 1000 valores: los 1000 valores de Y cuando $T=0$. Me faltan los otros 1000 valores (los valores de $Y(0)$ en los 1000 sujetos tratados ($T=1$). En el universo real, lo único que puedo calcular es la siguiente cantidad: $E(Y|T=0)$, es decir, la media de los valores de Y cuando los sujetos no son tratados. ¿Es posible afirmar que $E[Y(0)] = E(Y|T=0)$? Sería posible siempre que se cumpliera una condición: que los 1000 sujetos que nos faltan (los que han sido tratados) fueran "intercambiables" con respecto a los 1000 sujetos que tenemos (los que realmente no han sido tratados). Entonces se cumpliría la siguiente igualdad:

$E(Y|T=0) = E[Y(0)|T=1] = E[Y(0)]$, que se lee así: "el valor promedio de Y observado en los sujetos realmente no tratados es igual al valor promedio de Y que se habría obtenido si todos los sujetos que realmente han sido tratados no lo hubieran sido. Ambas expresiones son, por tanto, iguales al promedio de los desenlaces potenciales de Y cuando $T=0$."

A partir del razonamiento anterior, lo que queda claro es que tenemos un problema de datos faltantes: nos faltan 2000 desenlaces contrafácticos (potenciales): lo que habría pasado con Y en los 1000 sujetos tratados si no lo hubieran sido, y lo que habría pasado con Y en los 1000 sujetos no tratados si lo hubieran sido. ¿Cómo estimamos estos 2000 datos faltantes? Ya hemos dicho antes que, si los sujetos tratados y no tratados fueran intercambiables, no habría ningún problema, pues sabemos que ambos grupos se comportarían igual si se hubieran "intercambiado los papeles". Luego el problema se reduce a tratar de demostrar que los sujetos $T=1$ y $T=0$ son intercambiables, o buscar algún procedimiento para hacerlos intercambiables, en caso de que no lo sean. Vamos a ver las distintas situaciones que pueden plantearse en función del tipo de estudio.

2.1. En un estudio experimental con asignación aleatoria marginal

Si una muestra grande (y 2000 sujetos lo son), se reparte al azar en dos subgrupos (por ejemplo, $T=0$ y $T=1$), puede demostrarse que ambos subgrupos son intercambiables, y entonces se cumple que:

$$\begin{aligned} E[Y(1)] &= E(Y|T=1) \\ E[Y(0)] &= E(Y|T=0) \end{aligned}$$

De donde $E(Y|T=1) - E(Y|T=0)$ es un estimador insesgado de la medida de causalidad promedio $E[Y(1)] - E[Y(0)]$

Al estudio anterior se le denomina estudio experimental puro con asignación aleatoria marginal (marginal porque se reparte toda la muestra al azar en dos grupos, sin tener en cuenta ninguna otra característica de los sujetos incluidos en ella). Veamos la situación anterior con un ejemplo numérico:

Sigamos con el ejemplo anterior (2000 sujetos con dolor de cabeza son asignados al azar a recibir o no ibuprofeno: la mitad lo toman ($T=1$) y la otra mitad no ($T=0$)). Imaginemos que de los 1000 tratados curan 800 y no curan 200. Entre los no tratados, curan 300 y no curan 700. Necesitamos estimar dos cantidades:

$$\begin{aligned} E(Y|T=1) &= 800/1000 = 0,8 \\ E(Y|T=0) &= 300/1000 = 0,3 \\ E(Y|T=1) - E(Y|T=0) &= 0,8 - 0,3 = 0,5. \end{aligned}$$

Esta cantidad es un estimador insesgado de

$$E[Y(1)] - E[Y(0)]$$

y cuantifica el efecto causal promedio del ibuprofeno sobre la desaparición del dolor de cabeza. Epidemiológicamente, lo interpretaríamos diciendo que por cada 100 personas de esa muestra con dolor de cabeza que tomaran ibuprofeno, a 50 les desaparecería el dolor de cabeza gracias al efecto causal del ibuprofeno sobre el dolor de cabeza (un RAE). Así pues, en un estudio experimental con aleatorización marginal se cumple que:

$$E[Y(1)] - E[Y(0)] = E(Y|T=1) - E(Y|T=0) \quad [1]$$

2.2. En un estudio experimental con asignación aleatoria condicionada

Hay estudios experimentales puros en los que la asignación aleatoria no es marginal, sino condicionada en una tercera variable. Es decir, primero se clasifica a la población en función de los valores de una tercera variable, y seguidamente se hace un reparto al azar dentro de cada uno de los estratos así definidos. Por ejemplo, imaginemos que, de los 2000 sujetos del estudio anterior, a 800 les duele la cabeza de forma intensa (cefalea fuerte: CF), mientras que a los 1200 restantes les duele de forma leve (cefalea leve; CL). Los investigadores deciden hacer el siguiente estudio:

- Van a asignar al azar a los 800 sujetos CF a dos grupos: 400 van a recibir ibuprofeno y 400 no van a recibirlo.
- Van a asignar al azar a los 1200 sujetos CL a dos grupos: 300 van a recibir ibuprofeno y 900 no.

En este estudio, la asignación aleatoria difiere de la realizada en el estudio con aleatorización marginal en dos circunstancias:

1. Se ha aleatorizado separadamente dentro de cada grupo definido por una tercera variable (aleatorización condicionada).
2. Además, las probabilidades de recibir la intervención (el ibuprofeno), son diferentes en cada estrato de la variable de condicionamiento: 0,5 en CF y 0,25 en CL (suponemos que los investigadores han hecho esto porque han considerado que, puesto que la CL es más soportable, era posible ahorrarse algo de dinero en el estudio asignando en este grupo a menos personas a que recibieran el fármaco).

Así pues, este es un estudio experimental con asignación aleatoria condicionada y con una probabilidad diferente de recibir la intervención en función de la variable de condicionamiento.

¿Qué esperamos que ocurra cuando tratemos de cuantificar en este estudio el efecto causal promedio del ibuprofeno? Puesto que dentro de cada estrato de intensidad hemos realizado una asignación aleatoria, esperamos que el azar haya desempeñado bien su papel (esto es asumible, pues seguimos disponiendo de un elevado tamaño muestral en cada estrato), y que, dentro de cada estrato de intensidad, los sujetos que reciben y no reciben el ibuprofeno sean intercambiables. Si esto se cumple, lo que esperamos es lo siguiente (expresado en notación matemática):

Si llamamos G a la variable intensidad, con dos categorías:

G=1: CF

G=0: CL

en los sujetos del grupo CF (G=1):

$E[Y(1)|G=1] = E(Y|T=1, G=1)$: "el valor promedio de los desenlaces potenciales de Y si hubieran sido tratados en los sujetos con cefalea fuerte es igual al valor promedio observado de Y en los sujetos con cefaleas fuertes que han sido realmente tratados".

$E[Y(0)|G=1] = E(Y|T=0, G=1)$: "el valor promedio de los desenlaces potenciales de Y si no hubieran sido tratados en los sujetos con cefalea fuerte es igual al valor promedio observado de Y en los sujetos con cefaleas fuertes que no han sido realmente tratados".

En los sujetos del grupo CL (G=0):

$E[Y(1)|G=0] = E(Y|T=1, G=0)$: "el valor promedio de los desenlaces potenciales de Y si hubieran sido tratados en los sujetos con cefalea leve es igual al valor promedio observado de Y en los sujetos con cefaleas leves que han sido realmente tratados".

$E[Y(0)|G=0] = E(Y|T=0, G=0)$: "el valor promedio de los desenlaces potenciales de Y si no hubieran sido tratados en los sujetos con cefalea leve es igual al valor promedio observado de Y en los

sujetos con cefaleas leves que no han sido realmente tratados".

En conclusión, dentro de cada estrato de intensidad es posible estimar el efecto causal promedio del ibuprofeno. Sólo hay que usar la ecuación [1], pero de forma condicionada. Así, realmente estamos midiendo dos efectos causales promedio del ibuprofeno (uno para las cefaleas fuertes y otro para las leves).

Para las cefaleas fuertes:

$$E(Y|T=1, G=1) - E(Y|T=0, G=1)$$

Para las cefaleas leves:

$$E(Y|T=1, G=0) - E(Y|T=0, G=0)$$

Veamos qué es lo que ha ocurrido realmente en el estudio:

1. En los 800 sujetos con CF:

- De los 400 sujetos que recibieron el ibuprofeno curaron 100, luego $E(Y|T=1, G=1) = 100/400 = 0,25$

- De los 400 sujetos que no recibieron el ibuprofeno curaron 50, luego $E(Y|T=0, G=1) = 50/400 = 0,125$

$$E(Y|T=1, G=1) - E(Y|T=0, G=1) = 0,25 - 0,125 = 0,125$$

2. En los 1200 sujetos con CL:

- De los 300 sujetos que recibieron el ibuprofeno curaron 240, luego $E(Y|T=1, G=0) = 240/300 = 0,80$

- De los 900 sujetos que no recibieron el ibuprofeno curaron 608, luego $E(Y|T=0, G=0) = 608/900 = 0,675$

$$E(Y|T=1, G=0) - E(Y|T=0, G=0) = 0,80 - 0,675 = 0,125$$

Las dos cantidades anteriores son estimaciones causales promedio insesgadas. Observamos que el ibuprofeno está causalmente asociado a una reducción del dolor de cabeza, de igual magnitud en los sujetos con cefalea fuerte que en los sujetos con cefalea leve. Podríamos decir que, en ambos grupos, por cada 100 sujetos con cefalea que consumieran ibuprofeno, 12,5 iban a reducir su dolor de cabeza debido al efecto causal del ibuprofeno sobre el dolor de cabeza. Recordando nuestros conocimientos epidemiológicos, diríamos que la intensidad del dolor de cabeza no modifica, de forma aditiva, el efecto causal del ibuprofeno sobre dicho dolor. Ahora bien, ¿qué pasaría si uniéramos ambos grupos de sujetos (CF+CL), para estudiar el efecto causal promedio marginal (en todos los sujetos) del ibuprofeno sobre la cefalea? Vamos a verlo:

Sujetos tratados con ibuprofeno: 700 (400 CF+300 CL). Curaciones en ellos: 340 (100 CG+240 CL)
 $E(Y|T=1) = 340/700 = 0,486$

Sujetos no tratados con ibuprofeno: 1300 (400 CF+900 CL). Curaciones en ellos: 658 (50 CG+608 CL)
 $E(Y|T=0) = 658/1300 = 0,506$

De aquí se obtiene que:

$$E(Y|T=1) - E(Y|T=0) = 0,486 - 0,506 = -0,020$$

En la muestra total estudiada, el consumo de ibuprofeno se asocia de forma inversa a la desaparición de la cefalea. Dicho de otra forma, en la muestra, los sujetos con cefalea que no toman ibuprofeno tienen una probabilidad ligeramente mayor de curarse que los que sí lo toman. Obviamente, la expresión $E(Y|T=1) - E(Y|T=0)$ no es un estimador insesgado del efecto causal promedio del ibuprofeno. ¿Por qué no? Porque, en la muestra total, los sujetos $T=1$ y $T=0$ (tratados y no tratados), no son intercambiables, y por lo tanto:

$$E(Y|T=1) \neq E[Y(1)] \\ E(Y|T=0) \neq E[Y(0)]$$

¿Y por qué no son intercambiables? Porque los sujetos $T=1$ y $T=0$ son diferentes con respecto a una circunstancia (la intensidad de la cefalea), que es un factor que, por sí mismo (es decir, independientemente de tomar o no ibuprofeno), influye en la probabilidad de que ésta desaparezca a la media hora: las cefaleas fuertes curan menos que las leves. Y si entre los sujetos tratados hay más cefaleas fuertes que entre los sujetos no tratados (en los sujetos tratados el 57% tienen CF, frente a tan sólo el 31% en los sujetos no tratados), se lo estamos poniendo "más difícil" al ibuprofeno para que demuestre su "poder curativo". Usando la jerga epidemiológica, la intensidad de la cefalea es un factor de confusión que sesga la estimación causal promedio marginal del efecto del ibuprofeno sobre la curación.

2.3. En un estudio observacional

Imaginemos que observamos a 2000 sujetos con cefalea, 800 con una CG y 1200 con una CG. Entre los primeros, observamos que 400 deciden tomar ibuprofeno y otros tantos no. Al cabo de 30 minutos, a 100 y 50 sujetos, respectivamente, se les quita el dolor de cabeza. Entre los segundos, 300 deciden tomar ibuprofeno y 900 no. También al cabo de 30 minutos se producen 240 y 608 curaciones, respectivamente (he empleado a propósito los mismos datos que en el caso anterior). ¿Qué tipo de estudio es este? Es un estudio observacional de cohortes. Si calculáramos a partir de

estos datos el RAE marginal, obtendríamos un valor de $-0,020$. ¿Este RAE es causal? No, porque viendo los datos parece claro que la asociación marginal (cruda), entre ibuprofeno y curación está confundida por la intensidad de la cefalea. Si estratificáramos (condicionáramos) por intensidad, obtendríamos que el RAE del ibuprofeno en los sujetos con cefalea fuerte es de $0,125$, igual al que obtendríamos en los sujetos con cefalea leve. Y de nuevo nos preguntamos: los dos RAE de $0,125$ que acabamos de obtener ¿son causales? A diferencia de lo que ocurría cuando los datos procedían de un estudio experimental con aleatorización condicionada, la respuesta ahora es "no lo sé". ¿Por qué? Porque antes sabíamos que, dentro de CF y CL, la asignación de T había sido al azar, por lo que, dentro de cada nivel de intensidad, los sujetos $T=1$ (tratados) y $T=0$ (no tratados), eran intercambiables. Ahora el estudio es observacional y no sé si dentro de cada estrato de intensidad los sujetos tratados y no tratados son intercambiables.

Ahora bien, supongamos por un momento que la única razón distinta del azar que influyera en la decisión del sujeto de tomarse o no el ibuprofeno fuera la intensidad de su cefalea (de forma que los sujetos fueran más propensos a tomar ibuprofeno si su cefalea fuera fuerte). Entonces sí se cumpliría la intercambiabilidad de tratados y no tratados dentro de cada estrato de intensidad y entonces los RAE que calculamos más arriba sí serían causales o, dicho en el lenguaje contrafáctico, si estimarían el efecto causal promedio del ibuprofeno en ambos subgrupos de sujetos.

Así que, en resumen, un estudio observacional de cohortes en el que la asignación de la exposición dependa únicamente de un conjunto "n" de factores conocidos (en nuestro ejemplo $n=1$ -la intensidad de la cefalea-) es equiparable, en cuanto a su diseño, a un estudio experimental con aleatorización condicionada por las "n" variables, porque, en ambos casos, se cumple el *principio de intercambiabilidad*: dentro de cada estrato formado por la combinación de los "n" factores, los sujetos expuestos y no expuestos (tratados y no tratados), son asignados aleatoriamente. Adicionalmente, si el estudio es observacional, hay que exigirle otro principio, el de *positividad*: dentro de todos los estratos formados por la combinación de los "n" factores, la probabilidad de recibir el tratamiento debe ser mayor que 0 y menor que 1. Este principio no se exige en los estudios experimentales porque se da por hecho: no tendría sentido, en nuestro ejemplo, hacer un estudio experimental con aleatorización condicionada en el que, por ejemplo, todos los sujetos con cefalea fuerte fueran tratados (pues este estrato no sería informativo a la hora de estimar el efecto causal promedio del ibuprofeno en él).