

Higiene y Sanidad Ambiental, 20 (2): 1865-1869 (2020)

Causalidad en epidemiología (III): La causalidad contrafáctica. Técnicas de análisis

CAUSALITY IN EPIDEMIOLOGY (I): COUNTERFACTUAL CAUSALITY. ANALYSIS TECHNIQUES

Luis Miguel MARTÍN de los REYES, Mario RIVERA IZQUIERDO, 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

INTRODUCCIÓN

En el artículo anterior se expusieron los fundamentos teóricos del modelo causal contrafáctico, a partir del cual, y previa verificación de ciertas premisas, podía obtenerse una estimación insesgada de la causalidad promedio a partir de un estudio observacional. En este capítulo presentamos las técnicas básicas de análisis para alcanzar este objetivo. Partimos de la base de que, si se cumplen las condiciones SUTVA, de intercambiabilidad y positividad, las técnicas que se explican seguidamente son aplicables tanto a estudios experimentales con asignación condicionada como a estudios observacionales. A partir de aquí hay dos grandes grupos de técnicas:

- Técnicas basadas en la estimación de los sucesos contrafácticos. Si, tras condicionar por n variables, los sujetos tratados y no tratados son intercambiables, vamos a estimar, en cada sujeto, lo que habría pasado en la situación contraria a la observada (lo que habría pasado si los sujetos tratados hubieran sido no tratados, y viceversa). Hay muchas formas de hacer esto, pero aquí sólo vamos a ver dos: la estandarización y la ponderación por el inverso de la probabilidad de ser tratado (*Inverse Probability Weighting*: IPW) (en realidad ambas técnicas hacen exactamente lo mismo, solo que el método de cálculo es diferente).

- Técnicas basadas en el emparejamiento de los sujetos: Sabemos que, marginalmente, los sujetos $T=1$ y $T=0$ no son intercambiables, porque la probabilidad de recibir el tratamiento es diferente en función de " n " factores. Vamos a resolver este

problema construyendo subgrupos de sujetos dentro de los cuales la probabilidad de recibir el tratamiento sea la misma. Aquí también hay diversas técnicas para lograr este objetivo; la más empleada es la que se basa en el cálculo de las puntuaciones de propensión (*Propensity Score*: PS).

Aunque parezcan diferentes, el fundamento es el mismo en todas estas técnicas: asumir que, tras condicionar por un conjunto de variables (los confusores), los sujetos tratados y no tratados son intercambiables. Lo que ocurre es que, en la práctica, el elevado número de confusores que existe en los estudios observacionales desaconseja el uso de la estandarización y hace que los otros dos métodos no tengan por qué dar resultados exactamente iguales.

1. ESTANDARIZACIÓN

Vamos a partir de los datos que nos han servido para ilustrar el planteamiento teórico tanto de los estudios experimentales con aleatorización condicionada (ver apartado del artículo anterior), como de los estudios observacionales (apartado 2.3), es decir, los del ejemplo del ibuprofeno para el dolor de cabeza. En ambos estudios veíamos que la asociación marginal entre ibuprofeno y curación era un estimador sesgado del efecto causal promedio del ibuprofeno, pero que las estimaciones condicionadas por la intensidad de la cefalea eran estimadores insesgados (asumiendo en ambos casos que este era el único factor que influía en la probabilidad de ser tratado). La pregunta es: ¿es posible, a partir de esta información, obtener un estimador único del efecto causal promedio del

ibuprofeno? La respuesta es que sí, mediante la estandarización directa. El punto de partida es que:

En los sujetos tratados:

- Si son del grupo CF (G=1):
 $E[Y(1)|G=1] = E(Y|T=1, G=1) = 0,25$
- Si son del grupo CL (G=0):
 $E[Y(1)|G=0] = E(Y|T=1, G=0) = 0,80$

En los sujetos no tratados:

- Si son del grupo CF (G=1):
 $E[Y(0)|G=1] = E(Y|T=0, G=1) = 0,125$
- Si son del grupo CL (G=0):
 $E[Y(0)|G=0] = E(Y|T=0, G=0) = 0,675$

A partir de estos valores (que son causales), vamos con nuestro planteamiento contrafáctico:

1. ¿Cuántos sujetos se habrían curado si toda la población del estudio se hubiera tratado? (estamos tratando de estimar $E[Y(1)]$). La respuesta es sencilla, sabiendo que, en la población total, hay 800 pacientes con CF y 1200 con CL (esta sería nuestra población estándar):

$$0,25 \times 800 + 0,80 \times 1200 = 200 + 960 = 1160.$$

Dividiendo esta cantidad entre 2000 se obtiene que...

$$E[Y(1)] = 0,58$$

2. ¿Cuántos sujetos se habrían curado si nadie en la población del estudio se hubiera tratado? (estamos tratando de estimar $E[Y(0)]$). Igual que antes, sabiendo que, en la población total (nuestra población estándar), hay 800 pacientes con CF y 1200 con CL:

$$0,125 \times 800 + 0,675 \times 1200 = 100 + 810 = 910.$$

Dividiendo esta cantidad entre 2000 se obtiene que...

$$E[Y(0)] = 0,455$$

3. A partir de los cálculos anteriores,
 $E[Y(1)] - E[Y(0)] = 0,58 - 0,455 = 0,125.$

2. PONDERACIÓN POR EL INVERSO DE LA PROBABILIDAD

De nuevo partimos de los mismos datos de antes. Con ellos podemos construir una tabla con las curaciones (Y=1) que se han producido en todas las combinaciones de las dos variables que influyen sobre ella (el tratamiento y la intensidad), de la siguiente manera:

Intensidad	Tratamiento*	Observado	
		N	Resultados *
Fuerte N=800	Sí (0,5)	400	Y=1: 100 (0,25)
	No (0,5)	400	Y=1: 50 (0,125)
Leve N=1200	Sí (0,25)	300	Y=1: 240 (0,8)
	No (0,75)	900	Y=1: 608 (0,675)

* Entre paréntesis hemos puesto la probabilidad de cada valor de T (sí o no) en cada nivel de intensidad y de Y=1 (curación) en cada nivel de tratamiento.

Ahora vamos con nuestro razonamiento contrafáctico: ¿cuántas curaciones habrían ocurrido entre los 2000 sujetos si a todos ellos les hubiesen acaecido los dos sucesos (el observado y el contrafáctico); es decir, que los tratados también se hubieran no tratado, y viceversa:

Intensidad	Tratamiento*	Observado		Total (Observado + Contrafáctico)	
		N	Resultados *	N	Resultados
Fuerte N=800	Sí (0,5)	400	Y=1: 100 (0,25)	800 Tratados	Curan = $800 \times 0,25 = 200$
	No (0,5)	400	Y=1: 50 (0,125)	800 No Tratados	Curan = $800 \times 0,125 = 100$
Leve N=1200	Sí (0,25)	300	Y=1: 240 (0,8)	1200 Tratados	Curan = $1200 \times 0,8 = 960$
	No (0,75)	900	Y=1: 608 (0,675)	1200 No Tratados	Curan = $1200 \times 0,675 = 810$

* Entre paréntesis hemos puesto la probabilidad de cada valor de T (sí o no) en cada nivel de intensidad y de Y=1 (curación) en cada nivel de tratamiento.

Una vez que tenemos estimados todos los desenlaces en todos los sujetos, es fácil calcular $E[Y(1)]$ y $E[Y(0)]$ para toda la muestra:

$$E[Y(1)] = (200 + 960) / 2000 = 0,58$$

$$E[Y(0)] = (100 + 810) / 2000 = 0,455$$

$$E[Y(1)] - E[Y(0)] = 0,58 - 0,455 = 0,125$$

Este método (en todo equivalente a la estandarización directa), se llama del inverso de la probabilidad de tratamiento porque las cantidades de la columna de la derecha (las curaciones que habrían ocurrido en el universo contrafáctico), en la práctica se obtienen simplemente multiplicando los valores realmente observados de curaciones por el inverso de la probabilidad de ser tratado o no tratado en cada estrato de intensidad:

Intensidad	Tratamiento*	Observado		Total (Observado + contrafáctico)	
		N	Resultados *	N	Resultados
Alta N=800	Sí (0,5)	400	Y=1: 100 (0,25)	800 Tratados	Curan = 100*1/0,5=200
	No (0,5)	400	Y=1: 50 (0,125)	800 No Tratados	Curan = 50*1/0,5=100
Baja N=1200	Sí (0,25)	300	Y=1: 240 (0,8)	1200 Tratados	Curan = 240*1/0,25=960
	No (0,75)	900	Y=1: 608 (0,675)	1200 No Tratados	Curan = 608*1/0,75=810

Tanto mediante la estandarización como mediante el IPW hemos obtenido un estimador causal promedio marginal para toda la muestra (en nuestro caso, la suma de los realmente tratados y los realmente no tratados). Cuando se hace esto se habla de la estimación del ATE (*Average Treatment Effect*). Es posible también obtener dichos estimadores causales promedio marginales bien para el subgrupo de sujetos realmente tratados (ATT: *Average Treatment Effect on the Treated*), o bien para el subgrupo de sujetos realmente no tratados (ATU: *Average Treatment Effect on the Untreated*).

3. PUNTUACIONES DE PROPENSIÓN

Retomando el ejemplo del ibuprofeno y la cefalea, ya se vio que, tanto en el estudio experimental con asignación condicionada como en el estudio de cohortes, la asociación marginal entre ibuprofeno y curación era una estimación sesgada de la asociación causal entre ambas variables, debido a que la probabilidad de recibir el tratamiento dependía no sólo del azar, sino de una variable (la intensidad de la cefalea) que, a su vez, también afectaba a la probabilidad de curarse. Ello vulneraba el principio de intercambiabilidad en la muestra total del estudio. De aquí se desprende que una estrategia para obviar este problema consistiría en seleccionar subgrupos de sujetos en los que la probabilidad de recibir el tratamiento fuera la misma (independientemente del valor de esta probabilidad). De hecho, esto es lo que se obtiene con la asignación aleatoria condicionada:

- Todos los individuos en el subgrupo de cefaleas fuertes tenían una probabilidad de 0,5 de recibir el ibuprofeno.
- Todos los individuos en el subgrupo de cefaleas leves tenían una probabilidad de 0,25 de recibir el ibuprofeno.

Vamos a llamar "P" a la probabilidad de ser tratado, y "p" a un valor concreto de esa probabilidad. Si, en un estudio observacional, somos capaces de medir cuáles son los factores que, estando relacionados con el desenlace de interés, influyen en la probabilidad de recibir el tratamiento, ya tenemos el problema resuelto, pues podremos comparar los valores de $E(Y|T=1, P=p)$ ("valores medios del desenlace en los sujetos tratados que tenían una cierta probabilidad de

ser tratados") con los valores de $E(Y|T=0, P=p)$ ("valores medios del desenlace en los sujetos no tratados que tenían la misma probabilidad de ser tratados que los anteriores").

En el diseño experimental con aleatorización condicionada se cumple que, para los sujetos con cefalea fuerte...:

$$E(Y|T=1, G=1) = E(Y|T=1, P=0,5) \text{ y } E(Y|T=0, G=1) = E(Y|T=0, P=0,5), \text{ pues el estudio se diseñó para que cuando } G=1, p=0,5.$$

Por eso, en este estudio:

$$E[Y(1)|G=1] - E[Y(0)|G=1] = E(Y|T=1, G=1) - E(Y|T=0, G=1) = E(Y|T=1, P=0,5) - E(Y|T=0, P=0,5) \quad [2]$$

De igual forma, para los sujetos con cefalea leve se cumple que...:

$$E[Y(1)|G=0] - E[Y(0)|G=0] = E(Y|T=1, G=0) - E(Y|T=0, G=0) = E(Y|T=1, P=0,25) - E(Y|T=0, P=0,25) \quad [3]$$

pues el estudio se diseñó para que cuando $G=0, p=0,25$.

En el estudio observacional todo lo dicho anteriormente es válido siempre que asumamos que, una vez condicionada por todos sus determinantes, la probabilidad de ser tratado (P) sólo depende del azar (algo que en el estudio experimental no había que asumir, pues el estudio se había diseñado para que

fuera realmente así). Es decir, ahora hemos de asumir que no hay más determinantes de P que los que hemos medido en nuestro estudio. A partir de aquí, el análisis de un estudio observacional constará de tres pasos:

1. Identificar cuáles son los determinantes de P que, a su vez, están asociados con el desenlace.
2. Conocer cuál es la probabilidad que tiene cada sujeto de ser tratado, condicionada al conjunto de determinantes antes identificados: esta es la puntuación de propensión o *propensity score* (PS) de cada sujeto.
3. Comparar el promedio del desenlace Y en subgrupos de sujetos que tengan el mismo (o parecido) PS. Esto es lo mismo que decir: estudiar la asociación entre T e Y condicionada por el PS.

3.1. Identificar los determinantes (D) de P

Se trata de, normalmente gracias a una revisión de la literatura científica, unida al conocimiento de las condiciones en las que se ha hecho el diseño observacional que se quiere analizar, establecer los determinantes de la probabilidad de recibir la intervención. Para ello es muy útil diseñar un Gráfico Acíclico Dirigido (DAG). Es esencial asegurarnos de que los determinantes lo son realmente (es decir, que la asociación entre D y P tiene el sentido $D \rightarrow P$ y no a la inversa ($P \rightarrow D$), pues entonces T sería la causa de D y, en ese caso, D podría ser un mediador de la secuencia causal entre T e Y ($T \rightarrow D \rightarrow Y$), con lo que al condicionar por D estaríamos introduciendo un sesgo (estaríamos bloqueando un camino causal).

3.2. Estimación de los PS

Esta fase del análisis de un estudio observacional equivale a la fase de diseño de un ensayo con asignación aleatoria (en ambos casos se establece el modo de asignar a los participantes en el estudio a recibir o no la intervención, sin tener en cuenta el desenlace Y bajo estudio). El procedimiento es sencillo. Si T es el tratamiento ($T=1$ y $T=0$) y D_1, D_2, \dots, D_n son los determinantes potenciales de la probabilidad de ser tratado, se trata de construir un modelo de regresión múltiple que permita estimar la probabilidad de $T=1$ en nuestra muestra a partir de los valores de todos los determinantes en ella. Como T es una variable dicotómica, se suele emplear para ello la regresión logística:

$$\log [p(T=1)/[1-p(T=1)]] = a + b_1D_1 + b_2D_2 + \dots + b_nD_n$$

A la cantidad $\log [p(T=1)/[1-p(T=1)]]$ se le llama el logit de $p(T=1)$. En nuestro ejemplo, en el que sólo hay un determinante de la probabilidad de ser tratado (la intensidad de la cefalea), la ecuación de regresión logística que habremos de ajustar nos dirá que "el logaritmo del cociente entre la probabilidad de tomar el ibuprofeno y de no tomarlo es igual a un valor constante (a) más el producto de otra constante (b) por D_1 (la intensidad, que puede valer 0 -cefalea leve- ó 1 -

cefalea fuerte-). Si, a partir de la base de datos con los 2000 pacientes, aplicamos el siguiente modelo de regresión logística (los cálculos están hechos con Stata):

logit T G

Obtenemos el siguiente resultado:

```
Iteration 0: log likelihood = -1294.8933
Iteration 1: log likelihood = -1229.6636
Iteration 2: log likelihood = -1229.3199
Iteration 3: log likelihood = -1229.3199

Logistic regression              Number of obs   =   2,000
                                LR chi2(1)      =   131.15
                                Prob > chi2     =   0.0000
Log likelihood = -1229.3199      Pseudo R2      =   0.0506
```

T	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
G	1.098612	.0971825	11.30	0.000	.908138 1.289087
cons	-1.098612	.0666667	-16.48	0.000	-1.229277 -.967948

Así pues:

$$\log [p(T=1)/[1-p(T=1)]] = -1,098612 + 1,098612 * G$$

De esta ecuación se obtiene que, cuando $G=1$, el logit de $p(T=1) = 0$, de donde la $odds(T=1) = \exp(0) = 1$. Finalmente, $p = odds / (1+odds) = 1/2 = 0,5$, que es la probabilidad de que $T=1$ en los sujetos con $G=1$ (cefalea intensa).

De igual forma, cuando $G=0$, el logit de $p(T=1) = -1,098612$, de donde la $odds(T=1) = \exp(-1,098612) = 0,333$. Finalmente, $p = odds / (1+odds) = 0,333/1,333 = 0,25$, que es la probabilidad de que $T=1$ en los sujetos con $G=0$ (cefalea leve).

(Stata hace directamente estos cálculos con el comando predict PS).

Así pues, ya hemos definido en nuestro estudio PS, que no es más que el valor de $p(T=1)$, para los distintos niveles de intensidad de la cefalea. En ocasiones, en lugar de tomar $p(T=1)$ como el PS, se toma directamente el logit de $p(T=1)$, porque sus valores suelen ajustarse mejor a una distribución normal.

Aunque la regresión logística es el procedimiento habitual para estimar el PS, pueden emplearse otras muchas técnicas.

3.3. Estimar la asociación entre T e Y condicionada por PS

Como en nuestro estudio solo tenemos dos valores de PS (0,5 y 0,25), bastaría con estimar dos medidas de asociación causal promedio:

- Para $PS=0,5$: $E(Y=1|T=1, PS=0,5) - E(Y=1|T=0, PS=0,5) = 0,125$
- Para $PS=0,25$: $E(Y=1|T=1, PS=0,25) - E(Y=1|T=0, PS=0,25) = 0,125$

Como todos los sujetos con CF ($G=1$) tienen un $PS=0,5$ y todos los sujetos con CL ($G=0$) tienen un $PS=0,25$, se cumple lo expresado en las ecuaciones [2] y [3].

En la práctica, el valor de PS es una variable cuantitativa (y no dicotómica, como en este ejemplo), porque en un estudio observacional el número de determinantes de $P(T=1)$ es generalmente amplio y, además, cada uno no tiene por qué ser, como en nuestro ejemplo, una variable con solo dos categorías (CF y CL), sino que puede presentar más de dos categorías o, incluso, ser una variable cuantitativa. Por esa razón, en la práctica no es posible formar subgrupos de sujetos tratados y no tratados con idéntico PS. De hecho, lo más frecuente es que se dé una situación en la que el rango de valores de PS en los sujetos $T=1$ no coincide con el rango de valores de PS de los sujetos $T=0$. Por ejemplo, si el rango de los PS de los tratados va de 0,2 a 0,8 y el de los no tratados va de 0,1 a 0,6, sólo pueden incluirse en el estudio los sujetos con PS entre 0,2 y 0,6 (pues sólo en este rango hallaremos sujetos tratados y no tratados). Lógicamente, cuanto mayor sea el solapamiento entre los PS de tratados y no tratados, mejor (menos sujetos a excluir). Al rango de valores de PS solapados entre tratados y no tratados se la llama *región de soporte común*.

Teniendo en cuenta las limitaciones anteriores, para condicionar la asociación entre T e Y por los valores de PS, se puede elegir alguna de las siguientes estrategias (o probar varias de ellas):

a) Una vez construido el PS, se incluye esta variable en un modelo de regresión de Y sobre T. Esta alternativa ha sido bastante cuestionada, porque se asume que plantea los mismos problemas que los modelos convencionales de regresión.

b) Se establecen subgrupos de sujetos que tengan valores aproximadamente iguales de PS. Para definir este concepto (¿qué se entiende por PS aproximadamente iguales?), existen varias aproximaciones. Las más empleadas son las siguientes:

- Emparejamiento: elegir, para cada sujeto con $T=1$ y $PS=p$, un sujeto con $T=0$ y $PS\approx p$. El emparejamiento se puede hacer 1:1 o 1:n. Se puede hacer simplemente por vecindad (elegir al

más cercano), o por vecindad pero con un límite (que la distancia entre el PS del tratado y del no tratado no supere un valor máximo admisible o *caliper*). El emparejamiento puede hacerse sin o con reemplazamiento (en este último caso, un sujeto $T=0$ puede emparejarse a más de un sujeto $T=1$).

- Estratificación: transformamos el PS (originalmente cuantitativo), en una variable categórica y definimos estratos que contendrán sujetos tratados y no tratados con PS aproximadamente iguales. Lo más frecuente es emplear quintiles, porque se ha demostrado que una estratificación en 5 categorías permite eliminar el 90% del sesgo introducido por una variable cuantitativa (como es el PS). No obstante, si el tamaño de muestra es elevado, se puede aumentar el número de estratos (con lo que logrará una mayor similitud de PS entre tratados y no tratados dentro de cada estrato). Una vez obtenida la estimación del efecto causal promedio en cada estrato, deben combinarse los estratos para obtener una estimación marginal para el conjunto de estratos.

Cualquiera que sea la técnica que se emplee, es necesario tener en cuenta que, para cada subgrupo "igualado" con respecto a su probabilidad de ser tratado, debería ser esperable que la distribución de los determinantes de dicha probabilidad fuera similar. Esta hipótesis se debe comprobar siempre que se hace un análisis de PS. En caso de que no se cumpla para alguna variable, esta debe ser tomada en cuenta (ajustar por ella), en la estimación de las medidas de efecto causales.

Por lo demás, recordemos que tanto el IPW como el PS requieren que se cumpla el principio de positividad (en cualquier subgrupo de sujetos, la probabilidad de ser tratado debe ser >0 y <1). Imaginemos que $p(T=1)=0$ en un subgrupo de sujetos. En ese subgrupo el inverso de la probabilidad sería $1/0$, con lo que no se podría aplicar la fórmula IPW. Por lo demás, en un subgrupo de sujetos con un PS de 0 no se podría estimar ninguna medida causal promedio para el efecto del tratamiento, por razones obvias (en ese grupo no habría sujetos tratados).