

Sobre la Interpretación de la Evidencia Obtenida por los Estudios Cuantitativos de Investigación en Salud

Uno de los objetivos de la investigación en salud es generar evidencias sobre la efectividad de tratamientos o medidas de prevención. Este tipo de problema, que requiere la comparación de los desenlaces obtenidos, u observados, en dos o más grupos diferentes de sujetos, usualmente se aborda mediante métodos de investigación cuantitativos. El reto epistemológico es poder determinar si la causa putativa realmente genera el efecto estudiado, es decir si la intervención produce el resultado medido o si el factor estudiado es una causa del desenlace observado. A este proceso de razonamiento se le denomina, en la estadística y la filosofía de la ciencia, *inferencia causal*.^{1,2}

La inferencia causal es un proceso de razonamiento inductivo que permite concluir, con diferentes grados de certeza, que las estimaciones sobre los efectos causales de un estudio específico son válidas y generalizables.³ Para que lo anterior ocurra, se deben cumplir una serie de supuestos. Teóricamente, un estudio es válido cuando no se presentan ni errores aleatorios, ni errores sistemáticos (o sesgos).⁴ Los errores aleatorios se derivan del hecho de que

los estudios cuantitativos se realizan utilizando muestras al azar de sujetos. Para minimizar este error se debe garantizar que el proceso de selección sea aleatorio y se debe contar con un tamaño de muestra apropiado.⁵ Los sesgos se presentan por errores en la forma como se diseña y se desarrolla operativamente un estudio de investigación, los cuales afectan la comparabilidad de los datos y resultados que se obtienen de los sujetos de investigación.⁶ Pueden existir sesgos en la forma como se seleccionan los sujetos de estudio (sesgo de selección), sesgos en la forma en que se obtienen los datos del estudio (sesgo de información, incluyendo los sesgos de medición, memoria y detección) y sesgos mixtos combinando los anteriores.^{7,8}

Conceptualmente, se pueden estimar los efectos causales de una intervención a partir de los estudios experimentales, porque el control de la intervención mediante la asignación aleatoria de la misma, permite simular un escenario contra-factual en el cual los individuos bajo estudio son perfectamente intercambiables en relación al tratamiento recibido (intervención vs. control).⁹ De esta forma, el *experimento ideal* es aquél que cumple los siguientes criterios: i) se selecciona aleatoriamente un número grande de unidades de observación (ej. personas) a partir de una población de interés bien definida, ii) se identifican y se miden todos los factores de confusión conocidos, iii) se diseñan bloques de asignación utilizando esos confusores, y iv) el tratamiento o el control se asignan aleatoriamente a cada sujeto al interior de dichos bloques.² Pero en la vida real, es muy difícil que una investigación en salud cumpla con todas los criterios de un experimento ideal.⁴

Con relación a las investigaciones cuantitativas en salud sobre sujetos individuales, se ha considerado que el mayor nivel de

evidencia causal se obtiene a partir de los *Ensayos Controlados Aleatorizados* (ECA, la cual es la mejor traducción del término en inglés *Randomized Controlled Trial*), ya sean clínicos, comunitarios o de prevención, porque teóricamente la asignación de la intervención sólo está determinada por el azar y ninguna otra característica de los sujetos intervenidos, o del contexto, explicaría el resultado obtenido; simulando así un experimento contra-factual en el cual el desenlace es explicado solamente por la intervención bajo estudio.⁶ Con este tipo de diseño experimental no se deberían presentar errores de desbalance de los tratamientos (i.e. diferencias en los grupos de sujetos según su tratamiento asignado) si la muestra es lo suficientemente grande.³ Además, si la valoración del resultado se realiza de forma enmascarada respecto al tipo de tratamiento recibido, la posibilidad de ocurrencia de sesgos de información (i.e. errores sistemáticos en la medición) se minimiza.¹⁰ Finalmente, para que no se presenten sesgos en los ECA se requieren dos requisitos adicionales: que los sujetos de estudio cumplan con el tratamiento asignado según los protocolos y que no se presenten pérdidas de seguimiento.^{6,11} A pesar de que los ECA están ubicados en la cúspide de la jerarquía de la evidencia obtenida por las investigaciones en salud, solo por debajo de los meta-análisis,¹² estos ensayos presentan una serie de desventajas en relación con el *experimento ideal*, definido atrás: La selección de los sujetos que participan en la investigación generalmente no se realiza al azar, sino por conveniencia (ej. un grupo de pacientes que sufren la enfermedad en los hospitales del estudio); además esta selección está condicionada por los criterios de inclusión y exclusión de cada estudio y por la aceptación de participar

por parte de las personas que se contactan inicialmente. Estos problemas de selección llevan a limitaciones en la generalización de los resultados de los ECA más allá de las características propias de los participantes, ya que el error de estimación debido a sesgos en la selección de la muestra no es cero.^{2,3} Por otro lado, en la mayoría de los ECA los tamaños de la muestra son muy pequeños y la aleatorización no se realiza en bloques, situación que puede generar errores de estimación residuales por desbalances de las características (observadas y no observadas) de los sujetos en los grupos de tratamiento, que pueden afectar los hallazgos del estudio.^{2,3} Este último problema se puede solucionar parcialmente con la utilización de métodos de ajuste estadístico para controlar la confusión,¹³ como las regresiones múltiples;^{4,14} pero se debe tener en cuenta que el desbalance de los tratamientos respecto a variables no observadas no se resuelve con estos análisis estadísticos;⁴ entonces bajo esas circunstancias, la interpretación y generalización de los resultados de los ECA se verá limitada.³

Como se ha explicado, existen algunas precauciones en relación con la interpretación de los hallazgos de los ECA, pero estas precauciones deben ser más estrictas cuando se interpretan los hallazgos de los *estudios observacionales analíticos* (i.e. estudios de cohortes, de casos y controles, transversales y sus subtipos), porque el proceso de inferencia causal a partir de estos estudios es mucho más limitado.^{4,5} Con excepción de las investigaciones que se basan en muestras poblacionales aleatorias, todos los estudios epidemiológicos observacionales presentan algún grado de error de estimación debido al proceso de selección de la muestra (i.e. sesgos de selección), el cual es mucho mayor cuando las muestras se obtienen por

conveniencia (ej. controles escogidos a partir de sujetos hospitalizados) y no existe información disponible para corregir la selección diferencial de los sujetos de estudio (i.e. los pesos de muestreo u otras características del proceso de selección).³ Por otro lado, en la mayoría de los estudios observacionales no existe una intervención o tratamiento, sino que se estudia el efecto de una exposición o factor de riesgo (ej. uso de cigarrillo, consumo de β -carotenos) que no puede ser manipulada como en un experimento (ECA, en investigación en salud), sino que se *observa* tal como se presenta en los sujetos de estudio (ej. fuma o no fuma).⁷ De esta forma se presenta un fenómeno llamado *confusión*, en el cual se encuentra una asociación estadística espuria entre la variable de exposición (ej. consumo de café) y la variable resultado (ej. cáncer de pulmón), que en su lugar es realmente explicada por una tercera variable, o *confusor* (ej. uso de cigarrillo), que está relacionada no causalmente con la exposición y además está asociada independientemente con el resultado o desenlace.^{7,10} Como consecuencia de lo anterior, se presentan errores de estimación debidos a desbalances de las características (observadas y no observadas) de los grupos de personas expuestas y no expuestas, los cuales afectan los resultados del estudio.³

Por las razones aquí explicadas, en el análisis de los estudios observacionales no es suficiente estimar las relaciones bivariadas entre la exposición y el desenlace (ej. consumo de café y cáncer de pulmón), sino que se requieren métodos estadísticos más complejos, como la regresión múltiple, el emparejamiento o las variables instrumentales.^{4,7,12}

Los modelos de regresiones múltiples permiten estimar el efecto ajustado (o neto) de la exposición o intervención sobre la variable

resultado, teniendo en cuenta los efectos de las variables de confusión.^{4,13} Para elaborar un modelo de regresión apropiado se debe tener en cuenta el conocimiento previo sobre el fenómeno estudiado, para así poder incluir las variables de confusión indicadas en cada situación, las cuales pueden ser identificadas mediante técnicas como los diagramas causales;^{4,15} sin incluir más variables que las necesarias para no generar un sesgo de sobre-ajuste del modelo de regresión múltiple.¹⁶ Por otro lado, los modelos de regresión múltiple deben evitar el sesgo de ajuste por el efecto común, que ocurre cuando se incluye una variable de un efecto común (ej. fiebre), que es causado tanto por el desenlace estudiado (ej. gripa) como por otra variable del estudio (o covariable, ej. una intoxicación alimentaria), presentándose una asociación espuria entre la covariable y el desenlace estudiado (ej. entre la intoxicación alimentaria y la gripa).¹⁷ A pesar de las precauciones aquí indicadas, los modelos de regresión múltiple no pueden controlar la confusión residual que se presenta por variables no medidas, o desconocidas, las cuales pueden sesgar la magnitud o la dirección de los hallazgos del estudio.^{6,18}

Por otro lado, el emparejamiento es un método que permite obtener un balance estadístico entre los individuos intervenidos y no intervenidos (o expuestos y no expuestos) mediante la estimación del puntaje individual de la propensión (o probabilidad) de recibir o no la intervención, dadas las características de cada individuo antes de recibir la intervención (o exposición).¹⁹ Utilizando ese puntaje se emparejan los individuos tratados y no tratados, buscando balancear las características de las personas que reciben y no reciben el tratamiento, simulando el balance que se obtiene al asignar aleatoriamente los tratamientos o intervenciones, disminuyendo así el error

debido a la asignación no aleatoria de los tratamientos.^{18,20} Finalmente, a pesar de todos los métodos aquí descritos para el diseño y el análisis de los estudios observacionales, cuando se pretende estimar el efecto de intervenciones en salud mediante estudios observacionales, no es posible controlar adecuadamente la *confusión por indicación*, la cual se refiere a la imposibilidad de separar el efecto del conjunto de indicaciones del tratamiento (medidas y no medidas), del efecto real del tratamiento sobre la variable resultado estudiada.¹⁸

Como se ha descrito aquí, la obtención e interpretación de los resultados a partir de los estudios cuantitativos, con el propósito de generar evidencias en salud, es un proceso complejo que requiere tener en cuenta múltiples y complejos aspectos relacionados con el diseño y el análisis de estos estudios, ya sean experimentales u observacionales. Ningún estudio de investigación en la vida real es perfecto y tampoco ningún estudio individual es *per se* evidencia suficiente para cambiar las prácticas en salud. A pesar de estas limitaciones, cuando los estudios experimentales y observacionales se realizan siguiendo los métodos actuales de investigación en salud,^{4,7} ambos tipos de estudios se complementan en la tarea de acumular evidencias válidas y confiables, efectuada mediante los meta-análisis y las revisiones sistemáticas,²¹ para que esas evidencias finalmente sean útiles en mejorar la calidad de vida de los pacientes y sus familias.

Andrés Fandiño-Losada MD MSc. PhD.
Profesor Asistente y Coordinador, Maestría
en Epidemiología.
Escuela de Salud Pública.
Facultad de Salud. Universidad del Valle
Cali, Colombia.

Bibliografía

- 1 Pearl J, Glymour M, Jewell NP. 2016 Causal Inference in Statistics: A Primer (Wiley, New York)
- 2 Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005; 95(S1): S144-S150.
- 3 Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference. *J R Stat Soc Series A (statistics in society)* 2008; 171:481-502.
- 4 Rothman KJ, Greenland S, Lash T L(Eds.). 2008 Modern epidemiology.(Lippincott Williams & Wilkins)
- 5 Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. *The Lancet* 2001; 357: 373-380.
- 6 MacMahon S, Collins R. Reliable assessment of the effects of treatment on mortality and major morbidity, II: observational studies. *The Lancet* 2001;357: 455-462.
- 7 Szklo M, Nieto J. 2012 Epidemiology. (Jones & Bartlett Publishers).
- 8 Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health* 2004; 58: 635-641.
- 9 Höfler M. Causal inference based on counterfactuals. *BMC medical research methodology* 2005 5(1): 1.
- 10 Gluud LL. Bias in clinical intervention research. *Am J Epidemiol* 2006;163:493-501.
- 11 Frangakis CE, Rubin DB . Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999;86:365-379.
- 12 Guyatt GH, Haynes RB, Jaeschke R Z, Cook DJ, Green L, Naylor CD, et al. Evidence-Based Medicine Working Group. Users' guides to the medical literature: XXV. Evidence-based medicine: principles for applying the users' guides to patient care. *JAMA* 2000;284:1290-1296.
- 13 de Irala J, Martínez-González M Á, Grima FG. ¿ Qué es una variable de confusión?. *Medicina Clínica* 2001;117:377-385.
- 14 McNamee R. Regression modelling and other methods to control confounding. *J Occup Env Med* 2005; 62:500-506.
- 15 Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002; 155:176-184.
- 16 Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology (Cambridge, Mass.)* 2009; 20:488.
- 17 Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, Poole C. Illustrating bias due to conditioning on a collider. *Int J Epidemiol* 2010;39:417-420.
- 18 Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* 2010;63:64-74.
- 19 Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J* 2011; 32:1704.
- 20 Stuart EA. Matching methods for causal inference: A review and a look forward. *Statistical science: a review. Stat Sci* 2010; 25: 1.
- 21 Shrier I, Boivin JF, Steele RJ, Platt RW, Furlan A, Kakuma R, et al . Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007; 166:1203-1209.