

CAUSALIDAD EN LA CIENCIA

Pearl (Pearl & Mackenzie, 2018, # 16) propone la escalera de la causalidad para efectos prácticos en el trabajo científico. Esta escalera tiene tres escalones, empezando por la capacidad de la ciencia para analizar la causalidad mediante la recolección de información a partir de la observación sobre las variables causales y el efecto atribuible a ellas, datos que nos podrán evidenciar ya sea en forma de singularidad o de regularidad las relaciones entre las variables. En este primer escalón se debe tener cuidado con el peligro de no lograr aislar eficientemente la acción de la principal variable causal puesto que no todas las relaciones entre las variables son causales, pero las asociaciones pueden ser estadísticamente significativas.

En un segundo escalón se encuentra la intervención, con dos grandes vertientes; en la primera, las investigadoras inducen la intervención en forma programada y controlada (experimental), mientras que en la segunda se documenta una exposición a la que se han sometido las unidades de observación en forma espontánea o inducida (naturalmente o antropogénicamente). En este escalón la ventaja destacable en el modelo experimental es que las variables de confusión se controlan mediante la aleatorización en la asignación de las unidades de intervención a los grupos expuesto y no expuesto. Además se ha venido desarrollando a lo largo del siglo XX un aparataje de pruebas estadísticas capaces de dar certidumbre sobre las relaciones potencialmente causales y diferenciarlas de las relaciones simplemente de asociación estadística. Sin embargo, una respetable corriente de investigadores cuestiona la precisión de las conclusiones que se obtienen mediante este mecanismo de recolección de información, llegando a afirmar por ejemplo que *“muchas de las medicinas que tenemos en casa han sido desarrolladas sobre supuestos cuestionables”* (Pearl & Mackenzie, 2018, # 152).

En el escalón más alto (tercero) se encuentra la imaginación, mediante la cual se analizan matemáticamente los resultados potenciales, “prendiendo y apagando” (en los modelos de causalidad) las exposiciones en grupos de individuos y calculando los efectos para las diversas posibilidades de combinación entre tener la exposición y el efecto y no tener la exposición ni presentar el efecto (potential outcomes), todo esto contextualizado por el conjunto de variables de confusión, mediación y colisión que eventualmente se requieran incluir. Esta modelización matemática permite incluso la presencia de variables no medidas o desconocidas, que se incluyen en el modelo contextual con valores asignados hipotéticamente. Este es el análisis contrafáctico, en él se requiere respetar y ser capaces de procesar algunos requisitos que han venido evolucionado en los últimos tiempos y que seguirán apareciendo.

En los tres niveles de la escalera es indispensable declarar abiertamente los supuestos teóricos en los que se sustentan las relaciones entre las variables, no solo las relaciones de causalidad sino también la independencia de las variables entre las que no existe una relación causal. Estos supuestos han recibido un interés cada vez más preponderante en el desarrollo de los modelos de causalidad, se argumenta por parte de los innovadores de la nueva ciencia de la causalidad que los estudios basados tanto en la observación como en la intervención se concentran más en las relaciones estadísticas entre las variables que en ofrecer explicaciones teóricas sobre los supuestos en los que se basa su modelo de relación de variables. Por su parte, quienes abordan la causalidad desde la nueva propuesta de análisis necesitan explicitar gráficamente estas relaciones y realizar procedimientos lógicos que permitan

evidenciar cuáles relaciones son de causalidad y cuáles relaciones son de otro tipo de asociación.

“Irónicamente, la necesidad de una teoría de la causalidad empieza a surgir al mismo tiempo que lo hace la estadística. De hecho, la estadística moderna eclosiona a partir de las preguntas causales que Galton y Pearson elaboraron acerca de la herencia y de sus genuinos intentos para responderlas usando datos intergeneracionales. Desafortunadamente, ellos fallaron en su intento y en lugar de detenerse y preguntarse por qué, ellos declararon esas preguntas fuera de los límites y voltearon a desarrollar un emprendimiento prometedor, libre de causalidad, denominado estadística” (Pearl & Mackenzie, 2018, # 5)

“La mayoría de los estudios en las ciencias de la salud, sociales y del comportamiento tienen como objetivo responder preguntas causales en lugar de asociativas. Tales preguntas requieren cierto conocimiento del proceso de generación de datos, y no se puede calcular a partir de los datos solamente, ni de las distribuciones que gobiernan los datos. Sorprendentemente, aunque gran parte de los conceptos, el marco y las herramientas algorítmicas necesarias para abordar tales problemas están ahora bien establecidos, no son conocidos por muchas de las investigadoras que podrían ponerlos en uso práctico. Resolver problemas causales de manera sistemática requiere ciertas extensiones en el lenguaje matemático estándar de la estadística, y estas extensiones no suelen enfatizarse en la literatura convencional”. (Pearl, 2010, #)

La versión estadística (“frecuentista”) de la causalidad está ampliamente difundida en el mundo académico e incluso en campos como por ejemplo la industria relacionada con la salud. El conflicto entre estudios experimentales y observacionales es la discusión central sobre la causalidad presente en la literatura científica en español. A su alrededor se han desarrollado complejas estructuras de pensamiento como los esquemas de gradación de la evidencia científica, que han evolucionado a esquemas de decisiones basadas en evidencia en varios campos de la actividad científica y profesional.

La teoría de la causalidad como regularidad (rtc/tcr) se ve bien reflejada en el abordaje científico basado en la probabilidad, si se considera que la prioridad temporal, la conjunción constante y la contigüidad espacial son procesadas con bastante certidumbre mediante los procedimientos estadísticos más frecuentemente utilizados que estudian las relaciones entre variables mediante tablas de contingencia y gráficos de dispersión y regresión. Bajo los supuestos adecuados y con datos recolectados con rigurosidad, estos procedimientos pueden dar buena cuenta de las relaciones de causalidad entre las variables; sin embargo, existe el peligro de confundir las relaciones de asociación identificadas por estos procedimientos con relaciones de causalidad, en especial cuando estos procedimientos estadísticos no se basan en un modelo interpretativo previo, en el que los supuestos de causalidad estén claramente declarados.

En el escenario observacional de la causalidad la observación de la causa modifica la percepción de la investigadora sobre la probabilidad de la aparición del efecto. En el trabajo científico no es infrecuente que una vez que se observa la presencia de la causa seguida de la presencia del efecto, se procure realizar nuevas observaciones en las que se repita esta situación. Se podría decir que se pasa de una explicación de la causalidad como singularidad a la búsqueda activa de la causalidad como regularidad para la misma relación. La noción de búsqueda activa de la regularidad condujo a los

científicos occidentales a desarrollar las pruebas estadísticas de fortaleza de asociación, independencia, riesgo, regresión y tiempo hasta el evento (bondad de ajuste a la curva), que se utilizan para procesar los datos generados mediante una recolección observacional de los datos.

La consolidación de la estadística al inicio del siglo XX, según Pearl, produce un bloqueo (que dura un siglo) en el desarrollo de un instrumental científico que permita dilucidar las relaciones de causalidad de las de asociación.

La intuición humana está asentada en una lógica causal, no estadística (Pearl & Mackenzie, 2018, # 14).

Las pruebas de inferencia estadística de: estrechez de la asociación, independencia, riesgo, los procedimientos de regresión lineal, múltiple y logística y los análisis de “tiempo hasta el evento”, constituyen la columna vertebral del denominado “enfoque frecuentista” del análisis de datos estadísticos que las profesionales de salud aprendieron durante todo el siglo pasado y siguen siendo dominantes en la enseñanza actual. Pearl las denomina “un emprendimiento prometedor, libre de causalidad”, en especial por la probabilidad de que el resultado de las pruebas estadísticas muestre “correlaciones espúreas” entre las variables bajo análisis, simplemente por la falta de un esquema previo de relación de variables que proporcione una justificación teórica para esta relación o por los efectos de una (o varias) tercera variable “al acecho” que puede ser inapropiadamente incorporada en el análisis tanto lógico como estadístico.

El enfoque frecuentista establece la utilidad de cada prueba estadística para el tipo de variables a ser analizadas y para las posibles distribuciones de estos datos (paramétricas o no paramétricas) y es fácil encontrar matrices del uso de cada prueba en relación con el tipo de datos disponibles y el tipo de relaciones entre las variables que se asumen desde la teoría. Esto no es posible para el caso de la inferencia causal, en el libro “What if” se expresa con claridad que no se dispone de un listado de usos de procedimientos algebraicos o algoritmos para el procesamiento de las relaciones causales en un modelo de relación de variables.

*“El libro describe varios enfoques de análisis de datos para estimar el efecto causal de interés bajo un conjunto particular de suposiciones cuando se recopilan datos sobre cada individuo en una población. **Un mensaje clave del libro es que la inferencia causal no puede ser reducida a una colección de recetas para el análisis de datos.**” (Hernán & Robins, 2020, # vii)*

Múltiples objeciones surgen sobre la verdadera capacidad de la estadística para resolver el problema de la causalidad a partir de los datos, son objeciones similares a las que ha recibido la Teoría de la Causalidad como Regularidad. A continuación se presentan sólo tres de estas objeciones:

- ¿Cómo puede una relación basada en la observación de eventos similares ser objetiva? cada vez que se observa una causa, aparece el efecto y cuando esto se repite en forma constante es muy humano pensar que una vez que aparece la causa aparecerá el efecto o cuando aparece el efecto debió estar presente la causa. Esto se postula como un hecho objetivo, pero en realidad de lo que se trata es de situaciones similares. (Talbot, 2016, Lecture 1 1:03)
- La forma de recolectar los datos es permeable a la presencia de sesgos que deforman la intercambiabilidad entre las unidades de observación expuestas y no expuestas. La imposibilidad de comparación de una misma observación en la misma unidad antes y después de una intervención hace necesario recurrir a comparaciones estadísticas que se hacen entre los valores resumen de un

grupo de observaciones comparada con los valores resumen de otro grupo de observaciones y que terminan siendo aplicadas a los individuos de cada grupo y siendo inferidas a la población total. Esto es cierto no solo para los estudios observacionales sino también para los experimentales.

- ¿Qué pasa con las excepciones o cuando la aparición de la causa no se acompaña del efecto y viceversa? Es impensable la total consistencia en la que todas las unidades de observación que se encuentren expuestas a la causa presenten el efecto y todas las que no se encuentren expuestas no lo presenten (como en las leyes de la física). Esto abre una puerta para que las diferencias estadísticas entre los dos grupos no puedan ser mecánicamente extrapoladas a las poblaciones de las que provienen las observaciones y que cada vez sea más evidente que se requieren mecanismos metodológicos para individualizar los “beneficios” y los “daños” atribuibles a las intervenciones.

La existencia del paquete de terceras variables al acecho, de alguna manera incorporadas en las cadenas y trayectorias causales que había derrotado a la humanidad en el camino a la comprensión de la causalidad, fue abordada con bastante capacidad resolutoria por la estadística. Se pudo procesar la información recolectada en estudios transversales, se diseñaron los estudios de “casos y testigos” con su correspondiente instrumental estadístico. Estos avances no resuelven el problema de las variables de confusión ni de la introducción de sesgos en la recolección de los datos y en su análisis, pero con estas opciones “observacionales” de recolección y procesamiento de datos ya fue posible identificar potenciales candidatas a variables causales que permitieron grandes avances tecnológicos y en la práctica profesional.

Entran en el escenario otros elementos adicionales que complican la inferencia causal a partir de los tres antes mencionados (prioridad temporal, la conjunción constante y la contigüidad espacial); provienen de la versión contrafáctica de la causalidad: anticipación temporal (temprana y tardía), efecto eclipse y ausencia de medición. Para incorporar estos escenarios de la causalidad los procedimientos básicos de la inferencia estadística son cada vez menos potentes. El tratamiento de las “terceras variables al acecho” (variables de confusión) mediante el cierre de la “puerta de atrás” es muy diferente desde la perspectiva de los defensores de la estadística (frecuentistas), que desde la perspectiva de quienes demandan la explicitación de los supuestos teóricos de la relación de variables mediante el uso de un lenguaje gráfico y el tratamiento de las relaciones entre variables mediante modelos estructurales causales (cálculo). Adicionalmente, el manejo de las relaciones de mediación (que deben diferenciarse claramente de las de confusión) requiere de modelos matemáticos estructurales que no han recibido el mismo interés por parte de la estadística “frecuentista”. El lenguaje matemático que se requiere para el discernimiento de las relaciones entre variables que vaya más allá de las relaciones causales sustentadas en la teoría de la causalidad como regularidad y que incorporen las relaciones sustentadas en la teoría contrafáctica de la causalidad plantea ineludiblemente el uso de un enfoque de probabilidad condicional que deriva en la necesidad de utilizar el cálculo diferencial y otros recursos de las matemáticas avanzadas.

Finalmente se logró el control de las variables de confusión mediante el diseño de recolección de datos denominado Ensayo Aleatorio Controlado (*Randomized Control Trial*) (EAC/RCT), hasta convertirlo en el “estándar de oro” de la evidencia científica y de la identificación científica de la causalidad. En este escenario experimental, que utiliza una forma diferente de recolección de datos (prospectiva, aleatoria y cegada), el desarrollo de las alternativas de tratamiento estadístico de los datos es aún mayor y se acompaña también de una explosión de las formas de conformar y dar seguimiento a los grupos experimental y de control. La mayor dificultad que enfrenta esta forma

sistemática de recolección de información es que no siempre es posible llevarla a la práctica, tanto por temas logísticos como por problemas éticos y legales.

La declaración del EAC/RCT como en “estándar de oro” de la inferencia causal en la primera mitad del siglo XX distrajo a las investigadoras de otras posibilidades de abordaje de la causalidad, en especial de la novedad que trae a escena el redescubrimiento del planteamiento Humenano de la opción contrafáctica de la causalidad que fue propuesto al mismo tiempo que la opción de la causalidad como regularidad pero no desarrollado con la misma intensidad. Progresivos avances se fueron produciendo aisladamente, hasta que las nuevas necesidades de la realidad como la programación de instrumentos electrónicos, la abundancia de datos generada mediante registros regulares y los inconvenientes de aplicación de los EAC/RCT, hacen evidente la necesidad de ampliar la capacidad de los enfoques singularistas y de la regularidad. La opción contrafáctica acelera su lento desarrollo en la segunda mitad del siglo XX y se desarrolla exponencialmente en las últimas tres décadas.