CUESTIONAMIENTOS A LOS EAC/RCT

La serie de características enumeradas ciertamente ofrecen un modelo de recolección de datos y de asociación de variables bastante sólido y que ha permitido en la segunda mitad del siglo XX un desarrollo innegable de las ciencias de la salud. Sin embargo, existen múltiples autores que analizan los defectos de este modelo, algunos de los argumentos se recogen a continuación.

El solo hecho de que una relación causal no haya sido comprobada mediante un diseño experimental como el ECA/RCT no significa que no exista en la realidad. La rotación de la tierra, los problemas de salud derivados de la evolución de la especie, los conflictos sociales, etc., solo son factibles de estudiar mediante observación, a nadie se le habría ocurrido negar su existencia por el hecho de que no existe un modelo experimental que los compruebe. La noción de que una inferencia causal proveniente de un EAC/RCT (siempre que este diseño metodológico sea posible) será preferible a cualquier otra evidencia, deja de lado a la mayor parte de las necesidades y posibilidades humanas para establecer inferencias causales. Por lo tanto, la humanidad necesita entender mejor los supuestos sobre los que se basa la inferencia causal para disponer de otras alternativas que permitan develar las relaciones causa-efecto que permitan abordar la tríada explicación-predicción-manipulación como recurso para la supervivencia y progreso de la especie.

Una vez reconocidas las características que se argumentan para reclamar la validez del EAC/RCT como el “estándar de oro” de la inferencia causal, es necesario también reconocer las características que algunos autores han venido poniendo en evidencia recientemente sobre las limitaciones de esta argumentación. Un resumen al respecto de las deficiencias de los EAC/RCT los presenta Stephen John Senn (@stephensenn) en twitter:

- *Pocos ensayos son completamente aleatorios, cuando las participantes se incorporan a partir de un listado previo*
- *No se produce muestreo aleatorio en los casos en que las participantes se enrolan según van llegando*
- *Las pacientes no necesariamente empiezan simultáneamente la exposición*
- *Los resultados binarios verdaderos son raros*
- *Los efectos del estudio sobre las participantes pueden ser importantes*
- *La estimación de la incertidumbre es vital*

Otras autoras señalan que: las mediciones que se realizan durante un EAC/RCT son rigurosamente “coreografiadas” por las investigadoras (incluso si el estudio es doble ciego), por lo tanto los efectos y la magnitud de los efectos están influenciados por la buena adherencia de las unidades de observación a las intervenciones, lo que es difícilmente replicable en la realidad cotidiana. En cambio, los estudios observacionales recolectan la información en los medioambientes naturales y realidades sociales de las unidades de observación y bajo sus conductas habituales.

Los beneficios de la aleatorización según Pearl (Pearl & Mackenzie, 2018, # 147) son 1) la eliminación del efecto de confusión de las otras variables incluidas (diferentes a la exposición y a la consecuencia) en el diseño de los estudios experimentales y 2) la cuantificación de la incertidumbre. Esta segunda característica ha sido la fortaleza más reclamada por quienes proclaman que el EAC/RCT es el estándar de oro de los

diseños experimentales en ciencias de la salud, para confirmar esta ventaja se ha inventado un amplio menú de pruebas estadísticas y variantes de la conformación de los grupos expuesto y no expuesto.

> *Contrariamente a las afirmaciones frecuentes en la literatura aplicada, la aleatorización no iguala todo lo que no sea el tratamiento en los grupos de tratamiento y control, no brinda automáticamente una estimación precisa del efecto promedio del tratamiento (Average Treatment Effect ATE) y no nos libera de la necesidad de pensar en las covariables (observadas o no observadas). Averiguar si una estimación se generó por casualidad es más difícil de lo que comúnmente se cree. En el mejor de los casos, un ECA/RCT produce una estimación imparcial, pero esta propiedad tiene un valor práctico limitado. Incluso entonces, las estimaciones se aplican solo a la muestra seleccionada para el ensayo, a menudo no más que una muestra de conveniencia, y se requiere una justificación para extender los resultados a otros grupos, incluida cualquier población a la que pertenezca la muestra del ensayo, o a cualquier individuo, incluido un individuo en la prueba. Exigir "validez externa" no es útil porque espera demasiado de un ECA/RCT mientras infravalora su contribución potencial. De hecho, los ECA/RCT requieren suposiciones mínimas y pueden operar con poco conocimiento previo. Esto es una ventaja cuando se trata de persuadir a audiencias desconfiadas, pero es una desventaja para el progreso científico acumulativo, donde debe construirse sobre el conocimiento previo, no descartarse. Los ECA/RCT pueden jugar un papel en la construcción de conocimiento científico y predicciones útiles, pero solo pueden hacerlo como parte de un programa acumulativo, combinándolo con otros métodos, incluido el desarrollo conceptual y teórico, para descubrir no "qué funciona", sino "por qué funcionan las cosas".* (Deaton & Cartwright, 2018, #)

El hecho de que los iniciadores del uso de los EAC/RCT como R.A. Fisher, no dispusieron en su tiempo de la notación matemática que les permita realizar el control de confusión, no les permitió articular la comprensión de lo que observaron con sus conclusiones. Una renovada comprensión de que el principal objetivo de los EAC/RCT es el control de las variables de confusión, permitió a Sander Greenland y Jamie Robins develar la falta de una definición ampliamente aceptada de ¿en qué consiste el fenómeno de la confusión en la relación entre las variables en los estudios científicos? (Pearl & Mackenzie, 2018, # 155). Entonces, surge la pregunta sobre otras alternativas de ajuste por variables de confusión que no sea la aleatorización del EAC/RCT, en especial en el caso de que la ejecución de un EAC/RCT no sea factible por múltiples situaciones que no permitan su aplicación y que en general son bien conocidas por productoras y consumidoras de productos científicos. Según Paerl (Pearl & Mackenzie, 2018, # 150) la introducción de una medida de intervención no siempre es posible y en ocasiones tiene serios cuestionamientos éticos.

> ***Judea Pearl [****@yudapearl*], *(2023, 21 de marzo) Descubrí que la "aleatorización" se entiende bien porque es fácil imaginar un proceso de selección de muestras basado en los resultados de [lanzamiento de] monedas y porque la aleatorización es la base de las encuestas estadísticas. Lo que es más difícil de darse cuenta para las personas es que los ECA invocan "intervenciones", además de la aleatorización. La idea de que los pacientes se ven obligados (o aconsejados) a actuar de acuerdo con el resultado de una moneda, posiblemente en contra de su inclinación natural, eleva los ECA al nivel 2 de la escalera y nos permite inferir efectos causales; la aleatorización es insuficiente. [tweet], Twitter.*

https://twitter.com/yudapearl/status/1637877682573824000?t=Cd0emfGxB17C2264aDGGRw&s=03

Por otro lado, cada vez es más evidente que el diseño en general de un EAC/RCT está fuertemente influenciado por la definición del efecto/consecuencia/resultado con el que se diseñe un experimento. Muchos elementos de la relación causal se incluirán y definirán en función de la definición de la variable de efecto. En cuanto al argumento de que los EAC/RCT permiten una cuantificación precisa del efecto causal, también existe una discrepancia entre las autoras contemporáneas:

> *Las investigadoras entusiastas de los ensayos experimentales nunca nos dicen cuál es el efecto causal. Simplemente dicen que todo lo que evalúan los ensayos clínicos es un efecto causal. @yudapearl primero define el efecto causal y luego muestra que tipo de ensayo ideal puede evaluarlo. Boris Sobolev @soboleffspaces*
>
> *No conozco ninguna cantidad causal que permanezca indefinida en un SCM, suponiendo, por supuesto, que tomemos la relación "escucha a" como una forma primitiva para definir SCM. En cuanto a los ECA/RCT, la prueba en #Bookofwhy es la única prueba defendible que conozco de qué tipo de RCT ofrece (asintóticamente) el efecto causal promedio (ACE) del tratamiento en el resultado. De hecho, es extraño que ACE no se trate en la literatura de los ECA/RCT como el objetivo final del ejercicio de un ECA/RCT. Tomar la "comparación" como el objetivo de los ECA/RCT difícilmente justifica su estatura de "estándar de oro", sin mencionar sus costos. @yudapearl*
>
> *Se encontraron áreas de fortaleza metodológica (buena asignación al azar y ocultación de la asignación), pero áreas de debilidad con respecto al cegamiento de los participantes, las personas que administraron la intervención y los evaluadores de resultados. En un tercio de los ECA se observó una deserción sustancial (pérdidas por el control del punto de tiempo), lo que potencialmente conduce a un poder estadístico insuficiente para obtener estimaciones precisas del efecto de la intervención sobre los resultados primarios. La mayoría de los ECA mostraron que las intervenciones de alfabetización en salud tuvieron algún efecto beneficioso sobre los resultados del conocimiento, pero esto fue típicamente por menos de 3 meses después del final de la intervención. Hubo muchos menos informes de mejoras significativas en los resultados sustantivos orientados al paciente, como los efectos beneficiosos sobre el cambio de comportamiento o el estado de salud (clínico). La mayoría de los ECA incluyeron participantes de poblaciones vulnerables.* (Brainard et al., n.d., # 246)

Las defensoras de la calidad de estándar de oro de los EAC/RCT como evidencia experimental de la causalidad argumentan que el propósito de este tipo de estudios es comparar los resultados de un tratamiento en un grupo expuesto con los resultados en un grupo no expuesto a una intervención específica. Los resultados se definen en función de la modificación que se produce en las unidades de observación expuestas a un tratamiento definido; por lo tanto, la naturaleza de estos experimentos es estimar un valor en cada grupo y que la diferencia entre los grupos sea lo suficientemente importante como para atribuirle a la intervención. Argumentan que estas diferencias son el sustrato del que se ocupa la clínica, si el tratamiento en experimentación es superior a otro tratamiento o a ningún tratamiento en un grupo de unidades de observación.

No es el caso de los estudios en los que interesa saber si el resultado de una intervención es diferente entre los grupos expuesto y no expuesto y que además el valor de la diferencia puede ser extrapolado a toda la población que comparte las características del grupo intervenido. El efecto promedio del tratamiento (average treatment effect, ATE/EPT) que se obtiene mediante el análisis estadístico constituye el objetivo de muchos estudios no experimentales. Actualmente surge la discusión sobre los mecanismos tanto lógicos como estadísticos para asignar este valor promedio del grupo a cada unidad de observación participante en un estudio, o más aún, cómo hacerlo para individuos de la población general o de poblaciones similares a las que provienen las unidades de observación de un estudio.

> *Uno de los mantras menos discutidos de la inferencia causal es que no podemos acceder a efectos causales individuales; podemos observar una respuesta individual al tratamiento o al no tratamiento, pero nunca a ambos. En cambio, la mayoría de las investigaciones sobre inferencia causal se han centrado en el ATE, una cantidad de población que se puede estimar directamente a partir de ECA, pero que no proporciona información sobre cómo responderían los individuos que responden de una manera bajo un tratamiento bajo una alternativa. Nuestros resultados teóricos muestran que podemos ir más allá de ATE, para estimar (o limitar) la distribución completa de los efectos causales individuales. Los límites estimados pueden ser bastante estrechos y permitirnos tomar decisiones personalizadas precisas. En otras palabras, un individuo elegido al azar sería capaz de evaluar cómo respondería tanto al tratamiento como a su negación y, en consecuencia, decidir el curso de acción que mejor se ajuste tanto a sus preferencias personales como a sus necesidades sociales.* (Mueller & Pearl, 2023, #)

También preocupa a las investigadoras actuales la forma de combinar los resultados de los estudios observacionales con los resultados de EAC/RCT, para de esta forma profundizar las relaciones causales a partir de estudios no experimentales, disminuir la dependencia de estudios experimentales y aprovechar las grandes cantidades de datos provenientes de sistemas de registro contínuo o de bases de datos disponibles sobre muchas conductas humanas.

Pearl

> *Como veterano escéptico de la supremacía del ECA/RCT, doy la bienvenida al desafío de D&C de todo corazón. De hecho, The Book of Why (Pearl and Mackenzie, 2018, http://bayes.cs.ucla.edu/WHY/) me cita diciendo: "Si nuestra concepción de los efectos causales tuviera algo que ver con los experimentos aleatorios, estos últimos habrían sido inventado 500 años antes que Fisher." En este, así como en mis otros escritos, llego a afirmar que el ECA/RCT gana su legitimidad al imitar al "operador-do" [do-operator], y no al revés. Además, considerando las dificultades prácticas de realizar un ECA/RCT ideal, los estudios observacionales tienen una ventaja definitiva: interrogan a las poblaciones en sus hábitats naturales, no en entornos artificiales coreografiados por protocolos experimentales.*
>
> *El desafío de Deaton y Cartwright a la supremacía del ECA/RCT consta de dos partes: la primera (validez interna) trata sobre la maldición de la dimensionalidad y argumenta que, en cualquier ensayo individual, el resultado del ECA/RCT puede estar bastante lejos de la cantidad causal objetivo , que suele ser el efecto de tratamiento promedio (ATE). En otras palabras, esta parte se refiere al desequilibrio debido a muestras finitas y refleja la compensación tradicional de precisión de sesgo en el análisis estadístico y el*

aprendizaje automático. La segunda parte (validez externa) se ocupa de los sesgos creados por las inevitables disparidades entre las condiciones y poblaciones en estudio versus las que prevalecen en la implementación real del programa o política de tratamiento [intervención].

Los investigadores tardaron más de una docena de años en aceptar la noción de completitud en el contexto de la validez interna, tal como surgió del do-cálculo (ver Pearl (1995); Shpitser y Pearl (2008); Tian y Pearl (2002)). Aquí, la completitud nos dice qué supuestos son absolutamente necesarios para la identificación no paramétrica de los efectos causales, cómo saber si se satisfacen en cualquier descripción específica del problema y cómo usarlos para extraer parámetros causales de estudios no experimentales.

La integridad en el contexto de la validez externa es un resultado relativamente nuevo (ver Bareinboim y Pearl (2013)), que probablemente tomará algunos años más para que los investigadores ilustrados lo acepten, aprecien y utilicen por completo. Uno de los propósitos de este comentario es instar a la comunidad investigadora, especialmente a Deaton y Cartwright, a estudiar la reciente matematización de la validez externa y beneficiarse de sus implicaciones. (Pearl, 2018, #)