

Estudios aleatorizados, implementación y qué hacer con potenciales hallazgos

He solicitado al Dr Jorge Thierer su opinion sobre tomar conductas diagnósticas cuando no hay estudios aleatorizados acerca de su implementación, o acerca de qué hacer con los potenciales hallazgos.

Gentilmente me ha enviado su respuesta y considero oportuna su divulgación a través del espacio del consejo de enfermedad de Chagas de la Sociedad Sudamericana de cardiología

Un comentario general sobre estudios aleatorizados y registros

Cuando se clasifica la evidencia sobre la que se formulan recomendaciones en las guías de práctica clínica, la mayor fuerza de evidencia, rotulada como A, es la que corresponde a datos derivados de al menos 2 estudios aleatorizados o meta análisis. La evidencia B es la que se desprende de un solo estudio aleatorizado o estudios no aleatorizados de grandes dimensiones (cohorte, caso control), y la C la que corresponde a consenso de opinión de expertos, reporte de casos o estándar de cuidado. Está instalada, entonces, ya desde la clasificación, la preeminencia de los estudios de intervención aleatorizados sobre los estudios observacionales.

Estudios aleatorizados

Sin duda la gran ventaja de los estudios aleatorizados radica en que justamente *la asignación aleatoria de las intervenciones asegura la distribución balanceada de las características basales conocidas y desconocidas, y elimina el sesgo de selección*. Si se demuestra un efecto estadísticamente significativo de la nueva intervención respecto de su comparador (placebo, otra intervención), el mismo podrá ser atribuido a la intervención, y no a otros factores (confundidores) que por estar desigualmente repartidos entre ambas ramas pueden ser los verdaderos responsables de la diferente evolución.(1)

Además de eliminar el sesgo de selección los estudios aleatorizados tienen una serie de ventajas: *diseño y análisis rigurosos*; en muchos casos la instauración del *ciego* (doble ciego, simple ciego, e incluso la adjudicación ciega de eventos por personal no involucrado en las acciones cotidianas del estudio) *que evita el sesgo de observación*; *el control* (con placebo, cuando es posible, para descartar el efecto placebo que toda intervención acarrea, y si no lo es, con un control preferentemente concurrente); y el *seguimiento estrecho* de todos los incluidos para evitar pérdida de pacientes y por ende el sesgo de seguimiento diferencial.(2)

Pero es justo reconocer en los estudios aleatorizados un conjunto de “puntos débiles”. Por empezar, *las instituciones, médicos y pacientes que participan* de los mismos tienen la voluntad de participar y aceptan el proceso de asignación aleatoria y los controles frecuentes.

No son por lo tanto *similares a aquellos que por diferentes razones no lo hacen* (limitaciones por disponibilidad de tiempo o cercanía, deseo tanto del médico como del paciente de instaurar o no instaurar un tratamiento determinado). Y a ello se suma que en general los estudios aleatorizados tienen criterios de inclusión y exclusión estrictos, razón por la cual el grueso de los pacientes considerados inicialmente para participar terminan no haciéndolo. De allí que son estudios de *baja validez externa*. Hay por otra parte un presupuesto al que deben ceñirse, lo cual impone limitaciones económicas y de duración en el seguimiento. Y, por último, no son útiles para definir la incidencia de eventos o efectos adversos muy poco frecuentes.(3)

Algunas dudas surgen naturalmente al examinar la lógica de implementación de los estudios aleatorizados. Es cierto sí que la aleatorización elimina el sesgo de selección. Pero no es menos cierto es que *se selecciona la población que participará del ensayo*. Y esa población seleccionada es sin duda aquella donde *la intervención tiene más probabilidad de demostrar su efecto beneficioso*, si lo hubiera. Procedimientos de control y seguimiento estrecho, pacientes que acuden a todas las visitas, y libres de comorbilidades que amenacen la evolución y el cumplimiento más allá de la patología de interés, análisis frecuentes, etc. Los centros, procedimientos, médicos y pacientes de la vida diaria en general no aseguran tanta perfección.

La postura oficial es que un ensayo aleatorizado parte a priori de un nivel de evidencia alto. El instrumento GRADE plantea sin embargo una serie de puntos que deben ser tenidos en cuenta a la hora de juzgar la calidad de la evidencia. La falta de ocultamiento a la hora de asignar la intervención, la falta de ciego, el reporte selectivo de eventos, la pérdida en el seguimiento, la detención adelantada del estudio, son todas condiciones que disminuyen la fuerza de la evidencia. La inconsistencia o heterogeneidad en los hallazgos (efectos notables para algún punto final y no para otros, o muy diferente fuerza de asociación con diferentes resultados), la imprecisión en los resultados (intervalos de confianza que incluyen el 1 entre los valores posibles de un RR o un OR, o que incluyen el 0 cuando nos referimos a la diferencia en una variable continua) son otros criterios que restan credibilidad a las conclusiones.(4)

Estudios observacionales

Los estudios observacionales, en los que se comparan pacientes efectivamente tratados de diferente manera, pero en los que la indicación no ha sido aleatoria, merecen, como ya dijimos, una serie de críticas. Son *estudios sujetos a gran cantidad de sesgos*: sesgo de selección, de observación e inmortalidad (para empezar a recibir una medicación es necesario estar vivo; si se adjudica a los que reciben dicha intervención el tiempo desde que ingresaron a una cohorte hasta que iniciaron el tratamiento, se sobreestima el beneficio de la misma). Hay en estos estudios *confusión por indicación* (se indica determinada conducta o tratamiento en los que pueden recibirla, y por tanto tienen condiciones que se vinculan con la evolución más allá de lo que se haga), *confusión residual* (factores que influyen en el pronóstico y que no son tomados en cuenta en el análisis de las características basales). Por supuesto que también en el caso de los estudios observacionales hay limitaciones económicas: mantener un registro aceitado no es sencillo y demanda de grandes esfuerzos. La vinculación de diferentes bases

administrativas es una forma de implementar un seguimiento adecuado de los pacientes, pero por su naturaleza puede haber errores en la forma de caratular diversos eventos.(5)

Los estudios observacionales tienen desde ya una serie de ventajas: *alta validez externa*, definen la adherencia real a una estrategia o intervención, *mayor duración de seguimiento* y detección de eventos y efectos adversos poco frecuentes.(2)

En los últimos años se ha revalorizado el papel de los registros de pacientes. Se define a un registro como un sistema organizado que usa métodos de estudio observacional para recolectar datos uniformes (clínicos y otros) para evaluar resultados específicos para una población definida por una enfermedad, afección o exposición particulares y que sirve a uno o más propósitos predeterminados científicos, clínicos o de política sanitaria. La recolección de los datos se hace con un propósito definido, con definiciones pre-especificadas de datos coleccionados de manera uniforme, que reflejan el estado clínico del paciente. Al menos un elemento del registro es activo. Pueden enriquecerse al vincularse con otras bases de datos; en general son estudios de cohorte, pero pueden albergar otro tipo de estudios. Son esenciales para describir la historia natural de la enfermedad, definir efectividad y costo efectividad en el mundo real, definir seguridad y daño potencial de las intervenciones, describir calidad de cuidado y barreras en la atención y son fuente de datos para la investigación de resultados centrados en el paciente y, dato no menor, base para la realización de estudios de intervención.(6,7) Si se va a considerar solo la información proveniente de estudios observacionales, algunas aproximaciones estadísticas pueden intentar salvar las limitaciones mencionadas: análisis multivariado, puntaje de propensión (8), diseños de nuevo usuario, modelos estructurales marginales, variables instrumentales (9). Indudablemente el surgimiento del concepto de *“big data”*, con la posibilidad de manejar y analizar simultáneamente datos poblacionales de gran cantidad de bases de distinta índole, administrativas, de estudios y análisis, prescriptivas, de hospitalización, etc., implicará un avance notable y que por ahora solo vislumbramos, en nuestra capacidad de emplear datos observacionales y deducir de ellos reglas predictivas que permitan acercarnos a una medicina de precisión. (10)

¿Qué hacer frente a esta contradicción entre estudios observacionales y aleatorizados, y cómo potenciar las ventajas de cada uno? La realización de estudios aleatorizados pragmáticos, que conserven la aleatorización pero con un diseño simple y con criterios de inclusión y exclusión más laxos, y que reproduzcan las condiciones normales de seguimiento de los pacientes con la frecuencia habitual de visitas y prácticas es una opción.(11) Otra, la realización de estudios aleatorizados basados en registros, donde en principio todos los incluidos pueden ser sometidos al estudio con amplia validez externa. (12)

De acuerdo a las palabras de Sir Michael Rawlins, cabeza de NICE: “Experimento, observación y matemáticas individual y colectivamente tienen un rol crucial en proveer la base de evidencia para la moderna terapéutica. Argumentos sobre la importancia relativa de cada uno son una distracción innecesaria. Las jerarquías de evidencia deberían ser reemplazadas por aceptar e incluso abrazar una diversidad de aproximaciones”. (13) Observación y experimentación deben ser vistas como aliadas, y no como enemigas. Cada una potencia las ventajas de la otra, y disimula al menos en parte sus flaquezas. (14)

La respuesta específica a la pregunta del Dr. Mitelman.

En el mail, que tanto agradezco, en el que me invita a formular algunas apreciaciones sobre este tema, el Dr. Mitelman plantea la oposición de argumentos acerca de tomar conductas diagnósticas cuando no hay estudios aleatorizados acerca de su implementación, o acerca de qué hacer con los potenciales hallazgos.

Vayamos por partes. La inmensa mayoría de las conductas que tomamos a diario en la práctica médica no se basan en estudios aleatorizados. Cuando nos remitimos a las guías de práctica sucede algo similar. De 133 recomendaciones que formula la guía de práctica de Insuficiencia Cardíaca de la Sociedad Europea de Cardiología de 2016 para la insuficiencia cardíaca crónica (15), solo el 24,8% son A; el 30,8% son B y el 44,4 % son C. Y esto es particularmente notable cuando nos referimos a conductas diagnósticas. Si nos remitimos a este apartado en el Consenso de Insuficiencia Cardíaca de la Sociedad Argentina de 2016, (16) de 19 recomendaciones, 13 son C y 6 son B. No hay recomendaciones sobre diagnóstico con evidencia A. Es decir, que en general las conductas en este campo son en gran parte empíricas; no se sustentan ni siquiera en estudios observacionales. Pareciera no haber verdadero interés en encontrar justificación firme para la mayor parte de nuestras prácticas.

Muchas veces el examen cuidadoso de los datos de que se dispone puede darnos sorpresas. Por ejemplo, la realización de un ECG basal, o incluso de un ECG de esfuerzo, es parte habitual de la práctica cardiológica. Muchos médicos consideran que la información que brinda complementa la que entregan las variables clínicas. Una revisión sistemática y categorización de la evidencia disponible formulada a pedido de la Fuerza de Tareas de Servicios Preventivos de los Estados Unidos (USPSTF por sus siglas en inglés) acaba de publicarse. Y, para gran sorpresa de todos, resulta que no hay estudios que hayan explorado la utilidad del ECG de esfuerzo en personas de bajo riesgo; y en aquellos de riesgo moderado a alto no hay evidencia firme de repercusión favorable, y la mejoría de los modelos pronósticos tradicionales con el agregado de datos ECG es al menos discutible.(16) De allí que en estos pacientes se mantenga que no hay información consistente. Es tremendamente ilustrativo que con la cantidad incontable de ECG que se llevan a cabo en millones y millones de personas cada año, preguntas sencillas sobre su utilidad no puedan ser respondidas. Este es solo un ejemplo de cómo conductas que entendemos “indiscutibles” pueden ser sometidas a escrutinio, y no encontrar evidencia clara para implementarlas.

Podemos sí sostener que parece excesivo entender como “evidencia” solo lo que proviene de ensayos aleatorizados. En realidad, la Medicina Basada en la Evidencia fue, en sus orígenes, la idea de basar nuestra conducta en la mejor evidencia disponible (17). Esa información puede provenir entonces de estudios aleatorizados u observacionales. La asignación aleatoria, en virtud de todo lo expuesto, no es garantía indiscutible de evidencia fuerte. Hay estudios aleatorizados de mejor o peor calidad, más o menos creíbles, con conclusiones más o menos extrapolables a todos nuestros pacientes. En el otro extremo, hay estudios observacionales de mejor o peor calidad, y sobre los que podemos formular las mismas apreciaciones. Así como la aleatorización por sí sola no es garantía indiscutible, tampoco es la observación, si no es rigurosa, completa, y sólidamente constituida. Cada una de las fuentes de información debe ser adecuadamente evaluada, analizada y juzgada. Y seguiremos discutiendo hasta el hartazgo

cada aseveración; muchas verdades inconvencibles que se sostuvieron durante décadas se han derrumbado ante la aparición de nuevos hallazgos.

Pero a lo que no se debe renunciar de ninguna manera es al imperativo de tener siempre en cuenta la mejor información, la más pura y menos sesgada. En países como los nuestros, creo que en muchos aspectos estamos en una etapa previa de la discusión. No se trata del antagonismo entre datos de estudios observacionales o aleatorizados; *se trata de tener datos*. Es la falta de un registro elemental de lo que se hace (en forma integral, no aquello que surge de iniciativas parciales y bienintencionadas) el talón de Aquiles inicial de nuestra práctica. A remediar este pecado original deberían dirigirse nuestros esfuerzos. Si la necesidad de disponer de información se hace carne en nuestras sociedades, es de esperar que lo que sigue se dará por añadidura.

Jorge Thierer

Jefe de la Unidad de Insuficiencia Cardíaca de CEMIC

Director Asociado de la Revista Argentina de Cardiología.

Bibliografía

1. Diamond GA. Randomized trials, observational registries, and the foundations of evidence-based medicine. *Am J Cardiol* 2014;113:1436-41.
2. Hannan EL. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC Cardiovasc Interv* 2008;1:211-7.
3. Faraoni D, Schaefer ST. Randomized controlled trials vs. observational studies: why not just live together? *BMC Anesthesiol* 2016;16:102.
4. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-6.
5. Ieva F, Gale CP, Sharples LD. Contemporary roles of registries in clinical cardiology: when do we need randomized trials? *Expert Rev Cardiovasc Ther* 2014;12:1383-6.
6. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ* 2016;352:i493.
7. In: Glicklich RE, Dreyer NA, Leavy MB, editors. *Registries for Evaluating Patient Outcomes: A User's Guide*. Rockville (MD), 2014.
8. Kitsios GD, Dahabreh IJ, Callahan S, Paulus JK, Campagna AC, Dargin JM. Can We Trust Observational Studies Using Propensity Scores in the Critical Care Literature? A Systematic Comparison With Randomized Clinical Trials. *Crit Care Med* 2015;43:1870-9.
9. Hlatky MA, Winkelmayr WC, Setoguchi S. Epidemiologic and statistical methods for comparative effectiveness research. *Heart Fail Clin* 2013;9:29-36.
10. Van Poucke S, Thomeer M, Hadzic A. 2015, big data in healthcare: for whom the bell tolls? *Crit Care* 2015;19:171.
11. Ford I, Norrie J. Pragmatic Trials. *N Engl J Med* 2016;375:454-63.
12. Jones WS, Roe MT, Antman EM, Pletcher MJ, Harrington RA, Rothman RL et al. The Changing Landscape of Randomized Clinical Trials in Cardiovascular Disease. *J Am Coll Cardiol* 2016;68:1898-1907.
13. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet* 2008;372:2152-61.

14. Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer* 2014;110:551-5.
15. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;37:2129-2200.
16. Marino J, Barisani J, Thierer J, Liniado G, Pereiro S, Francesia A y col. Consenso de Insuficiencia Cradíaca Crónica de la Sociedad Argentina de Cardiología. *Rev Argent Cardiol* 2016; 84: supl 3
16. Jonas DE, Reddy S, Middleton JC, Barclay C, Green J, Baker C et al. Screening for Cardiovascular Disease Risk With Resting or Exercise Electrocardiography: Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA* 2018;319:2315-2328.
17. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312:71-2.