

RESPUESTA

Agradezco enormemente las apreciaciones emitidas por el colega Dr. Álvaro Javier Idrovo, con respecto al artículo “La cúspide de la pirámide de la evidencia” publicada en el reciente número de *Acta Médica Colombiana*, puesto que me permite hacer algunas aclaraciones, que por lo extenso del tema era imposible abarcar.

La pirámide de la *medicina basada en la evidencia* es una representación gráfica propuesta para comprender los diversos niveles de la evidencia científica que permite dar recomendaciones en salud. Su propósito es destacar las fortalezas y debilidades de cada uno de los diseños, especialmente la incertidumbre generada en sus resultados, puesto que es imposible conocer con 100% de certeza la “verdad”. Por lo tanto, no debe concebirse como una estructura rígida, aunque sí sólida, que descarta de plano el conocimiento generado por estudios preclínicos o por estudios con diseños menos robustos, puesto que estos son la base para generar una hipótesis que indefectiblemente debe ser probada en un ensayo clínico, en la mayoría de los casos.

La crítica más importante que se le ha hecho a esta estructura es el excesivo protagonismo del diseño de metaanálisis y a las revisiones sistemáticas, ubicándolo en la cúspide de la evidencia (1) como fuente para tomar decisiones en salud, a pesar de ser un estudio secundario. Actualmente, se sugiere utilizarlos como una lente con el cual observar y analizar detenidamente el cuerpo de la evidencia científica, dándole mayor relevancia a los estudios primarios, especialmente a los ensayos clínicos, concepto que intenta esbozar la Figura 2 del artículo. Se han propuesto modificaciones a esta estructura con diversas estrategias, siendo una de las más recientes la que propone asumirlo como una rueda que represente la totalidad de la evidencia (2). Sin embargo, no logra ser convincente dado lo anteriormente expuesto.

El diseño de metaanálisis ha tenido muchas críticas bastante conocidas en epidemiología clínica, tales como que un número no puede resumir todo un campo de investigación, combinación de estudios heterogéneos (mezcla de peras con manzanas); basura entra, basura sale, la exclusión de estudios importantes y el sesgo de publicación que invalida los resultados, así como una mala conducción metodológica. Aunque la mayoría de estas críticas no se ajusta a la realidad, este tipo de estudios debe afrontar el problema de la heterogeneidad y dilemas metodológicos, por lo que incluso podrían considerarse como estudios observacionales cuyos resultados son interpretados como asociaciones más que efecto causal (3). A pesar de estos hechos incontrovertibles, son muy importantes en la definición de la dirección y fortaleza de las recomendaciones en las guías de práctica clínica.

En epidemiología clínica, es ampliamente conocido el estudio de LeLorier J et al (4), publicado en *The New England Journal of Medicine*, que reportó una baja concordancia entre los desenlaces primarios y secundarios entre metaanálisis

y los ensayos clínicos subsecuentes. Se encontró un valor predictivo positivo de 68% y negativo de 67%, lo que implica que, en hasta en el 35% de los casos, los resultados de metaanálisis no predicen los del ensayo clínico. Este dato puede interpretarse desde distintas perspectivas, algunos lo utilizarán para demostrar la importancia de los metaanálisis, mientras que otros por el contrario buscarán resaltar su poca fiabilidad, algo similar a ver el vaso medio lleno o medio vacío. Otros estudios demuestran resultados similares, lo que sugiere que la evidencia extraída de metaanálisis debería considerarse como generadora de hipótesis, la cual debe ser comprobada en un ensayo clínico, como es usual en el ámbito científico (5).

El sistema GRADE orienta sobre el uso de diversos tipos de estudios en forma individual para ser incluidos en la síntesis de la evidencia en guías de práctica clínica. Considera que los ensayos clínicos proporcionan la mejor fuente de evidencia sobre los efectos de una intervención para definir directrices en salud (6). Los estudios no aleatorizados, representativos de la población, son considerados la mejor evidencia en cuanto a pronóstico, riesgo basal, exactitud de pruebas diagnósticas y para estimar la utilidad de los valores y preferencias de los pacientes. No obstante, esto no quiere decir que, en intervenciones, en algunos escenarios específicos, puedan reemplazar a los ensayos clínicos, por ejemplo cuando no los hay, o ser secuenciales aportando información valiosa o aún complementaria (6).

Por esta razón, en las diversas guías de práctica clínica el nivel de evidencia y su fortaleza va de la mano del análisis de ensayos clínicos con asignación aleatoria. Su permanencia en versiones subsecuentes de las guías depende precisamente del diseño del estudio. Un análisis de las guías del *American College of Cardiology/American Heart Association* demostró que de las recomendaciones clase I soportadas por múltiples ensayos clínicos el 90.5% (95%CI, 83.2%-95.3%) son retenidas en las siguientes versiones, en comparación con el 81.0% (95%CI, 74.8%-86.3%) de las soportadas por un solo ensayo clínico o datos observacionales (7). La probabilidad de que las recomendaciones basadas en opinión de expertos, un solo ensayo o datos observacionales sean degradadas, reversadas u omitidas es alta (Odds ratio 3.49; 95% CI, 1.45-8.41; P = 0.005). La opinión de expertos es la que menor confiabilidad tiene por estar plagada de sesgos y conflictos de interés.

Existe un viejo adagio que reza: “El sentido común es el menos común de los sentidos”, y en evidencia científica sí que es cierto. La literatura cardiovascular ofrece múltiples ejemplos que lo demuestran (8). Por ejemplo, el magnesio, en estudios experimentales parecía limitar el tamaño del infarto de miocardio, generando un menor flujo de calcio, lo que en teoría disminuía la posibilidad de arritmias, incluso ensayos clínicos pequeños y sus respectivos metaanálisis demostraron que podía reducir la mortalidad casi a la mitad.

Sin embargo, el ensayo clínico ISIS 4 de 58.050 pacientes, publicado en la revista *Lancet* en 1995 demostró que no hubo reducción de la mortalidad en el grupo total ni en ningún subgrupo analizado (tratados en forma temprana o tardía, en presencia de trombolisis o antiplaquetarios, o de alto riesgo) (9). Otro caso es el levosimendan, un inotrópico con mecanismo de acción novedoso por ser sensibilizador de los miofilamentos al calcio, demostró en ensayos pequeños previos como el LIDO (10) reducción de la mortalidad a 180 días como desenlace secundario al compararse con tratamiento con dobutamina. Sin embargo, el subsecuente estudio confirmatorio SURVIVE, publicado en *JAMA* en 2007 con 1327 pacientes hospitalizados por falla cardiaca aguda descompensada no logró demostrar diferencias entre el manejo con levosimendán versus dobutamina (11).

Finalmente, la calidad de los estudios es importante para confiar en sus resultados. Es preferible un estudio observacional bien conducido que un ensayo clínico sesgado o un metaanálisis desarrollado por investigadores con escasa experticia en su metodología. De ahí la insistencia dentro del artículo en el análisis crítico de ellos para descubrir posibles inconsistencias, análisis inadecuados o limitaciones no reportadas o no expresadas abiertamente por los autores.

Ioannidis P. ha insistido desde hace algún tiempo en la conducción de estudios en los que se pueda confiar y ha señalado algunos puntos cruciales a tener en cuenta, denominando su artículo en forma disruptiva como “Por qué la mayoría de los hallazgos de investigación publicados son falsos” (12).

En el artículo mencionado por el colega, del grupo del mismo autor, debe analizarse con detenimiento, puesto que sugiere exactamente lo opuesto a lo expresado al resaltar que “dado que los ensayos pequeños y sus metaanálisis pueden dar estimaciones poco fiables e infladas de beneficio, es posible que sea necesario realizar megaensayos o, al menos, considerar y realizar con mayor frecuencia ensayos sustancialmente grandes y con potencia suficiente” (discusión página 11 primer párrafo) (13).

El secreto radica en definir qué se considera un ensayo pequeño (14), y los propios autores confiesan que los resultados están afectados por el relativo gran tamaño de muestra de sus “ensayos pequeños”. Es indudable que el cálculo inadecuado de la muestra afecta la potencia (15).

Superamos recientemente una etapa de oscurantismo científico durante la pandemia de COVID-19, en la que se recomendaron terapias ineficaces y algunas hasta dañinas, como hidroxicloroquina, ivermectina, claritromicina, remdesivir y otras tantas, basadas en estudios preclínicos,

observacionales y metaanálisis espurios, que luego fueron retractados o desmentidos gracias a las publicaciones de los ensayos adaptativos del grupo colaborativo RECOVERY.

Por ello, mi invitación es a abrir la pirámide en su cúspide, sin dejarla caer, y darle el verdadero valor a los ensayos clínicos.

Dr. Juan Manuel Senior, MD, MSc

Internista. Cardiólogo intervencionista, Magíster en Epidemiología Clínica, Universidad de Antioquia. Jefe de Posgrado Cardiología Clínica y Cardiología Intervencionista, Hospital Universitario San Vicente Fundación. Medellín (Colombia).

E-Mail: juan.senior64@gmail.com

Referencias

1. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med.* 2016;21(4):125-7.
2. Aldous C, Dancis BM, Dancis J, Oldfield PR. Wheel Replacing Pyramid: Better Paradigm Representing Totality of Evidence-Based Medicine. *Ann Glob Health.* 2024;90(1):17.
3. Berlin JA, Golub RM. Meta-analysis as evidence: building a better pyramid. *JAMA.* 2014;312(6):603-5.
4. LeLorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between Meta-Analyses and Subsequent Large Randomized, Controlled Trials. *N Engl J Med.* 1997;337:536-542.
5. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. Examples from the management of acute myocardial infarction. *Ann Intern Med.* 1995;123(11):873-7.
6. Cuello-García C, Santesso N, Morgan RL, Verbeek J, Thayer K, Ansari MT et al. GRADE guidance 24 optimizing the integration of randomized and non-randomized studies of interventions in evidence syntheses and health guidelines. *J Clin Epidemiol.* 2022;142:200-208.
7. Neuman MD, Goldstein JN, Cirullo MA, Schwartz JS. Durability of class I American College of Cardiology/American Heart Association clinical practice guideline recommendations. *JAMA.* 2014;311(20):2092-100.
8. Fonaroff AC, Califf RM, Harrington RA, Granger CB, McMurray JJV, Patel MR et al. Randomized Trials Versus Common Sense and Clinical Observation: JACC Review Topic of the Week. *J Am Coll Cardiol.* 2020;76(5):580-589.
9. ISIS 4 collaborative group. ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. *Lancet.* 1995;345(8951):669-85.
10. Follath F, Cleland JGF, Just H, Papp JGY, Scholz H, Peuhkurinen K et al. Efficacy and safety of intravenous levosimendan compared with dobutamine in severe low-output heart failure (the LIDO study): a randomised double-blind trial. *Lancet.* 2002;360(9328):196-202.
11. Mebazaa A, Nieminen MS, Packer M, Cohen-Solal A, Kleber FX, Pocock SJ et al. Levosimendan vs dobutamine for patients with acute decompensated heart failure: the SURVIVE Randomized Trial. *JAMA.* 2007;297(17):1883-91.
12. Ioannidis P. Why most published research findings are false. *PLoS Med.* 2005;2(8):e124.
13. Kastrati L, Raesi-Dehkordi H, Llanaj E, Quezada-Pinedo HG, Khatami F, Ahanchi NS, et al. Agreement between mega-trials and smaller trials: a systematic review and meta-research analysis. *JAMA Netw Open.* 2024;7(9):e2432296.
14. Hackshaw A. Small studies: strengths and limitations. *European Respiratory Journal.* 2008; 32(5): 1141-1143.
15. Pocock S, Clayton TC, Stone GW. Design of Major Randomized Trials: Part 3 of a 4-Part Series on Statistics for Clinical Trials. *J Am Coll Cardiol.* 2015 Dec 22;66(24):2757-2766.

