

Comentario

¿Conocimiento No Comunicado? Interpretación y Aplicación de los Resultados de los Ensayos Clínicos y del Meta-Análisis

George Davey Smith y Matthias Egger

DEPARTAMENTO DE MEDICINA SOCIAL, UNIVERSIDAD DE BRISTOL, BRISTOL, REINO UNIDO

Las evaluaciones de las terapéuticas médicas deben enfrentarse a la distinción entre principios generales, los beneficios (o daños) que producen los tratamientos a nivel de grupo y las decisiones acerca del cuidado clínico particular que debería recibir un paciente determinado. Esta tensión se puede percibir en las discusiones surgidas en el momento de la introducción de métodos formales de evaluación de cuidados sanitarios. Los enfoques estadísticos aplicados en medicina, que a principios del siglo diecinueve aparecían referidos como el "método numérico"[1], fueron considerados como un ataque al "arte" de la medicina y algo que apartaba a los médicos de su propia tarea, que era adaptar el tratamiento a las características de los individuos particulares que constituían sus pacientes. En 1835, Jean Civiale utilizó el método numérico para resaltar que en el tratamiento de las piedras en la vejiga, la litotricia (trituration y expulsión) era superior a la litotomía (extracción quirúrgica). Murieron seis (2,3%) de los 257 pacientes tratados con litotricia, en comparación con los 1.141 (20%) muertos entre los 5.715 pacientes del grupo control que fueron tratados con litotomía. Estos resultados no convencieron a una comisión de la Academia de Ciencias encargada de revisar ese trabajo[2]. Un miembro de la comisión, el clínico François Joseph Double, dijo:

En asuntos estadísticos, cuando se intenta apreciar los hechos de forma numérica, la primera preocupación por encima de todo es dejar al individuo fuera de lugar, considerándolo meramente como una fracción de su especie. Se le debe privar de su individualidad para eliminar cualquier rasgo de la misma que pudiera introducirse accidentalmente en la pregunta de interés. Por el contrario, en la medicina aplicada, el problema siempre es individual, los hechos que contribuyen a solucionarlo se presentan uno por uno; nosotros nos enfrentamos exclusivamente a la personalidad del paciente y en último término a un ser humano particular con toda su idiosincrasia al que el médico debe tratar. Para nosotros las masas quedan completamente fuera de la cuestión[2].

El punto de vista que aquí se comenta tiene una larga historia en la medicina. Tal como escribió antes de Double, hace ya más de 500 años, el cirujano Henri de Mondeville, "Cualquiera que crea que se puede acomodar la misma cosa a cualquier persona es un inmenso tonto, ya que la medicina no se practica sobre el género humano en general, sino sobre cada individuo en particular" [3]. El papel aparentemente menor que juegan las habilidades clínicas de acuerdo con los que proponen la evaluación estadística de la eficacia del tratamiento, visto como un fallo para comprender la importancia del "conocimiento no comunicado" en la práctica clínica[4], es un objetivo de ataque ahora por los oponentes a la actual moda de la medicina basada en la evidencia [5]. Identificar cuando es apropiado generalizar desde lo abstracto al individuo particular constituye el problema central de cualquier forma de práctica clínica científica.

EL PACIENTE INDIVIDUAL CONSULTA A CHARLES STEIN

El médico que aconsejó a los estudiantes de medicina en la mitad del siglo diecinueve que "el conocimiento más valioso que poseemos es el que hemos adquirido a través de una larga y continuada observación y experiencia. De ninguna manera se puede comunicar a los otros de forma escrita u oral" [4], estaría sorprendido al conocer que ahora se recomienda una búsqueda de MEDLINE como paso esencial para unos óptimos cuidados sanitarios [6]. Se sugiere que lo que se puede extraer de un ensayo aleatorio controlado, o mejor aún de un meta-análisis de ensayos aleatorios controlados, proporciona rápidamente la evidencia que se requiere.

Si el tema bajo consideración fuera el manejo de un paciente que ha sufrido un infarto de miocardio, se localizaría en este caso un meta-análisis sobre el uso de β -bloqueantes en prevención secundaria (figura 1). El meta-análisis estima que en conjunto existe una reducción del 22% en la mortalidad cuando se utilizan β -bloqueantes después de un ataque cardiaco. Una prueba estadística no rechaza la hipótesis nula que sostiene que los ensayos son homogéneos ($p = 0,16$) y un completo análisis de sensibilidad sugiere que los hallazgos del meta-análisis son robustos y se mantienen cuando se

cambian los métodos estadísticos, se ponderan los estudios de acuerdo con la calidad metodológica, o cuando se excluyen ensayos cuyos resultados pueden estar influidos por una finalización temprana [7].

La aplicación de los resultados del meta-análisis a un paciente en particular requiere aceptar la asunción de que los pacientes incluidos en los ensayos clínicos son un grupo de referencia apropiado para ese paciente en cuestión. ¿Es aceptable asumir que el efecto de un determinado tratamiento es idéntico a lo largo de los diferentes grupos de pacientes? ¿Responden de la misma forma el joven que el anciano, los hombres que las mujeres, los que padecen una enfermedad leve que aquellos que tienen una enfermedad severa? La alternativa a aplicar los resultados globales, es decir, la suma de la evidencia de los ensayos disponibles, es localizar subgrupos con características bastante similares a las del paciente en consideración. El arte de la medicina puede ser, en realidad, la aplicación de la habilidad que los expertos clínicos tienen, o piensan que tienen, para reconocer las sutiles diferencias y trazos que permiten la adecuación exacta del tratamiento y la investigación al paciente.

Sin embargo, parear las características del paciente con los subgrupos puede estar seriamente errado. Consideremos a un médico que trabaje en Alemania y que consulte el meta-análisis de β -bloqueantes después de un infarto de miocardio. Se puede tener en cuenta el único ensayo exclusivo que reclutó a una proporción substancial de pacientes alemanes, el Estudio Europeo sobre Infarto [8] (EIS en la figura 1) y desechar la reducción global de la mortalidad debida al tratamiento con β -bloqueantes. Este estudio no demostró beneficio, e insinuó un aumento de la mortalidad, para el grupo aleatorio que utilizó β -bloqueantes. ¿Cómo se deben manejar los pacientes alemanes? Es bastante inverosímil que estos pacientes no se beneficien realmente de un tratamiento que ayuda a los pacientes en otros países y la estimación más válida del resultado para los pacientes alemanes puede venir, esencialmente, de desestimar el único ensayo que reclutó en números substanciales a pacientes alemanes.

FIGURA 1. Mortalidad total en los 17 ensayos control-placebo de β -bloqueantes en prevención secundaria después de un infarto de miocardio. El cuadrado negro y la línea horizontal corresponde a la razón de odds y a los intervalos de confianza del 95% de los ensayos. El diamante representa la razón de odds combinada y su intervalo de confianza del 95%, que indica una reducción del 22% en los odds de la mortalidad (Fuente: Egger et al. [7].)

Esta conclusión puede parecer paradójica: de hecho la expresión estadística de este razonamiento se conoce como paradoja de Stein [9]. La aplicación de resultados procedentes de un meta-análisis, o el hallazgo global de un ensayo clínico más que lo que se observa en un análisis de subgrupos, se basa en la asunción de que las inconsistencias entre ensayos o entre subgrupos dentro de los ensayos se deben al azar. Charles Stein demostró que se puede hacer una estimación más fiable de una medida teniendo en cuenta los hallazgos procedentes de situaciones similares en vez de confiar completamente en un suceso en particular. La aplicación del método de Stein implica "contraer" los puntos de datos individuales hacia la gran media. El grado de ajuste de un único punto hacia la gran media dependerá de la precisión de su valor, esencialmente de la cantidad de datos en los que se basa ese valor. En el caso del meta-análisis de los β -bloqueantes, cuanto mayor sea el tamaño de la muestra de un estudio particular, mayor será la precisión estimada de su efecto y menor sería la contracción hacia la gran media. Un valor extremo que se basa en una cantidad substancial de datos se ajustará menos hacia la gran media que un valor extremo basado en pequeñas cantidades de datos. Dado que el EIS sólo es un contribuyente menor de la evidencia global respecto a los efectos de los β -bloqueantes después de un infarto de miocardio, su efecto estimado se ajustará de forma considerable hacia el resultado global.

LOS RIESGOS DE LOS ANÁLISIS DE SUBGRUPOS

Un factor que contribuye a la aplicación inadecuada de los hallazgos procedentes de subgrupos dentro de ensayos clínicos es que los informes a menudo se concentran más en los análisis de subgrupos que en el resultado global. En el caso de los β -bloqueantes después de un infarto de miocardio, los autores de los ensayos individuales (generalmente poco poderosos) a menudo procuran identificar algunos subgrupos sobre los que el tratamiento parece ser particularmente beneficioso o inútil. Sin embargo, al comprobar los hallazgos de estos subgrupos en ensayos posteriores y en un proyecto conjunto de diferentes ensayos se encontró que eran grandemente espúreos [10,11].

Un ejemplo actual lo ha proporcionado la presentación particular de los datos del Ensayo Beta-Bloqueantes en Ataque Cardíaco (BHAT), que comprendía 3.837 pacientes hospitalizados por infarto de miocardio en 31 centros médicos de Estados Unidos y Canadá [12]. Los pacientes recibieron de forma aleatoria propanolol o un placebo y estuvieron en seguimiento durante un promedio de 25 meses. En conjunto, las tasas de mortalidad estimadas fueron 7,2% para el propanolol y 9,8% para el placebo (RR 0,73; IC del 95%, 0,59-0,9). Horwitz et al [13] se sorprendieron al ver que los resultados en un tercio de los centros de estudio eran discordantes con respecto a los resultados principales, en donde salía más favorecido el placebo que el

TABLA 1. Tasas de mortalidad por tipo de tratamiento en los 31 centros del Ensayo Beta-Bloqueantes en Ataque Cardíaco

Centro	Número de pacientes	Tasa de mortalidad (%)		Diferencia proporcional (%)
		Propranolol	Placebo	
	97	0	6,3	-100
1	115	1,8	12,1	- 80
2	113	1,8	8,9	- 80
3	84	2,4	9,5	- 75
4	130	4,7	15,2	- 69
5	141	4,3	11,3	- 62
6	131	3,0	7,7	- 61
7	110	5,5	12,7	- 57
8	111	5,5	12,5	- 56
9	88	4,5	9,3	- 51
10	155	5,2	10,3	- 49
11	117	3,4	6,8	- 49
12	140	5,7	10,0	- 43
13	196	7,2	12,1	- 40
14	96	6,3	10,4	- 40
15	105	13,5	21,2	- 36
16	129	7,6	11,1	- 32
17	84	7,0	9,8	- 29
18	94	10,4	13,0	- 20
19	125	6,5	7,9	- 19
20	119	8,3	10,2	- 18
21	251	8,0	7,9	0,8
22	115	17,5	17,2	2
23	65	12,5	12,1	3
24	88	7,0	6,7	5
25	140	11,4	10,0	14
26	193	7,4	6,1	20
27	150	10,7	6,7	60

1/6/2018	Asociación Española de Evaluación de Tecnologías Sanitarias			
28	110	12,7	7,3	75
29	127	12,5	4,8	162
30	118	8,5	1,7	400
31				

Fuente: Horwitz et al. [13].

propranolol (Tabla 1). Compararon 20 centros en los cuales el propranolol era superior al placebo (RR combinado 0,5; IC del 95%, 0,38-0,67) con 11 centros en que ocurría lo contrario (RR combinado 1,33; IC del 95%, 0,95-1,88) y concluyeron que, basándose en la falta de solapamiento de los intervalos de confianza, "los distintos efectos del tratamiento con propranolol en los centros dominantes y divergentes no se podía atribuir exclusivamente al azar" [13].

Horwitz et al intentaron entonces explicar las diferencias de los efectos del tratamiento en términos de características de los pacientes, las cuales diferían entre los centros que informaron acerca de un efecto beneficioso del tratamiento y aquellos que informaban sobre un efecto perjudicial. Encontraron varias diferencias: en los centros que informaban acerca del efecto beneficioso de los β -bloqueantes estaban en mayor grado representados los pacientes que tenían una historia de infarto de miocardio previo, que tenían insuficiencia cardiaca congestiva y que no eran de raza blanca, que en los centros en los que los β -bloqueantes estaban asociados con un resultado adverso. Horwitz y colaboradores sugirieron que las recomendaciones sobre tratamiento deberían modificarse de acuerdo con sus hallazgos.

Horwitz y colaboradores consideraron una "observación extraordinaria que un tercio de los centros BHAT tuvieran resultados favorables para el placebo" [14]. Sin embargo, es bastante probable que estos hallazgos se deban enteramente al azar y cualquier conclusión obtenida a partir de ellos, espúrea. Esto se puede demostrar mediante un sencillo ejercicio de simulación. Si se asume un RR constante de 0,73 a lo largo de los 31 centros BHAT (multiplicando el número de muertes observadas en el grupo placebo por 0,73 para obtener el número esperado de muertes en el grupo de propranolol) y se añade un error aleatorio (mediante muestreo de la correspondiente distribución de Poisson con una media igual al número de muertes), aparecen resultados similares a los observados en BHAT (Tabla 2). Había 20 centros en donde el propranolol era superior al placebo (RR combinado 0,58; IC del 95%, 0,44-0,78) y 10 centros en los cuales el placebo era superior al propranolol (RR combinado 1,76; IC del 95%, 1,21-2,54). Las observaciones que Horwitz y colaboradores querían utilizar para modificar las recomendaciones de tratamiento es probable que sólo reflejen el lugar que ocupa el azar.

Otro ejercicio de simulación ha mostrado que, si el efecto del tratamiento es estadísticamente significativo ($p = 0,05$) y los pacientes se distribuyen de forma aleatoria entre dos grupos, en una de cada tres ocasiones el efecto del tratamiento será grande y altamente significativo, estadísticamente hablando, en un grupo y no significativo e irrelevante en el otro [15]. Por tanto, el análisis de subgrupos puede llenar la literatura de hallazgos espúreos, aunque aparentemente de relevancia clínica, que hará desviar el rumbo de la práctica clínica más que ayudarla.

EL PROBLEMA DE GENERALIZAR LOS HALLAZGOS DE LOS ENSAYOS

El otro lado de la moneda respecto a la demostración de que los análisis de subgrupos producen probablemente hallazgos espúreos es que es fácil imaginar situaciones en las cuales el mismo tratamiento puede beneficiar a un grupo de pacientes y no tener efecto, o incluso producir daño, en otro grupo. Durante un amplio rango de conferencias en 1.921, Sir James Mackenzie demostró

de forma sofisticada la comprensión de este aspecto [16]. Mackenzie, cuyas contribuciones a la cardiología, la epidemiología clínica y a la práctica de la medicina general son ampliamente reconocidas [17, 18], puso en conocimiento cómo de limitada es la habilidad de la práctica médica en relación con el juicio pronóstico cuando se realizan exámenes médicos para pólizas de seguros de vida [16]. Mackenzie aplicó su método, que combinaba observaciones clínicas extensas y detalladas, principalmente procedentes de la práctica de medicina general, con el conocimiento fisiológico (como el contrario al puramente anatómico) de la función del corazón,

TABLA 2. Tasas supuestas de mortalidad por tipo de tratamiento en los 31 centros del Ensayo Beta-Bloqueantes en Ataque Cardíaco^a

Diferencia
proporcional (%)

Tasa de mortalidad (%)

Centro	Número de pacientes	Propranolol	Placebo	
	88	0	6,8	-100
10	118	0	1,7	-100
31	110	3,6	12,7	- 71
8	127	1,6	4,8	- 67
30	125	4,8	12,9	- 63
20	94	4,3	10,6	- 60
19	97	4,1	8,3	- 51
1	84	4,8	9,5	- 50
4	141	7,0	12,9	- 45
6	251	4,0	7,2	- 45
22	155	9,0	15,6	- 42
11	88	6,8	11,4	- 40
25	110	3,6	5,5	- 33
29	196	7,1	10,2	- 30
14	115	8,6	12,3	- 30
23	193	3,1	4,2	- 26
27	84	7,1	9,5	- 25
18	111	7,1	9,1	- 21
9	115	17,2	19,3	- 11
2	117	5,1	5,2	- 2
12	96	1,3	1,3	0
15	113	15,4	14,1	+ 9
3	105	6,1	4,6	+ 31
16	140	5,7	4,3	+ 33
26	65	8,3	5,1	+ 64
24	119	27,3	15,6	+ 75
21	130	13,8	7,7	+ 80
5	140	8,6	4,3	+100
13	131	24,5	11,5	+113
7	129	10,5	3,6	+195
17	150	10,4	2,1	+400
28				

a Resultados de un ejercicio de simulación en el que se asume una reducción constante del riesgo relativo del 27% (RR = 0,73) en todos los centros, después de añadir un error aleatorio de Poisson.

en la controversia acerca del uso adecuado de digital. El observó que [16]:

Desde hace más de cien años se reconoce que el digital tiene una acción benéfica en determinadas formas de enfermedad cardiaca, pero no existe un conocimiento preciso acerca de cual tipo de enfermedad cardiaca puede beneficiarse de la droga, por tanto, a cada persona que se le supone una afección del corazón se le da digital. Cuidadosas investigaciones han revelado que la droga se debe utilizar exclusivamente en un pequeño porcentaje de casos ...

El pequeño porcentaje de casos que el tenía en mente se referían básicamente a los pacientes con fibrilación del atrium y con una rápida respuesta ventricular [17], aunque dijo que algunas veces podría haber una mejoría de los síntomas en fallo cardiaco con ritmo sinusal [19]. La visión restrictiva de Mackenzie acerca del terreno apropiado del uso de la terapia digital se extendió rápidamente entre las ideas de sus contemporáneos, principalmente en Europa y Estados Unidos, que consideraban que la digital se debería utilizar en todas las formas de fallo cardiaco, incluidos los pacientes con ritmo sinusal [18]. Un artículo en el primer volumen del recientemente bautizado *British Heart Journal* en 1939 marcó la adopción masiva de la perspectiva de Mackenzie en su propio país [20]. Esta visión fue repetida de forma frecuente a lo largo del siguiente medio siglo [21].

Un ensayo aleatorio controlado a gran escala sobre la terapia digital en pacientes con fallo cardiaco y en ritmo sinusal, publicado 76 años después de la conferencia de Mackenzie [22], no encontró beneficios sobre la mortalidad global y una pequeña reducción en un resultado que combinaba hospitalización o mortalidad en casos de empeoramiento del fallo cardiaco. Los beneficios fueron mayores, en términos absolutos y relativos, para los pacientes que tenían un riesgo mayor cuando comenzaban el tratamiento. Esto confirma y asienta el argumento general de Mackenzie acerca del objetivo del tratamiento, que se basaba tanto en el nivel de riesgo y en la fisiopatología de la enfermedad.

Mackenzie consideró otras diversas situaciones donde los resultados de los tratamientos podrían variar entre los grupos de pacientes. Para un paciente sospechoso de apendicitis, señaló que una intervención podría o bien salvarle la vida o ser un riesgo innecesario, dependiendo si tenía en ese momento un apéndice inflamado o no. La cirugía en el caso de la úlcera péptica presentaba el mismo caso: "He visto pacientes enormemente beneficiados por la operación, algunos han empeorado por la misma y he visto incluso morir pacientes como consecuencia de la intervención". En ese momento el tratamiento de la úlcera péptica era un asunto de intenso debate, en el que unos abogaban por la cirugía y otros por el tratamiento médico. En una serie de pacientes tratados quirúrgicamente sólo se curaron el 27% [23]. Esto se vió como "el resultado evidente de la interferencia de la cirugía en aquellos casos que eran tributarios de tratamiento médico" [23]. En realidad, tal como consideró un eminente cirujano, "pueden encontrarse síntomas similares a una úlcera gástrica en otras enfermedades y las intervenciones sobre el estómago se realizan frecuentemente sin encontrar ninguna justificación local" [24]. Otros, sin embargo, han informado de altas tasas de éxito (80%) de la cirugía y fallos del tratamiento en la mayoría de los pacientes tratados médicamente [25,26]. En esa época la mortalidad de la operación estaba alrededor del 5%, mientras que la mortalidad de las úlceras gástricas no tratadas se estimaba en un 20% [23,25].

Esta situación hace clara la tensión entre la perspectiva individual y la de grupo. Imaginemos un ensayo aleatorio, llevado a cabo en los tiempos en que Mackenzie escribía eso, comparando el tratamiento médico con el quirúrgico. En ese tiempo, cuando no había gastroscopias, el ensayo hipotético habría incluido una mezcla de pacientes, con unos que probablemente se beneficiarían de la intervención (casos quirúrgicos) y muchos casos que se tratarían de forma conservadora. Tal como aparece en la tabla 3, los resultados generales (con un beneficio significativo de la cirugía) habrían sido desafortunados cuando se aplicaban a los pacientes individuales, para quienes la cirugía produciría o una reducción substancial mayor o un aumento del riesgo de muerte.

TABLA 3. Resultados de la mortalidad de un hipotético ensayo realizado en la década de los 20 donde se comparan tratamientos médicos y quirúrgicos entre 800 pacientes con úlcera péptica de estómago

Resultados globales*	Casos quirúrgicos				Casos médicos	
	Muertos	Vivos	Muertos	Vivos	Muertos	Vivos
Tratamiento quirúrgico						
Tratamiento médico	20	380	10	190	10	190
	42	358	40	160	2	198
Riesgo relativo	0,48		0,25		5,0	
Intervalos de confianza del 95%	0,28-0,80		0,13-0,49		1,11-22,5	

* Los resultados generales indican un efecto beneficioso de la cirugía, reduciendo el riesgo de muerte en un 52%. Este es un promedio erróneo que infraestima el beneficio que se obtiene en los casos "quirúrgicos" y falla a la hora de reflejar el daño que causa a los casos "médicos".

Existen muchos ejemplos de terapias que tienen diferentes efectos en pacientes con diferentes niveles de riesgo [27]. Si tomamos un ejemplo de cirugía más reciente y real, los datos de pacientes individuales de un meta-análisis de ensayos de by-pass de arterias coronarias [28], se observa

que mientras que los pacientes de alto riesgo consiguen aumentar de forma significativa el tiempo de supervivencia después de la cirugía, los pacientes de bajo riesgo tienen en la actualidad periodos de supervivencia más cortos si se someten a cirugía (figura 2). De forma similar, el valor de la escleroterapia para reducir el riesgo de sangrado en pacientes cirróticos con varices esofágicas se debería restringir a los pacientes de alto riesgo [29].

Los hallazgos aquí presentados resaltan los problemas de la generalización a partir de los resultados de los ensayos clínicos. El problema de la generalización ha estado tradicionalmente relacionado, sin embargo, con la extensión de los hallazgos procedentes de un grupo incluido en un ensayo particular a otros grupos no incluidos en él: por ejemplo, generalizar los resultados de ensayos llevados a cabo en hombres a mujeres, o los llevados a cabo en personas en edades medias a ancianos. Sin embargo, tal como se ha visto previamente, también pueden existir problemas a la hora de generalizar los resultados globales de un ensayo a grupos de pacientes incluidos en el propio estudio. Tomemos el hipotético ejemplo del ensayo sobre úlcera péptica (ver tabla 3) en el que la cirugía parece reducir la mortalidad general, incluso aunque para un grupo importante de participantes tiene efectos dañinos. La aplicación de los criterios de inclusión del ensayo a vuestros pacientes se traduciría en la identificación de algunas personas como candidatas al tratamiento, aquellas que concuerdan con los criterios del ensayo por los cuales el tratamiento reduce la mortalidad, y que realmente los situarían en un lugar donde tendrían un riesgo elevado de sufrir un daño al recibir el tratamiento.

DE LO GENERAL A LO PARTICULAR

La naturaleza potencialmente engañosa de los resultados agregados de los ensayos está ilustrada en la discusión de Ioannidis y Lau [30] acerca de la contribución de los pacientes de alto riesgo a los hallazgos de los ensayos clínicos. El tratamiento que dieron al tema del efecto del magnesio intravenoso en el infarto agudo de miocardio también suministra una demostración de cómo no existe una solución sencilla para determinar si se debería aceptar el resultado global o se deberían buscar subgrupos para ello. A pesar de que los diversos análisis de subgrupos hechos en el gran ensayo ISIS-4 no encontraron ningún beneficio con el magnesio en el subgrupo de alto riesgo [31,32], permanece la creencia de que los resultados generales obscurecen el valioso beneficio en los pacientes de alto riesgo [30,33].

FIGURA 2. Aumento del tiempo de supervivencia de acuerdo con la puntuación de riesgo de la Administración de Veteranos procedente de siete ensayos que comparan la cirugía de bypass coronario mediante injerto con el tratamiento médico en pacientes con enfermedad coronaria estable (Fuente: Yusuf et al. [28]).

Las discusiones acerca de cuando es apropiado generalizar desde lo abstracto a pacientes particulares o a subgrupos de pacientes revelan que los que proponen diferentes puntos de vista no pueden apreciar las posiciones de los otros [14,34]. El nivel de debate no ha avanzado grandemente desde que Civiale y Double cruzaron sus espadas, con los clínicos defendiendo su desdén de los principios básicos de la estadística en nombre de la "biología clínica" [14] y los estadísticos presentando argumentos a nivel de abstracción [34] que produce un rechazo comprensible por parte de los clínicos.

Lo que se requiere es un grado de humildad en un tema para el que no hay solución clínica o estadística. Con el beneficio de conocer lo que ya ha ocurrido, descubriremos probablemente que en muchas de las situaciones clínicas actuales nos encontramos en la misma situación que Mackenzie al principio del siglo para determinar cuales de las características del paciente influirían en las decisiones sobre el tratamiento. El desarrollo de los ensayos clínicos aleatorios desde los tiempos de Mackenzie ha proporcionado una base mucho más sólida para tomar decisiones acerca de pacientes abstractos y, si se incluyen muestras representativas de pacientes en los ensayos, decidir si el efecto global de un tratamiento sobre la salud de la población es beneficioso o dañino. Sin embargo, los ensayos aleatorios no han contestado todavía la pregunta de cuáles individuos se benefician actualmente de las intervenciones médicas. Esto, seguramente, es el tema clave de la investigación clínica para el próximo milenio.

El Departamento de Medicina Social de la Universidad de Bristol forma parte del Consejo de Investigación Médica para la Colaboración en la Investigación de Servicios Sanitarios

Society for Social Medicine

META-ANÁLISIS/HSR

Ensayos con doble placebo, creencias firmes y meta-análisis: lecciones procedentes de la homeopatía y la cardiología

M. EGGER, G. DAVEY SMITH (*Departamento de Medicina Social, Universidad de Bristol*)

Objetivos- Revisar posibles fuentes de sesgos en meta-análisis; ofrecer métodos gráficos y estadísticos para detectar tales sesgos; y aplicar estos métodos a meta-análisis de ensayos aleatorios recientes.

Antecedentes- Diversos mecanismos pueden conducir a la inclusión de una muestra sesgada de estudios en una revisión sistemática o meta-análisis. Entre ellos se incluyen los sesgos de publicación, sesgos del idioma, sesgos de citas y sesgos múltiples de publicación. Una heterogeneidad real en el tamaño de las estimaciones del efecto procedentes de diferentes estudios, por ejemplo cuando los efectos difieren según el riesgo subyacente, o cuando la intensidad de la intervención varía a lo largo de los estudios, puede producir unas estimaciones combinadas del efecto sin sentido. Las irregularidades en los datos, que pueden ir desde un pobre diseño metodológico a un análisis inadecuado y un fraude evidente, son otra fuente de sesgos. Se discutirá el fundamento teórico de estos sesgos y se mostrará la evidencia empírica.

Métodos- Se dispone de métodos gráficos y estadísticos para detectar los sesgos y explorar las fuentes de heterogeneidad en meta-análisis. El gráfico de chimenea es un sencillo gráfico de dispersión de las estimaciones del efecto en relación con el tamaño de la muestra. Este gráfico es útil para detectar la heterogeneidad cuya causa puede estar en que no hay pequeños estudios negativos (sesgos de publicación) o en una baja calidad en ensayos pequeños. Las pruebas estadísticas examinan si existe asociación entre las estimaciones del efecto y sus precisiones. Entre los métodos se incluyen una prueba de correlación del rango, un enfoque de regresión de efectos fijos y una meta-regresión de efectos al azar.

Marco- Meta-análisis de 89 ensayos sobre homeopatía, de 15 ensayos sobre magnesio intravenoso en infarto de miocardio y de 22 ensayos de estreptoquinasa intravenosa en infarto de miocardio.

Resultados- Los exámenes visuales indican una asimetría en el gráfico de chimenea y por tanto un sesgo potencial en los ensayos sobre homeopatía y magnesio pero no en los ensayos sobre estreptoquinasa. Esto se confirmó con los análisis estadísticos habituales. En el caso de la homeopatía, también se asoció con las estimaciones del efecto el idioma de publicación y la calidad del estudio. Una búsqueda manual de publicaciones especializadas en homeopatía y magnesio indicó que existe una comunidad de investigadores que mantienen unas creencias firmes en la eficacia de estos tratamientos. Se mostrarán ejemplos ilustrativos.

Conclusiones- La utilización de métodos gráficos y técnicas de meta-regresión para escrutar de forma puntillosa la presencia de sesgos y otras fuentes de heterogeneidad en las revisiones sistemáticas y en el meta-análisis, debería convertirse en un procedimiento de rutina en el meta-análisis. Esto es particularmente importante cuando existen creencias fuertes sobre el tema de estudio a priori. Estas creencias parecen ser bastante comunes tanto en la medicina dominante como en la alternativa.

BMJ, 1.994; 308: 72-74.

¿Quién se beneficia de las intervenciones médicas?

Tratar pacientes de bajo riesgo puede ser una estrategia de alto riesgo

A menudo, los resultados de los ensayos clínicos se expresan en términos relativos, por ejemplo, un tratamiento particular reduce en un 40% el riesgo de un resultado adverso. Incluso conocer que el tratamiento reduce el riesgo de tal resultado desde un 5% a un 3% (una reducción absoluta de un 2%), puede ser clínicamente más útil¹.

Los efectos del tratamiento antiplaquetario son una referencia en este caso. El meta-análisis de ensayos aleatorios controlados de tales tratamientos realizado por la Colaboración de Ensayistas Antiplaquetarios muestra que el riesgo de morir por causa vascular varía substancialmente según el grupo de pacientes (p. 81)². El rango va desde un 10% en un mes para pacientes que entraron en el ensayo en el momento de un infarto agudo de miocardio, a un 9% a los dos años de supervivencia después del infarto y baja hasta sólo un 2% en cinco años en el caso de ensayos de prevención primaria.

No es por tanto sorprendente que, con tal variación de riesgos de futuros sucesos vasculares, los efectos absolutos de los agentes antiplaquetarios difieran substancialmente según los grupos de pacientes. Las reducciones en términos de riesgos relativos pueden ser similares, las drogas antiplaquetarias reducen el riesgo futuro de infarto no letal de miocardio en un 30% tanto en ensayos de prevención primaria como de secundaria. Pero cuando los resultados se presentan en términos de número de pacientes que necesitan ser tratados para evitar un infarto no letal de miocardio, los números parecen muy diferentes.

Para prevenir un infarto no letal de miocardio, se necesitan tratar 50 pacientes durante dos años en prevención secundaria, y 200 pacientes durante cinco años en prevención primaria. En otras palabras, se necesitan 100 pacientes-año de tratamiento en prevención secundaria o 1.000 pacientes-año de tratamiento en prevención primaria para obtener el mismo resultado beneficioso de un infarto no letal de miocardio menos.

El hecho de que el riesgo en que se encuentran los pacientes cuando comienzan el tratamiento influye en el beneficio absoluto del tratamiento, se ha discutido ampliamente en relación con el tratamiento de la hipertensión³⁻⁵. De acuerdo con los ensayos aleatorios controlados, el tratamiento reduce de forma consistente el riesgo de accidente cerebrovascular en un 40% (expresado en términos relativos)⁶. Sin embargo, la variación de hasta 10 veces en las tasas de accidentes cerebrovasculares en los pacientes de estos ensayos, produce un rango del número de pacientes-año de tratamiento necesitados para prevenir un accidente cerebrovascular que varía entre alrededor de 100 y 1.000. Para prevenir una muerte por proceso vascular, las cifras equivalentes son aproximadamente el doble. Las características de los pacientes en la línea de base, edad, sexo, presión sanguínea, otros factores de riesgo para enfermedad cardiovascular y el daño final en el órgano, pueden discriminar grupos con amplias diferencias de riesgo de enfermedad coronaria o cerebrovascular. El balance entre beneficios y efectos secundarios del tratamiento será más favorable en pacientes con mayor riesgo de enfermedad vascular.

Consideremos, por ejemplo, la hipertensión sistólica en el programa del anciano (SHEP), en el que el 28% del grupo de tratamiento informaban sobre lo que ellos consideraban un "problema intolerable" durante el tratamiento, una proporción considerablemente mayor que el 21% que informaba sobre ese mismo suceso en el grupo placebo⁷. El balance entre reducir el riesgo de un accidente cerebrovascular o de una muerte por causa vascular y el incremento, que afecta a 7 de cada 100 pacientes, en el riesgo de experimentar un "problema intolerable" sería claramente valorado de forma diferente según los distintos niveles de riesgo cardiovascular de los pacientes.

El balance entre los beneficios y los efectos secundarios indeseados es particularmente firme cuando los efectos secundarios son serios. El balance será más favorable para aquellos pacientes con un riesgo mayor de sufrir el evento y que el tratamiento trata de prevenir. En el caso de los agentes antiplaquetarios, existe cierta evidencia de que pudieran incrementar ligeramente el riesgo de accidente cerebrovascular hemorrágico. Para el caso de participantes de bajo riesgo en ensayos de prevención primaria, esto favorece el hecho de que las reducciones en muertes por proceso vascular debidas al tratamiento son pequeñas y no significativas, lo que se traduce en que la mejor estimación es la de una muerte evitada por cada 1.450 personas-año de tratamiento. Esto es desfavorable cuando se compara con una muerte vascular evitada por cada 75 personas-año de tratamiento en los supervivientes de un infarto de miocardio y de una muerte evitada por cada 3 personas-año de tratamiento en pacientes tratados durante un infarto agudo de miocardio.

Consideraciones similares se pueden aplicar al consumo apropiado de drogas que disminuyen el colesterol⁸. Para pacientes que tienen alto riesgo de morir por enfermedad coronaria, la reducción de las concentraciones de colesterol sanguíneo mediante fármacos reduce la mortalidad. Para pacientes con riesgo bajo, por ejemplo, aquellos con concentraciones moderadamente altas de colesterol sanguíneo pero sin síntomas de enfermedad cardiovascular, las drogas que reducen el colesterol pueden suponer un aumento del riesgo de muerte. Para estos pacientes, un pequeño aumento absoluto del riesgo de muerte atribuible al tratamiento, de aproximadamente 1-2 muertes por 1.000 pacientes-año de tratamiento, es suficiente para rechazar cualquier tratamiento relacionado con la disminución del riesgo de enfermedad coronaria.

Pequeños riesgos pueden tener mayor peso que pequeños beneficios

Comentando este fenómeno, Geoffrey Rose subrayó que "en la prevención de masas, cada individuo tiene generalmente sólo una pequeña expectativa de beneficio y este pequeño beneficio puede ser fácilmente sobrepasado por un pequeño riesgo"⁹. Esto debería servir como precaución contra el consumo indiscriminado de drogas para prevenir enfermedades hasta que se hayan establecido sus beneficios globales mediante ensayos controlados¹⁰.

Un último ejemplo en el campo de la medicina cardiovascular concierne a los inhibidores de la enzima convertidora de la angiotensina utilizados para tratar y prevenir el fallo cardíaco. Mientras que la utilización de estas drogas en el infarto agudo de miocardio en apariencia produce exclusivamente un pequeño beneficio¹¹, su utilización a largo plazo en pacientes con, o en riesgo de desarrollar, fallo cardíaco ha tenido resultados más impresionantes. Los resultados de cinco grandes ensayos¹²⁻¹⁶, que aparecen en la figura, muestran que tratar pacientes de alto riesgo produce un beneficio considerablemente mayor que tratar pacientes de bajo riesgo. Dado que el rango en el número de pacientes que se necesitan tratar en un año para prevenir una muerte varía entre 2 y 330, dependiendo de su riesgo de muerte por enfermedad cardíaca previa, las evaluaciones de riesgo-beneficio y coste-efectividad variarán ampliamente entre diferentes situaciones clínicas.

Figura: Efecto de los inhibidores de la enzima convertidora de angiotensina sobre la mortalidad en pacientes con disfunción ventricular: resultados de cinco ensayos controlados

Fuera del campo de la medicina cardiovascular, es evidente la misma relación, por ejemplo, cuando se evalúa el beneficio de tratar la infección por VIH con zidovudina. El primer gran ensayo en pacientes con SIDA, o complejos relacionados con SIDA, con alto riesgo de morir muestra efectos impresionantes: 42 muertes menos por 100 personas-año de tratamiento¹⁷. Incluso en pacientes con una infección menos avanzada de VIH que participaron en el ensayo Concorde, el uso temprano de zidovudina se asoció con un efecto adverso no significativo sobre la mortalidad¹⁸. Cuando se examinó la relación entre el beneficio del tratamiento y la mortalidad entre los controles de seis ensayos publicados¹⁷⁻²², el aumento de riesgo de muerte en los controles se asoció con el incremento en el retraso del número de muertes debido al tratamiento con zidovudina. Para las personas que tienen una infección asintomática por VIH, los efectos secundarios de la zidovudina junto con los inciertos efectos en la mortalidad hacen que la opción del tratamiento no sea muy atractiva. Una demostración formal reciente acerca de que el tratamiento temprano con zidovudina en el curso de la enfermedad no es coste-efectivo apoya esta conclusión²³.

El mayor beneficio (y la disminución de la probabilidad de daño) experimentado por pacientes en alto riesgo tiene diversas implicaciones en la práctica clínica y en las políticas de salud pública. En primer lugar, la forma de presentación de los datos del ensayo puede influir las percepciones del médico acerca de las recomendaciones de tratamiento. Convertir las mediciones relacionadas con el beneficio en mediciones tales como el número de pacientes que necesitan ser tratados para prevenir un evento puede modificar la respuesta de los facultativos a los resultados de los ensayos, presumiblemente en la dirección adecuada²⁴. Se debería potenciar que tanto los investigadores como las empresas farmacéuticas proporcionen medidas absolutas y relativas del beneficio producido por el tratamiento.

En segundo lugar, las guías de prácticas clínicas deberían reflejar la asociación entre el grado de beneficio y el nivel de riesgo. Esto no ha sido así en el caso del primer conjunto de guías de práctica clínica del programa nacional de educación sobre colesterol en los Estados Unidos²⁵. En este caso se recomendaba el tratamiento intensivo para determinados grupos de pacientes con un considerable riesgo menor de evento coronario que para otros grupos de pacientes que se consideró que no requerían una atención especial²⁶. De forma prometedora, tanto la última revisión de estas guías²⁷ como el nuevo conjunto de guías de Nueva Zelanda para tratar el aumento de la presión arterial²⁸ han tenido un enfoque basado en el riesgo a la hora de hacer decisiones acerca del tratamiento. En tercer lugar, el meta-análisis debería incorporar estimaciones del riesgo de los ensayos incluidos para examinar si las diferencias sistemáticas observadas en el beneficio están relacionadas con el grado de riesgo de los participantes de los ensayos⁸. En último lugar, dado que aquellos que más se benefician del tratamiento son los que más tienen que ganar, se debería reconocer la irracionalidad de las recomendaciones actuales que aconsejan que no se trate a los grupos de población con un aumento del riesgo de enfermedad, tal como es el caso de los fumadores.

GEORGE DAVEY SMITH

Senior lecturer en epidemiología y salud pública
Departamento de Salud Pública,
Universidad de Glasgow,
Glasgow G12 8RZ

MATTHIAS EGGER

Senior research fellow
Departamento de Medicina Social y Preventiva
Universidad de Berna,
CH-3012 Berna,
Suiza.