Una escala pronóstica, o un chivato como "llevar tabaco", quiere predecir el futuro, anticiparlo. Pero para cambiarlo, la relación debe ser causal

Cobo / Ventura

Falacias en investigación epidemiológica

Esteve Fernández

Andábamos ya enfrascados en la preparación de esta Jornada cuando un buen compañero me envió, como regalo de cumpleaños, un artículo todavía en prensa de Keneth J. Rothman, padre –o uno de los padres– de la "epidemiología moderna". Con un título más que sugerente *(Six persistent research misconceptions)* (1), Rothman sistematiza seis conceptos erróneos que, a su parecer, persisten en el imaginario y en la práctica de los investigadores a pesar de haberse probado su falsedad. Unas semanas más tarde pude asistir a un seminario impartido por el propio Rothman, en el cual amplió estos errores a 10 y los etiquetó como falacias.

No son nuevas estas falacias, ni seguramente están todas, y tanto Rothman como Porta (2) y otros epidemiólogos habían y habíamos predicado ya en este sentido, pero siempre con un común denominador: nos habíamos dirigido a investigadores, a estudiantes de posgrado y aprendices de investigadores, tanto en el ámbito clínico como en el epidemiológico. A modo de recuerdo, la serie «Epidemiología para clínicos» que abrió las puertas del valor de la epidemiología a una incontable cohorte de médicos y otros profesionales clínicos, incluyendo jóvenes aprendices de epidemiólogo, con artículos sobre el razonamiento epidemiológico (3), los diseños epidemiológicos (4) o los errores más frecuentes en este tipo de investigación (5-7).

Me pareció que esas "falacias" no solo estaban arraigadas en el imaginario de los investigadores, sino que también podían estarlo en el de los periodistas de salud que se enfrentan casi a diario con diseños epidemiológicos y conceptos estadísticos a través de la lectura de notas de prensa y artículos científicos, y en el de los propios investigadores, sin haber realizado, por lo general, una formación específica en los principios epidemiológicos básicos.

Por todo ello, tras procesar la información y poner los debidos filtros, me pareció muy claro

Tabla 1. Falacias en investigación epidemiológica y clínico-epidemiológica.

1) El ensayo clínico controlado es el mejor diseño de estudio.
2) Los estudios prospectivos son mejores que los retrospectivos.
3) El tamaño importa.
4) La representatividad importa.
5) La significación estadística da validez al estudio.
6) Mucha evidencia significa buena evidencia.
7) Las revistas buenas garantizan estudios buenos.

que algunas de estas falacias, y otras de cosecha propia basadas en la experiencia al tratar con comunicadores científicos y al comunicar directamente al público, merecían una exposición en una jornada sobre epidemiología y periodismo (Tabla 1).

1) El ensayo clínico controlado es el mejor diseño de estudio

Esta afirmación, basada en el conocimiento que tenemos de los diferentes tipos de estudio usados en epidemiología, es esencialmente cierta y ha calado profundamente entre investigadores y periodistas. La presencia de un grupo control, a semejanza del control que usaríamos en cualquier experimento en un laboratorio, parece la panacea que sustenta la validez de los ensayos clínicos, que de esta forma superan las limitaciones de los estudios observacionales, etiquetados casi a priori de «más débiles». Gonzalo Casino nos ha recordado en este Cuaderno el reto que supone para los informadores «conocer y diferenciar los distintos tipos de diseños epidemiológicos y la confianza que merecen». Los informadores no son ajenos a este problema, y en un seminario anterior sobre *Bioestadística para periodistas y comunicadores* ya se abordó esta cuestión (8). En ese seminario, Pablo Alonso-Coello presentó el sistema GRADE (9), en el cual a los ensayos clínicos se les asigna de entrada una calidad alta, pero pueden ser penalizados, y a los estudios observacionales se les asigna de entrada una calidad baja, pero en algunas ocasiones la confianza en ellos puede aumentar.

Así, aunque el ensayo clínico tenga en principio las mejores cualidades para ofrecer resultados válidos, sin los errores sistemáticos más propios de los estudios observacionales, no debemos caer en la falacia de pensar automáticamente que «el ensayo clínico es el mejor diseño de estudio». Para no caer en esta falacia deberíamos añadir a la proposición anterior algo así como «siempre que esté bien diseñado y ejecutado». En efecto, los ensayos clínicos pueden hacerse bien o no tan bien, e incluso acumular tantos errores (en el diseño, en su ejecución, en el análisis de los datos) que sus resultados no sean válidos o sean menos válidos que los de un estudio observacional, por ejemplo un estudio de cohortes prospectivo, pulcramente diseñado y ejecutado.

El primer ensayo clínico sobre la eficacia de los cigarrillos electrónicos para dejar de fumar, publicado en la prestigiosa revista *The Lancet,* concluyó que «los cigarrillos electrónicos, con o sin nicotina, eran moderadamente eficaces en ayudar a los fumadores a dejar de fumar, con una abstinencia similar a la de los parches de nicotina» (10). Esta conclusión se basaba en las tasas de abstinencia de consumo de tabaco a los 6 meses: 7,3% de los usuarios de cigarrillos electrónicos con nicotina, 4,1% de los usuarios de cigarrillos electrónicos sin nicotina y 5,8% de los usuarios de parches de nicotina, sin diferencias estadísticamente significativas. ¿Debemos creer a pies juntillas estos resultados e interpretarlos dogmáticamente porque derivan de un ensayo clínico y además aleatorizado? Estos resultados, al margen del pequeño tamaño muestral (que muchas veces sí que importa), deben ser discutidos a la luz de otros aspectos del diseño y la ejecución del ensayo. Este se diseñó para contrastar la superioridad de los cigarrillos electrónicos sobre los parches de nicotina, no para demostrar que tienen un efecto similar (que es la interpretación de los autores), para lo que no está preparado. Además, sorprende la baja abstinencia en los

usuarios de parches, cuando en ensayos clínicos previos y en estudios de resultados en la práctica real se alcanzan abstinencias a los 6 meses del 15% al 20%. También sorprende cómo se administró el tratamiento en los grupos, que bien podría condicionar la adherencia al tratamiento: mientras que los participantes asignados a usar cigarrillos electrónicos los recibieron directamente en sus casas, junto con las recargas, los participantes asignados al parche de nicotina recibieron unos cupones para retirarlos en su farmacia.

Para acabar con esta falacia, también vale la pena recordar que algunas hipótesis no pueden ser evaluadas mediante un ensayo clínico, por razones logísticas o sencillamente éticas. Sabemos que el tabaco causa cáncer de pulmón gracias a estudios observacionales de casos y controles y de cohortes, y no ha sido necesario hacer, ni hubiese sido ético, un ensayo clínico en el que, tras una asignación aleatoria, tuviéramos un grupo de personas fumando y otro grupo sin fumar durante 20 años para analizar las respectivas tasas de incidencia de cáncer de pulmón.

2) Los estudios prospectivos son mejores que los retrospectivos

Esta falacia deriva del común error de haber "asimilado" durante muchos años un estudio retrospectivo con un estudio de casos y controles, que en el sistema GRADE tiene un menor grado de evidencia, junto con la concepción errónea de considerar de peor calidad la información recogida retrospectivamente que la información recogida de manera prospectiva.

La característica de prospectividad o retrospectividad de un estudio puede aplicarse tanto a estudios de cohortes (o longitudinales) como a estudios de casos y controles, puesto que se refiere a cómo se sitúa el investigador respecto a las observaciones a hacer, es decir, a la información que se va a recoger. Así, si el investigador anticipa qué información va a recoger, estamos ante un estudio prospectivo. Podemos diseñar un estudio prospectivo de cohortes, cuando el seguimiento se hace prospectivamente (hacia delante) tras la determinación de la información a recoger y cuando los sucesos a observar aún no han aparecido, pero también podemos hacer un seguimiento retrospectivo de la cohorte (hacia atrás) cuando la información ya existe y los sucesos ya se han producido. En los estudios de casos y controles, la información sobre las exposiciones es retrospectiva, pues aun tratándose de nuevos casos van a proporcionar información de sus exposiciones pasadas. Sin embargo, el reclutamiento de casos y controles puede ser tanto prospectivo (a medida que van apareciendo) como retrospectivo (si se buscan en registros ya existentes).

Por tanto, lo que tiene importancia es cómo se obtiene la información, y cuán exactos y válidos son los métodos de medida, ya sean cuestionarios, signos clínicos, parámetros bioquímicos o diagnósticos basados en pruebas complementarias. Por ejemplo, en un estudio de cohortes prospectivo en el que se recoja información al inicio del seguimiento sobre la exposición solar de los participantes mediante un cuestionario para relacionarlo con la incidencia futura de melanoma maligno, será crucial que dicho cuestionario sea válido y esté adaptado a la población del estudio. De lo contrario, el cuestionario no mide lo que debe medir, es decir, no es válido, y por ello la información obtenida estará sesgada, y a pesar de ser un diseño de cohortes y prospectivo los resultados serán erróneos. Esta misma cohorte, o incluso un estudio de casos y controles, a priori más débil que la supuesta cohorte anterior, tendrían más valor si utilizaran un cuestionario debidamente validado y adaptado.

En muchas ocasiones, la disponibilidad de la información condiciona la direccionalidad del estudio. Los estudios de cohortes retrospectivos y los estudios de casos y controles han sido muy importantes para caracterizar las causas de múltiples enfermedades, sobre todo en el ámbito de la salud laboral, gracias a los registros detallados de los puestos de trabajo desempeñados por los trabajadores. Por tanto, los estudios prospectivos no son siempre garantía de resultados más válidos.

3) El tamaño importa

Aunque hayamos simplificado provocadoramente el enunciado de esta falacia, queda claro que nos referimos al tamaño de la muestra que incluimos en los estudios. La estadística justamente nos permite eso: estudiar y sacar conclusiones válidas de una muestra (abordable, finita, localizable) en vez de tener que utilizar toda la población. Y por población debemos entender un conjunto de sujetos en un marco temporal y geográfico determinado, y con unas características determinadas. Por ello, epidemiológicamente son poblaciones "la población española", "la población andaluza de 65 o más años de edad", "la población con cáncer de colon", "los adolescentes fumadores", etc. Y cuando trabajamos con muestras de esas poblaciones debemos tener en cuenta su tamaño (de eso hablamos en esta falacia) y también su representatividad (de eso hablaremos en la siguiente falacia).

Normalmente desconfiamos de un estudio "pequeño" y, al contrario, tendemos a confiar en estudios con un gran número de participantes. Es curioso, porque tampoco sabemos a ciencia cierta si 10, 100 o 1000 sujetos son pocos o muchos. Por ejemplo, si el año 2014 deseamos hacer una revisión sistemática sobre los efectos del tabaco sobre el cáncer de pulmón, dada la gran cantidad de estudios acumulados desde mediados del siglo xx, podríamos decidir excluir a priori aquellos estudios (de casos y controles, o de cohortes) con menos de 100 casos, y seguramente sería muy correcto. ¿Sería correcta esta exclusión si la revisión sistemática nos la hubiésemos planteado, por ejemplo, en el año 1950, cuándo había poca investigación sobre tal asociación? De haber sido así, el Informe del Surgeon General de 1954 que estableció, tras revisar la literatura disponible hasta la fecha, que el tabaco causaba cáncer de pulmón, seguramente no hubiera llegado a esa conclusión, pues habría excluido la mayoría de los estudios existentes hasta entonces.

Un estudio "pequeño", pero bien realizado, puede tener las mismas garantías de validez que uno de mayor tamaño, sobre todo si los hallazgos son relevantes y a pesar del reducido tamaño de la muestra se hacen evidentes. En algunas ocasiones, un reducido tamaño de la muestra puede ser la explicación de la falta de significación estadística, elemento necesario, aunque no imprescindible, para establecer asociaciones. Y al contrario, en un estudio con un gran tamaño de muestra puede aparecer una asociación estadísticamente significativa, pero irrelevante desde el punto de vista clínico, epidemiológico o social. Por echar mano de un ejemplo clásico, imaginemos un gran ensayo clínico en el cual un nuevo tratamiento reduce de forma estadísticamente significativa la presión arterial sistólica una media de 3 mmHg. Las preguntas entonces son: ¿es relevante esa reducción?, ¿debemos dejar el tratamiento convencional e instaurar el nuevo tratamiento en miles de pacientes? (6). Algunas de estas paradojas ya fueron comentadas en la anterior Jornada (11).

El tamaño importa, pero no tanto. Por ejemplo, la caracterización en 1981 del síndrome de inmunodeficiencia adquirida por la infección por un virus hasta la fecha desconocido se debe al estudio de cinco casos ocurridos en tres hospitales diversos. Estos hallazgos se comunicaron como «caso clínico» (un artículo de apenas tres páginas) en *Morbidity & Mortality Weekly Report* (12), la revista de los Centers for Disease Control and Prevention de los Estados Unidos.

El tamaño de la muestra depende de cuatro características del estudio: 1) la magnitud del efecto que se quiere estudiar (cuanto mayor sea su magnitud, necesitaremos mayor número de participantes); 2) el riesgo de un error alfa, es decir, el riesgo de concluir erróneamente que existe una diferencia (habitualmente este error se establece en el 5%); 3) el error beta, es decir, la probabilidad de no encontrar diferencias verdaderas, que habitualmente se fija en el 20%; y 4) la naturaleza de los datos, es decir, el tipo de escala en que se midan los resultados (p. ej., con variables numéricas la variabilidad entre los participantes puede hacer requerir un mayor tamaño mues-

tral, pero esto es algo que no está en manos del investigador).

4) La representatividad importa

Esta falacia, con un enunciado mimético al de la tercera falacia por compartir un cierto origen común derivado del muestreo estadístico, consiste en creer que para hacer generalizaciones válidas de un estudio es necesario que los participantes en el estudio sean una "muestra representativa" de la población diana. Esta falacia se deriva de la idea de que la generalización (lo que también llamamos "validez externa") de un estudio implica la extrapolación mecánica de los resultados de la muestra a la población fuente de la muestra (1). Pero la generalización científica es realmente el proceso de construir una proposición correcta sobre cómo funciona la naturaleza. Este tipo de generalización es el fin último de la ciencia, y por ello los científicos en los laboratorios intentan maximizar la validez interna controlando todas las variables en estudio que pudieran distorsionar los resultados. En los laboratorios experimentales no se seleccionan muestras de animales representativas ni líneas celulares representativas de supuestas poblaciones de animales o células. El investigador básico controla todas las variables, usando animales, microorganismos o líneas celulares genéticamente idénticos y sometidos a las mismas variables ambientales, y que sólo difieren en la variable manipulada experimentalmente. A diferencia de algunas encuestas (electorales o incluso las encuesta de salud) cuya finalidad es estudiar una muestra para poder extrapolar a toda la población diana, la generalización científica se mueve por conjeturas informadas, pero sólo desde la plataforma de estudios válidos.

Un ejemplo clásico son las cohortes del *Nurses' Health Study*, un estudio iniciado en 1976 que actualmente incluye 238.000 enfermeras colegiadas de los Estados Unidos (13). Inicialmente diseñado para estudiar el riesgo del consumo de contraceptivos orales, ha aportado información relevante sobre el efecto de estos fármacos y multitud de otros factores (dieta, tabaquismo, ejercicio físico, etc.) en la incidencia y la mortalidad por numerosas enfermedades. Los resultados obtenidos mediante esta cohorte han sido ampliamente generalizados, a pesar de estar formada por mujeres, enfermeras, con un rango de edad limitado y de nivel socioeconómico bastante homogéneo. La cohorte en sí misma es poco o nada representativa de la población general, e incluso de la población femenina, pero sus resultados son válidos y aplicables científicamente.

5) La significación estadística da validez al estudio

Un error común entre los investigadores con menos formación o experiencia, y quizás en informadores y periodistas mal asesorados, es utilizar la significación estadística como *proxy* de la validez del estudio. La validez de una medición es el grado en que el resultado de esta se corresponda realmente con lo que está sucediendo. De forma análoga, hablamos de validez interna de un estudio cuando los resultados obtenidos en los sujetos de estudio son correctos para esos sujetos. Es decir, la validez de un estudio depende de la ausencia de sesgos o errores sistemáticos que pueden introducirse tanto al seleccionar los sujetos que participan en el estudio (los miembros de la cohorte, los casos y los controles, la población diana…) como al recoger los datos (o realizar las observaciones); o bien por la falta de control de variables que pueden estar interfiriendo las asociaciones estudiadas, los llamados "factores de confusión". Estos errores sistemáticos poco tienen que ver con el azar o error aleatorio (o variación aleatoria) derivado del hecho de estudiar una muestra en vez de toda la población. Justamente, la finalidad de la bioestadística es cuantificar la probabilidad de que el azar o variación aleatoria sea la causa de los resultados en los estudios de investigación.

Por tanto, en muchas ocasiones se obtienen resultados estadísticamente significativos (p. ej., una diferencia entre dos tratamientos de la que estamos muy seguros que no se

debe al azar, o la asociación estadísticamente significativa entre un factor de riesgo y una determinada enfermedad) derivados de estudios con errores sistemáticos que invalidan realmente cualquier resultado. A pesar de ser estadísticamente significativa, la diferencia entre los fármacos podría deberse a que los tratamientos no se asignaron al azar, sino en función del grado de evolución de la enfermedad, que puede determinar a su vez la curación. O, en el otro ejemplo, la asociación causal podría ser justamente la inversa, y tratarse de un factor de protección, puesto que en el grupo de casos solo se incluyeron personas que habían sobrevivido a la enfermedad en estudio.

La significación estadística, ligada normalmente a la precisión estadística o grado de variación aleatoria, no informa ni da ninguna pista sobre la validez interna del estudio.

6) Mucha evidencia significa buena evidencia

Esta falacia, descrita por Petr Skrabaneck y James McCormik en 1989 en su obra *Sofismas y desatinos en la medicina* (14) como «falacia del haz de leña», continúa vigente, y más aún cuando la información científica disponible y accesible se ha multiplicado en las últimas décadas, gracias a la eclosión de Internet. En esencia, consiste en «creer que múltiples elementos aislados de evidencia (simples sospechas o argumentos débiles) poseen mayor fuerza demostrativa cuando se consideran en conjunto». Existe cierta analogía con la frase «una mentira repetida adecuadamente mil veces se convierte en una verdad», atribuida a J. Goebbels, ministro de propaganda del III Reich hitleriano.

Desde hace unas décadas se ha formalizado el metaanálisis como un tipo especial de estudio de síntesis del conocimiento de especial relevancia, con metodología propia para la búsqueda bibliográfica y la selección de los estudios, y también métodos de síntesis estadística propios, en el cual a partir de los datos publicados de los estudios se obtienen nuevos resultados cuantitativos. Esta falacia no tiene tanto que ver con el metaanálisis formalmente bien diseñado y ejecutado, sino con la tentación de algunos investigadores de agrupar resultados no conclusivos de diferentes estudios para llegar a conclusiones por agregación, bastante típico en las revisiones bibliográficas "clásicas". La agrupación cualitativa (como en una revisión clásica) o cuantitativa (como en el metaanálisis) de resultados de diferentes estudios debe basarse en estudios individuales válidos, en los términos de validez (ausencia de errores sistemáticos) ya comentados anteriormente.

7) Las revistas buenas garantizan estudios buenos

La última de las falacias tiene que ver con las revistas, que son el canal de difusión principal de los resultados de la investigación. Las revistas científicas son, en principio, el garante de la investigación. De hecho, una investigación está inconclusa si sus resultados no ven la luz en formato de artículo científico. El funcionamiento de las revistas científicas se basa en el *peer review* o revisión por expertos, que es el proceso mediante el cual los artículos son revisados por expertos (*peers,* pares o semejantes a los autores del artículo) para valorar su validez y mejorarlos (15).

La revisión externa por expertos se ha generalizado como criterio de calidad de las revistas y de su contenido. La existencia de este proceso se tiene en cuenta a la hora de incluir una revista en los repertorios bibliográficos o en los baremos de valoración de los trabajos publicados por un investigador. A pesar del escaso conocimiento empírico que todavía existe sobre su utilidad, sus efectos y sus limitaciones, la revisión por expertos se considera un elemento indispensable para garantizar la calidad en las publicaciones científicas.

Sin embargo, unas revistas tienen más prestigio y otras menos. Este prestigio o relevancia puede intentarse cuantificar a partir del impacto bibliográfico de los artículos que publican (medido a través del factor de impacto, que da cuenta del promedio de citas obtenidas por artículo en una revista en un periodo de tiempo determinado) (16), de la influen-

cia de los artículos sobre el sistema de ciencia e investigación, del impacto de la usabilidad y citación de los artículos en las redes sociales (el movimiento *altmetrics*) (17) o de otros factores acaso altamente intangibles, como el prestigio de los editores o de la sociedad científica o institución promotora de la revista.

«Las buenas revistas llaman a buenos artículos» sería la traslación del «efecto Mateo» (por la frase «al que más tiene, más se le dará» de la parábola de los talentos en el Evangelio de San Mateo) al mundo de las publicaciones científicas. Las buenas revistas atraen no muchos, sino muchísimos manuscritos (*BMJ*, por ejemplo, recibe alrededor de 8000 manuscritos cada año), por lo que tienen mucho donde escoger y pueden ser muy exigentes en términos de validez, relevancia y originalidad (siguiendo con *BMJ*, sólo acepta un 7% de esos manuscritos). Por todo ello, es muy fácil caer en la falacia de atribuir a los artículos las características de la revista donde se publican, lo que en términos generales en epidemiología denominamos la "falacia ecológica" (atribuir al individuo una característica del grupo).

Evidentemente, la mayoría de los artículos publicados en las revistas más prestigiosas serán buenos artículos, pero no siempre será así. Las buenas revistas también se equivocan, y más de una vez encontraremos artículos más que cuestionables (por su metodología, por sus interpretaciones y conclusiones) publicados, que sería lo mismo que decir "refrendados", por una buena revista. Si tomamos como indicador de calidad las citas recibidas por los artículos (un artículo útil, en principio, debería ser más usado y por lo tanto citado por otros investigadores) podemos llevarnos alguna sorpresa. Pues bien, en ciencias médicas, el 12% de los artículos publicados no había recibido una sola cita a los 5 años de haberse publicado, y ese porcentaje era del 27% en las revistas de ciencias naturales (18). Un estudio sobre la no citación en revistas del campo de la inmunología aporta interesantes datos (19): el porcentaje mediano de artículos no citados fue del 17,6%, y el porcentaje de no citación se reducía al 15% o menos solo en las revistas con un factor de impacto mayor de 3. En efecto, a pesar de existir una gran correlación inversa entre el factor de impacto y el porcentaje de artículos no citados (a mayor factor de impacto, menor porcentaje de no citación), llama tremendamente la atención que una revista como *Nature Immunology,* con el mayor factor de impacto de la especialidad en el año 2012 (27,868), tenga una veintena de artículos que han recibido menos de 15 citas.

A pesar de que la mayoría de las revistas de primera fila suelen publicar artículos válidos y relevantes, conviene no bajar la guardia porque un artículo venga refrendado por su publicación en una gran revista. Los investigadores, los informadores y los periodistas no deben dejarse deslumbrar por el prestigio de las revistas. Por todo ello, es recomendable ser siempre críticos en la valoración de cualquier artículo, se publique donde se publique.

Bibliografía

1. Rothman KJ. Six persistent research misconceptions. J Gen Intern Med. 2014;29:1060-4.
2. Porta Serra M. Métodos de investigación clínica: errores, falacias y desafíos. Med Clin (Barc.). 1990; 94:107-15.
3. Porta Serra M. La observación clínica y el razonamiento epidemiológico. Med Clin (Barc). 1986;87:816-9.
4. Álvarez Dardet C, Bolúmar F, Porta Serra M. Tipos de estudios. Med Clin (Barc). 1987;89:296-301.
5. Porta Serra M, Álvarez Dardet C, Bolúmar F, Plasencia A, Velilla E. La calidad de la información clínica (I): validez. Med Clin (Barc). 1987;89:741-7.
6. Plasencia A, Porta Serra M. La calidad de la información clínica (II): significación estadística. Med Clin (Barc). 1988;90:122-6.
7. Porta Serra M, Plasencia A, Sanz F. La calidad de la información clínica (y III): ¿estadísticamente significativo o clínicamente importante? Med Clin (Barc). 1988;90:463-8.
8. Alonso Coello P. La confianza en los resultados de la investigación y el sistema GRADE. En: Casino G, editor. Bioestadística para periodistas y comunicadores. Cuadernos de la Fundación Dr. Antonio Esteve nº 26. Barcelona: Fundación Dr. Antonio Esteve; 2013. p. 25-32.
9. Alonso-Coello P, Rigau D, Solà I, Martínez García L. La formulación de recomendaciones en salud: el sistema GRADE. Med Clin (Barc). 2013;140:366-73.

10. Bullen C, Howe C, Laugesen M, McRobbie H, Parag V, Williman J, et al. Electronic cigarettes for smoking cessation: a randomised controlled trial. Lancet. 2013;382:1629-37.
11. Francescutti P. Diálogo 2. Sobre los estudios observacionales y su tratamiento periodístico. En: Casino G, editor. Bioestadística para periodistas y comunicadores. Cuadernos de la Fundación Dr. Antonio Esteve nº 26. Barcelona: Fundación Dr. Antonio Esteve; 2013. p. 39-45.
12. Centers for Disease Control and Prevention (CDC). Pneumocystis pneumonia – Los Angeles. MMWR Morb Mortal Wkly Rep. 1981;30:250-2.
13. The Nurses' Health Study. Website. (Consultado el 2 de noviembre de 2014.) Disponible en: http://www.channing.harvard.edu/nhs/
14. Skrabanek P, McCormick J. Sofismas y desatinos en la medicina. Barcelona: Doyma; 1992.
15. Plasencia A, García AM, Fernández E. La revisión por pares: ¿buena, mala o todo lo contrario? Gac Sanit. 2001;15:378-9.
16. Porta M, Fernández E, Bolúmar F. Commentary: the "bibliographic impact factor" and the still uncharted sociology of epidemiology. Int J Epidemiol. 2006;35:1130-5.
17. Priem J, Groth P, Taraborelli D. The altmetrics collection. PLoS One. 2012;7:e48753.
18. Remler D. Are 90% of academic papers really never cited? Reviewing the literature on academic citations. The impact blog. London School of Economics and Political Science. (Consultado el 3 de noviembre de 2014.) Disponible en: http://blogs.lse.ac.uk/impactofsocialsciences/2014/04/23/academic-papers-citation-rates-remler/
19. Weale AR, Bailey M, Lear PA. The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor. BMC Med Res Methodol. 2004;4:14.