



El intervalo de confianza se interpreta mejor que el valor de p

Cobo / Ventura

Diálogo 1

Herramientas estadísticas, buenos consejos y cierta intuición periodística

Ainhoa Iriberrí

Los periodistas que acudimos a la jornada de bioestadística organizada el pasado 14 de febrero por la Asociación Española de Comunicación Científica y la Fundación Dr. Antonio Esteve pedíamos algo imposible: que en menos de 12 horas los expertos nos enseñaran a los comunicadores a informar bien sobre un estudio publicado en una revista científica. Se trataba de aprender a evitar algunos de los grandes males del periodismo científico: la exageración, la falta de rigor y la descontextualización, entre otros.

Así, estos conceptos salieron a la luz en el primer diálogo entre periodistas y bioestadísticos, que sirvió también como punto de encuentro. Porque ni los bioestadísticos son esas personas que, con sus enrevesados términos, dificultan muchísimo la comprensión de un estudio, ni los periodistas los irresponsables que algunos

de los primeros pueden pensar. Informamos lo mejor que podemos con nuestros conocimientos y las opiniones de expertos acreditados.

Quizás el error, como quedó de manifiesto en la jornada, es que la bioestadística es una disciplina nueva y algunos investigadores –y la mayoría de los periodistas– carecen muchas veces de conocimientos suficientes sobre la materia. Pero el diálogo entre ambos agentes también demostró que compartimos un deseo común: informar con rigor de los avances científicos.

Pirámides y confianza

Los periodistas estamos acostumbrados a trabajar con pirámides. Desde la famosa pirámide invertida, o regla de las «5 w», sabemos que la información hay que clasificarla y jerarquizarla.

Por ello no nos extrañó cuando se habló del concepto de «pirámide de la evidencia». No todos los estudios científicos son iguales, y no debemos confiar en todos ellos por igual. De nuevo, hemos de recurrir a una pirámide; en este caso, la de la evidencia.

La pirámide ilustra que hay una jerarquía de la confianza que nos merecen los distintos tipos de estudios, y el interés de un grupo de trabajo que comenzó en 2000, el grupo GRADE (siglas en inglés de Clasificación de la Evaluación, Desarrollo y Valoración de las Recomendaciones). El objetivo de este colectivo no era otro que desarrollar un sistema común y razonable para calificar la calidad de la evidencia y la fuerza de las recomendaciones. Un sistema que, obviamente, puede servir también a los periodistas a la hora de elaborar sus informaciones sobre un avance científico.

En la base de la pirámide están los estudios experimentales realizados con animales de laboratorio, luego se encuentran los distintos tipos de estudios observacionales, y en la cúspide los mejores ensayos clínicos y las revisiones sistemáticas (con o sin su correspondiente metaanálisis).

A partir de ahí, podríamos pensar que está todo dicho. A la hora de valorar si informamos sobre un estudio, sólo hemos de situarlo en la pirámide. Si la confianza es baja, no se informará sobre él; si es alta, intentaremos que vaya en portada. Esto que dicta la lógica está, sin embargo, muy lejos de la realidad.

Como los ponentes dejaron muy claro desde el principio, tal jerarquía es flexible, como lo es la clasificación de la confianza que nos ofrece. Estudios observacionales que a priori merecen poca confianza pueden aumentarla si reúnen una serie de requisitos, que la mejoran. Del mismo modo, estimaciones procedentes de un ensayo clínico, incluso aleatorizado –lo que es clave para empezar a confiar en este tipo de estudio– pueden no ser buenas si el ensayo está mal diseñado y ejecutado.

Sin embargo, la confianza en los resultados no es el único criterio que los periodistas seguimos a la hora de informar. Obviamente, el primero será si la información en cuestión nos pare-

ce interesante, es decir, de interés periodístico. Otros factores, como la revista científica en que se publica un determinado estudio, también pesarán en nuestra decisión, ya que no es lo mismo que un trabajo aparezca en una revista de alto impacto o en una de tercera fila. Para saberlo, hay un *ranking* que clasifica las revistas por su factor de impacto (citas recibidas), el *Journal Citation Reports* del Institute for Scientific Information, actualmente Thomson Reuters, por la empresa que lo elabora. Un obstáculo para los periodistas: su acceso es de pago, aunque muchas instituciones, entre ellas la Fundación Española para la Ciencia y la Tecnología, lo ponen a disposición de los investigadores. Una razón añadida para que se cuente entre nuestras fuentes más consultadas.

Revisión frente a metaanálisis

Uno de los primeros conceptos que se discutieron en este diálogo fue la diferencia entre revisión sistemática y metaanálisis. Las revisiones sistemáticas (con o sin metaanálisis) pueden cambiar la práctica clínica y, por tanto, son más que dignas de merecer un hueco destacado en los periódicos. Pero como explicaron los especialistas, no todas las revisiones pueden tener metaanálisis. ¿Por qué? Lo primero que hay que tener claro es que las revisiones “narrativas” del pasado han dado lugar a las revisiones sistemáticas, con una metodología explícita, idealmente especificada en un protocolo. Una vez obtenidos los estudios pertinentes, los incluidos en la revisión en ocasiones pueden metaanalizarse de manera conjunta, pero no siempre, por diversas circunstancias. Asimismo, los estudios pueden no haber medido una variable de manera similar, o haber reportado los resultados numéricos de distintas maneras o incluso de forma incompleta. Una cosa está clara: siempre que sea razonable, los autores llevarán a cabo un metaanálisis, algo que suma valor a la investigación porque aumenta la precisión de los resultados.

En la definición de «metaanálisis» hubo algo de controversia entre los participantes en la jornada. Según se explicó, este tipo de estudio

supone poder combinar pacientes de diferentes trabajos previos. ¿Se trata, entonces, de un artificio estadístico? Algunos de los ponentes prefirieron cambiar esa palabra (artificio) por una más neutra: método. Su argumento fue que proporciona respuesta a dos preguntas cruciales: 1) cuál es el efecto de la intervención en estudio y 2) cuál es el grado de homogeneidad de este efecto entre los diferentes estudios.

En cualquier caso, los expertos nos ofrecieron algunas claves para detectar una buena revisión sistemática. En primer lugar, la revisión de los trabajos ha de ser sistemática, es decir, se ha de buscar todo lo que se ha publicado sobre una determinada hipótesis. «Nos debe alertar que alguien realice un metaanálisis sin haber hecho una revisión sistemática; si no, no hay manera de trazar lo que hay metido ahí dentro, puede que sean los estudios favoritos de los autores. Es la principal cautela que hay que tener a la hora de mirar una revisión», señaló Pablo Alonso.

Al hablar de homogeneidad de estudios para poder incluirlos en un metaanálisis, tampoco hay que exagerar. No se trata de que sean trabajos exactamente iguales, pero sí tienen que intentar responder a la misma pregunta. Por ejemplo, si el metaanálisis evalúa el valor de una intervención terapéutica, ésta ha de ser similar en todos los trabajos, pudiendo variar la dosis, el comparador (por ejemplo, incluyendo placebo y no intervención) e incluso las características de los pacientes, aunque sólo ligeramente. Como dijo Pablo Alonso, «tiene que tener un sentido biológico y clínico a la hora de combinar los estudios, que no chirríe al incluir pacientes o intervenciones completamente diferentes; por eso hay unos criterios de inclusión y exclusión consensuados en las revisiones sistemáticas».

Los periodistas insistían: ¿cómo distinguir qué metaanálisis es el mejor? De nuevo, la clave está en la calidad. «El drama del metaanálisis es la heterogeneidad», dijo Erik Cobo.

Al final se consiguió la receta que los comunicadores estábamos buscando. Cuando se busca la respuesta a una pregunta científica polémica, una buena estrategia es averiguar, en primer lugar, si existe una revisión de la Cola-

boración Cochrane. Ésta es una entidad en la que colaboran más de 28.000 investigadores de más de 100 países (incluido España), y que se dedica a hacer revisiones sistemáticas sobre los temas más variados. Si hay una revisión de este tipo sobre un tema polémico, sus conclusiones son dignas de figurar en la información periodística.

Tenemos así una herramienta que los periodistas científicos hemos de consultar con regularidad, ante informaciones que llegan a la redacción, el estudio de radio o el plató de televisión. Si entramos en la web de la Colaboración Cochrane (www.cochrane.org) y accedemos a *Top 50 Reviews*, veremos algunos ejemplos interesantes, como revisiones sobre si el uso de estatinas previene la enfermedad cardiovascular en las personas sanas, si los suplementos con antioxidantes previenen la mortalidad o si la vitamina C es útil para prevenir los tan frecuentes catarros. Seguro que, como periodistas, hemos recibido informaciones sobre estos temas a lo largo de nuestra carrera; la próxima vez, sabremos qué consultar antes de decidir incluirlas o no. Es aconsejable mirar otras revisiones sistemáticas, pues hay vida más allá de la Cochrane. Una buena fuente es el metabuscador TripDatabase (<http://www.tripdatabase.com/>), que clasifica los resultados en diferentes categorías.

Sin embargo, está claro que los bioestadísticos no son partidarios de las soluciones fáciles. Así, aunque la Cochrane goza de su confianza, Pablo Alonso indicó que «no todas sus revisiones son de igual calidad». Pero la realidad, señalaron, es que hay estudios empíricos que comparan revisiones de esta organización y otras, y concluyen que, en términos generales, las primeras tienen una mayor calidad.

Una revisión sistemática debería tener un protocolo, hacer una pregunta específica y llevar a cabo una búsqueda rigurosa, bien descrita en la publicación. Ejemplos de buenas prácticas son las que consultan más de una base de datos, expertos, referencias bibliográficas, registros de ensayos clínicos e incluso inquieran a la industria farmacéutica. Además, han de hacer una evaluación del riesgo de sesgo, es decir, evaluar de alguna manera la calidad de los estudios

que incluyen. Si queremos profundizar más en la interpretación de una revisión sistemática, es aconsejable consultar una fuente con experiencia. Al fin y al cabo, la metodología de las revisiones sistemáticas, y la estadística en general, es toda una ciencia.

Que la bioestadística sea una ciencia no quiere decir que los periodistas no nos podamos formar en ella. Así, Erik Cobo fue más allá y recomendó a los comunicadores presentes la lectura y el estudio de la guía Prisma,¹ elaborada para evitar la inercia de los investigadores y consistente en una serie de ítems que éstos deben preguntarse para saber si están elaborando correctamente un metaanálisis. Por ejemplo, todo buen metaanálisis ha de incluir información sobre cuándo se empezaron a recoger los datos y cuando cesó su recopilación. Es útil también, por tanto, para evaluar la calidad de los metaanálisis ajenos y no sólo los que se están llevando a cabo.

Erik Cobo recordó por qué se creó esta guía y otras similares, que pueden consultarse en la web de la Red Equator (<http://www.equator-network.org>): «Se supone que los investigadores son los profesionales que rompen fronteras, que consiguen que la sociedad vaya más lejos, pero hay casos clarísimos de cosas absurdas que se hacían en todas las revistas de investigación sobre las que los estadísticos llevábamos años advirtiendo». Antes de que surgieran este tipo de iniciativas, no era raro que los investigadores utilizaran como argumento para emplear una determinada metodología que era la que se utilizaba en una revista de referencia. Por eso, los metodólogos se unieron con placer a la iniciativa para buscar transparencia en las investigaciones científicas, iniciada por el grupo de investigadores y editores médicos herederos del llamado Grupo de Vancouver, que lideró la mejora de la calidad de los originales científicos. ¿Era esto necesario? «Hay una argucia que se utiliza mucho: un investigador pide al estadístico que le haga el análisis lo más complicado posible, para que revisores y editores no entiendan nada y no les quede más remedio que tragar; esto se usa y es contrario a las ideas originales de los estadísticos», apuntó Erik Cobo.

La p y el intervalo de confianza

El bioestadístico añadió algunas afirmaciones sin duda curiosas. Por ejemplo, que p (ese número que establece hasta qué punto una hipótesis puede ser debida a la casualidad) «ha de jubilarse» y que es mejor fijarse en otro parámetro, como es el intervalo de confianza. «Los investigadores, al publicar, siguen la tradición. Por eso, esta jubilación sólo se conseguirá si disponemos de guías sobre cómo publicar artículos. No olvidemos que las revistas científicas no publican artículos buenos, sino los mejores que les llegan», afirmó.

Los periodistas presentes en la jornada preguntamos por ejemplos prácticos. Al fin y al cabo, teníamos que salir de esta reunión sabiendo cazar al vuelo un metaanálisis y diferenciarlo, además, de otros tipos de estudios.

El ejemplo clásico, señaló Pablo Alonso, es una revisión sistemática publicada en la década de 1990 que evaluaba la eficacia de un fármaco, la lidocaína, para prevenir arritmias en los pacientes que habían sufrido un infarto agudo de miocardio. El medicamento se aplicaba desde 20 años atrás. Algunos estudios observacionales parecían demostrar un posible efecto beneficioso en la disminución del riesgo de muerte, pero algunos investigadores empezaron a cuestionar la idea y se iniciaron ensayos aleatorizados, que compararon la evolución de los pacientes tratados con lidocaína con la de aquellos que no recibían el fármaco. Aunque estos estudios por separado apuntaban (de manera no concluyente) que los enfermos que tomaban el medicamento morían más, se continuaron haciendo ensayos clínicos y hubo que esperar a que alguien realizará una revisión sistemática (con metaanálisis) para que cambiara la práctica clínica y se abandonara el uso de lidocaína con este fin. Dicho metaanálisis se publicó en la década de 1990, pero si se hubieran metaanalizado los estudios llevados a cabo en la década de 1980 ya se hubiera visto que el efecto era claramente perjudicial. Conclusión: durante muchos años, miles de personas fueron tratadas innecesariamente y con consecuencias muy graves. ¡Como para no valorar las revisiones sistemáticas!

Estudios observacionales

Cuando parecía que ya todo estaba explicado, el epidemiólogo José Luis Peñalvo, del Centro Nacional de Investigaciones Cardiovasculares, añadió un nuevo dato en el cual fijarse: que no es lo mismo un metaanálisis de estudios experimentales que uno de trabajos observacionales.

Los estudios observacionales ocupan muchas veces los distintos medios de comunicación. Cada vez que leemos informaciones que asocian un determinado factor de riesgo a un efecto, es más que probable que la información venga de un trabajo de este tipo, por lo que conviene saber un poco más sobre ellos. A fin de cuentas, como observó Cobo, la búsqueda de causas es la pasión favorita de las mentes más racionalistas, a diferencia de la confirmación de efectos que persiguen las más pragmáticas y empíricas.

La primera clave sobre los estudios observacionales es básica para el periodismo. Con los estudios observacionales no puede inferirse causalidad. «Lo único que podemos deducir es una asociación o una relación y, dentro de ésta, todos los grados de gris: fuerte, modesta, no significativa, una tendencia...», señaló Peñalvo. En la escala de fuerza de asociación, hay tres tipos de diseño: de casos y controles, transversales y de cohortes.

Los estudios de casos y controles responden a un tipo de diseño observacional que es muy atractivo, porque es muy rápido. No hace falta mucha infraestructura y son perfectos para estudiar hipótesis sobre enfermedades de baja prevalencia. ¿Cómo se llevan a cabo? Se acude a un hospital y se seleccionan todas las personas que ingresen con una determinada enfermedad, a las que se denominará «casos». Después se establecen una serie de criterios para escoger unos «controles», que estarán emparejados por tener similitudes con los casos. Posteriormente, se utilizan registros o se hacen preguntas a los participantes, por ejemplo sobre elementos a los que están expuestos, y se establecen asociaciones. Son, en definitiva, estudios rápidos y fáciles de hacer que, en ocasiones, se realizan dentro (anidados) de estudios de cohortes.

Las cohortes tienen, por definición, una duración larga. El ejemplo clásico es la cohorte del estudio Framingham, un trabajo que empezó en 1948 en el pueblo de la costa este de Estados Unidos del mismo nombre y que sirvió para establecer los factores clásicos de riesgo cardiovascular. Actualmente se está estudiando la tercera generación, los nietos de los primeros voluntarios que accedieron a que un equipo de cardiólogos les vigilara de por vida. Ahora, los factores que se estudian están más centrados en la herencia.

Por último, son importantes los estudios observacionales transversales, también muy habituales y muy presentes en los medios de comunicación. Un ejemplo claro son las encuestas nacionales de salud (en España se realizan cada 2 años). Se trata de establecer tendencias con datos a partir de estudios observacionales. Un ejemplo de este tipo de estudio son las informaciones que, a partir de combinar la información de varios «cortes» transversales en la población, relacionan la entrada en vigor de la Ley Antitabaco con la disminución de episodios cardiovasculares o de muertes por esta causa.

La importancia de la causalidad

En realidad, destacaron los expertos, nunca podrá decirse con total seguridad que una circunstancia ha llevado a la otra, y como ocurre siempre con la epidemiología observacional, la relación no necesariamente será causal. El debate se complicó aún más cuando Erik Cobo reflexionó en alto: «¿Y si, en ocasiones, no importara realmente si es causal o no?». Así, este estadístico habló de un ejemplo concreto: los estudios que asocian las modificaciones de las leyes de seguridad vial con la reducción del número de muertos en la carretera. «Si decidimos titular *La nueva Ley ha evitado 100 muertes* no lo hacemos bien; si titulamos *Desde la entrada en vigor de la Ley ha descendido el número de muertes*, el titular es neutro y correcto», explica Cobo, que sin embargo añadió: «A mí, como ciudadano, los dos titulares me son útiles. A veces los estadísticos nos ponemos muy antipáticos», concluyó.

De este primer diálogo entre periodistas y estadísticos queda claro que existen muchas herramientas que los primeros podemos utilizar para informar correctamente sobre las investigaciones científicas. No obstante, es evidente que sigue haciendo falta formación, una cierta intuición y, sobre todo, consultar a los expertos. Sólo que ahora, en la agenda, habrá que añadir un apartado al de investigadores, cardiólogos, físicos o cualesquiera categorías que utilicemos

habitualmente: los bioestadísticos, que sin duda tienen mucho que decir.

Bibliografía

1. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6:e1000097. Disponible en: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000097>