



Un riesgo que crece del 1% al 2% puede significar un aumento del 1% o del 100%, según se mire

González / Ventura

Talleres Análisis de *papers*, comunicados de prensa y artículos periodísticos

Esperanza García Molina

En la jornada de bioestadística para periodistas, después de los diálogos se llevaron a cabo talleres prácticos para analizar informaciones científicas con contenido estadístico, valorar su difusión en los medios de comunicación e identificar los problemas que estos casos podrían plantear a los periodistas en el ejercicio de su profesión. Para trabajar sobre estas cuestiones, los participantes en los talleres recibieron un *dossier* con varios ejemplos, cada uno de ellos compuesto por un artículo científico o *paper*, el comunicado de prensa redactado por la institución científica responsable de la investigación y artículos periodísticos reales escritos sobre el estudio. Además, en el *dossier* se planteaban una serie de cuestiones que los participantes debían resolver en cada ejercicio. Las preguntas eran las siguientes: ¿qué tipo de estudio es?, ¿cuál es el *outcome*

o criterio de valoración?, ¿es adecuado el diseño del estudio?, ¿qué resultados ofrece?, ¿cuál es su mensaje principal?, ¿cuál es su nivel de confianza?, ¿es consecuente el *press release*?, ¿merece la pena informar?, ¿qué cautela habría que tener al informar? y ¿es correcto el titular periodístico?

Los participantes se organizaron en grupos de tres o cuatro personas con un representante para responder a las cuestiones que se planteaban en los talleres. Erik Cobo y Pablo Alonso fueron los dos expertos que dirigieron los talleres, junto al periodista Gonzalo Casino.

En este capítulo sobre los talleres prescindiremos de publicar íntegramente los textos que se utilizaron para no incurrir en violaciones de derechos de reproducción de los artículos, tanto los periodísticos como los científicos. No obstante,

facilitamos las referencias bibliográficas suficientes para encontrarlos. En el resumen de estos talleres se incluyen párrafos textuales, con las referencias a sus publicaciones, para poder seguir los ejemplos.

Aunque el *dossier* incluía más ejemplos para trabajar, durante los talleres sólo se escogieron tres. Por falta de tiempo, únicamente en el primero se completaron las 10 cuestiones planteadas; en los otros dos, nos limitamos a analizar en detalle el artículo científico, pero no su difusión en los medios.

Primer ejemplo: relación entre la miopía y la exposición a la luz en niños menores de 2 años

El primer ejemplo analizado en los talleres se basó en un artículo publicado en *Nature* en el año 1999 sobre miopía y medio ambiente, con el título *Myopia and ambient light at night*.¹

Gonzalo Casino sugirió hacer primero una lectura del artículo en diagonal, pensando en una situación real de trabajo en la redacción, con pocas horas para valorar si hacerse eco o no de ese artículo y, en caso afirmativo, cómo hacerlo. Más tarde se respondieron las 10 preguntas para valorar el paso del *paper* a la noticia.

El artículo trataba sobre un estudio en padres de niños de 2 a 16 años de edad, pacientes de una clínica oftalmológica pediátrica. Los padres completaron un cuestionario acerca del nivel de exposición a la luz que habían tenido sus hijos antes de cumplir 2 años. Concluía que la prevalencia de la miopía en la infancia tenía una fuerte asociación estadística con la exposición a la luz durante el sueño en los primeros 2 años de vida, y que esta relación era dependiente de la dosis.

La primera cuestión planteada en el taller consistía en delimitar de qué tipo de estudio se trataba. Los participantes estaban de acuerdo en que no era un estudio de intervención, sino observacional. ¿Pero de qué tipo? ¿Cohortes, series de casos, o casos y controles? Se recordó que el criterio básico de un estudio de cohortes es que la muestra esté seleccionada por la exposición. Es un criterio clave que ayuda a distinguir el tipo de estudio. En este caso, la muestra se selec-

cionó tomando en cuenta a los pacientes que habían acudido a la consulta de un oftalmólogo y se les pidió que cumplimentaran una encuesta, en la cual se reflejaban hechos que habían sucedido con anterioridad. Erik Cobo aclaró que seleccionar a los participantes por la exposición no significa seleccionar sólo a los expuestos, sino escoger un grupo de individuos y medir en ese momento su grado de exposición.

Está claro, por tanto, que no se trata de un estudio de casos y controles. No se seleccionan los casos y después se buscan los controles, sino que se eligen pacientes, y de ellos unos tienen miopía y otros no. Más tarde se pregunta si de pequeños han estado expuestos a la luz durante el sueño. Como no han sido seleccionados en función de su exposición ni de su resultado (miope/sano), es una serie de casos. Las series de casos lo son independientemente de su tamaño: pueden ser de dos o tres casos, o de muchos. Aunque las dos variables se recogen en el mismo momento del tiempo, una de ellas hace referencia al pasado, por lo que más investigadores lo clasificarían como longitudinal retrospectivo que como transversal.

¿Cuál es el criterio de valoración en esta serie de casos o qué se está midiendo? La relación entre la miopía y la exposición a la luz. ¿Y qué se encuentra? Primero, que hay una asociación fuerte de la miopía con la exposición a la luz durante el sueño. Además, que es dependiente de la dosis, una de las cosas que pueden aumentar la credibilidad del estudio: cuanto mayor es la exposición a la luz, mayor es el riesgo de desarrollar miopía.

¿Es adecuado el diseño? Es un diseño poco fiable. En la pirámide que jerarquiza la calidad de los tipos de diseño está cerca de la base, y no permite establecer causalidad. Su nivel de confianza es muy bajo. Además, el estudio se ha hecho con una muestra de participantes obtenida de una clínica terciaria, que es una muestra difícilmente extrapolable a la población general.

¿Qué se echa de menos en la metodología? Que el estudio no se ajusta por la herencia familiar de miopía, pues no se estudia en cada caso si los padres son miopes o no. Además, el estudio se basa en lo que han respondido los

padres en un cuestionario sobre la exposición de sus hijos a luz hace muchos años; padres que, además, han llevado a sus hijos a un oftalmólogo. Una recogida de datos retrospectiva con un intervalo temporal en algunos casos de hasta 16 años es poco fiable. Gonzalo Casino señala que en este estudio interviene el sesgo de memoria. Una de las asistentes a la jornada añade que se está preguntando a padres cuyos hijos tienen problemas de visión si les dejaban la luz encendida de pequeños, y es fácil que cambien sus recuerdos en busca de justificaciones para la miopía de sus hijos.

En este punto, Erik Cobo interviene para señalar que el sesgo de memoria puede ser inocente, en el sentido de añadir solamente más variabilidad y ruido, o puede no ser inocente, «porque es posible que los padres de los niños que han desarrollado miopía recuerden detalles de ese tipo con mayor facilidad, puesto que se fijaron en ellos». La clave para hacer un estudio más fiable, añade Cobo, reside en que la recogida de las dos piezas de información sea lo más independiente posible, que no pueda haber una conexión entre ellas. Esto se consigue mediante enmascaramiento. En un estudio de cohortes prospectivo se recoge la información de las dos partes de forma enmascarada. Cuando se recogen datos sobre la exposición nadie sabe lo que pasará más tarde, y quienes toman los datos de la evolución no tienen delante los datos de exposición inicial.

¿Por qué hay relación entre miopía y dosis? Alonso comenta que puede ser causal o, por el contrario, deberse a que aquellos padres más miopes expusieron a sus hijos a mayores intensidades de luz porque ellos mismos la necesitaban. Es otro factor de confusión. ¿Por qué se publicó este estudio en *Nature*, una de las revistas científicas de mayor índice de impacto? Alonso responde que la revisión por pares de una revista no es un salvoconducto que garantice la fiabilidad de un estudio.

Entre el público, la periodista Ainhoa Iriberrri insiste en que el hecho de que la muestra se haya recogido en una clínica resta validez al estudio. Cobo aclara que de esta manera se está reduciendo la ventana experimental, es decir, la

perspectiva que se tiene en los datos. El espectro es más reducido y resulta más difícil encontrar relaciones, pero si se encuentran, la siguiente pregunta debe ser si la relación encontrada es aplicable a una ventana más amplia. Por otra parte, el análisis estadístico y el valor de p podrían ser correctos, ya que no se distingue entre relación y relación casual.

A continuación se analiza la nota de prensa que se preparó para difundir el estudio. En la nota no se informa de que los casos han sido seleccionados en una clínica. Un participante en el taller cree preocupante que la nota, por un lado, diga que no puede establecerse una relación de causa-efecto entre la exposición a la luz y la miopía, y sin embargo que por otro lado sugiera que no es aconsejable exponer a los niños a la luz mientras duermen. Cobo responde que si la nota dijera «los investigadores interpretan que los niños deben dormir con poca luz» sería impecable, porque es cierto que lo han interpretado así. Alonso añade que el titular de la nota de prensa es correcto: *Near sightedness in children linked to light exposure during sleep before age 2. Linked*, «relacionado» en inglés, es un término muy suave, pero no debería aconsejar un cambio de actitudes ante un estudio con tantos problemas de fiabilidad.

Seguidamente se pasó a analizar los artículos periodísticos. El periódico español *El País* dedicó el 17 de mayo de 1999 un reportaje a este estudio, con el título *Los bebés no deben dormir con luz*, un titular que, a la vista del análisis anterior, parece desafortunado. Alonso señala que la fuente independiente escogida por la autora del reportaje para contrastar la fiabilidad de los resultados, Eduard Estivill, es un médico especializado en temas de sueño, pero no oftalmológicos. En este punto, Casino incide en la importancia de consultar fuentes competentes en estadística. En este caso, haría falta un experto que supiera indicar al periodista el principal punto flaco del estudio: que no se ha controlado la herencia.

Casino llama la atención sobre el artículo publicado en el periódico británico *The Guardian*, con el título *Babies left in the dark see way to a brighter future* (13 de mayo de 1999), que contiene errores importantes de interpretación. «Está escrito por

Tim Radford, considerado uno de los mejores periodistas científicos de los últimos años; una prueba de que todos podemos meter la pata». Para Casino, lo más interesante de los errores que se cometieron con esta noticia es observar cómo a la prensa se le fue de las manos la información sobre un estudio que presentaba muchas debilidades. Alonso añade que hay que tener especial cuidado con noticias sobre mensajes muy novedosos con diseños observacionales pobres.

El artículo de *The Guardian* no contiene fuentes independientes del estudio, sólo cita a los propios autores. Es más, se atribuye declaraciones propias del autor que no son tales, porque son copia literal de la nota de prensa. Casino continúa poniendo como ejemplo de buena cobertura periodística el tratamiento que da el periódico estadounidense *The New York Times* a este artículo científico, al cual dedicó una noticia el 13 de mayo de 1999, a la vista del interés periodístico del tema, pero en el penúltimo párrafo indicaba que el estudio era prematuro, incompleto y que no había tenido en cuenta un factor obvio como la herencia. Es decir, se hace eco del estudio, pero lo analiza con fuentes independientes, especialistas en oftalmología, que supieron leer el artículo con perspectiva y ofrecer una visión crítica al periodista. En este sentido, Alonso opina que el periodista puede obtener una ayuda más valiosa del especialista al que consulta si, en lugar de darle a leer el *paper* sin más, le señala los puntos donde sospecha que puede haber problemas.

¿Los periodistas deberían haberse hecho eco del estudio o no? Casino opina que es cuestión de perspectivas, pero lo que hizo el *The New York Times* es perfecto. Además, el artículo se publicó sin firmar, lo cual es una lección de periodismo. Un participante reflexiona que, si medios como *El País* y *The Guardian*, con suficientes recursos, metieron la pata, ¿qué no pasará en medios menores o sin especialistas en ciencia en las redacciones? Para Casino, «lo fundamental es recordar que, en este caso, lo que falta es un buen conocimiento del contexto del problema, que no sabemos todavía cuáles son las razones que provocan la miopía». Una manera correcta de enfocar este estudio, según el periodista, es utilizarlo como percha para hablar de la

miopía, recordar que no se conocen los factores que la provocan y plantear que este artículo ofrece una nueva hipótesis aún sin confirmar.

A favor de *El País* se recuerda que, más tarde, publicó un segundo reportaje sobre una revisión posterior en *Nature*, diciendo que no se habían podido demostrar las conclusiones del primer estudio y señalando sus limitaciones.

Segundo ejemplo: las pelirrojas sufren más miedo al ir al dentista

El artículo de este segundo caso se publicó en diciembre de 2012 en la revista *Journal of Endodontics*, con el título *Anesthetic efficacy of the inferior alveolar nerve block in red-haired women*.² El objetivo del estudio, que en su resumen se define como aleatorizado, era medir la posible relación entre la presencia del alelo de un gen (que tienen las personas pelirrojas) y la eficacia de la anestesia para el nervio alveolar inferior. Participaron en el estudio 62 mujeres pelirrojas y otras 62 con el pelo oscuro. Se les hizo responder un cuestionario para medir la ansiedad. Después se les administró la anestesia dental y se midió su efecto. No se encontraron diferencias en el efecto de la anestesia entre mujeres pelirrojas y morenas. Sin embargo, las pelirrojas manifestaron mayores niveles de ansiedad.

Este ejercicio se plantea, en primer lugar, con el objetivo de que los participantes en el taller sean capaces de averiguar el nivel de evidencia del estudio. Cobo recuerda la pirámide de la evidencia para analizar qué tipo de estudio es, y avisa de que tiene trampa. ¿Qué tipo de estudio es? Se trata de un estudio experimental. Es fácil reconocerlo porque en el texto aparece la palabra «aleatorizado».

¿Cuál es el *outcome*? La medida entre 0 y 100 del nivel de ansiedad de las participantes en el estudio mediante una escala de ansiedad. ¿Es adecuado el diseño? En principio, si es aleatorizado, parece que sí. ¿Qué resultados ofrece? Que las pelirrojas tienen más ansiedad al ir al dentista. ¿Pero en qué se sustenta esta afirmación? Realmente sólo en los resultados del test previo al experimento. En el estudio intentan es-

tablecer una relación entre la presencia de un gen concreto en las personas pelirrojas y una menor efectividad de la anestesia, pero los resultados de esta parte del ensayo indican que no existe dicha relación. La anestesia les hace el mismo efecto que a las morenas. Por lo tanto, el único resultado que hallan es la respuesta ansiosa que revela el test. Al leer el artículo se empieza a notar que hay algo raro.

Las conclusiones del artículo indican que el pelo pelirrojo y la presencia de este gen están ligados de manera significativa a un mayor nivel de ansiedad. Se comparan los resultados de una muestra en la cual la mitad de las mujeres son pelirrojas y la otra mitad no lo son; hasta ahí todo parece correcto, está equilibrado. Prosigue Cobo recordando qué significa la causalidad entre dos variables A y B: que al cambiar la variable A se consigue una modificación en la variable B. Pero, si no se puede variar A, ¿para qué preguntarse si la relación con B es causal? La pregunta clave en este caso es: ¿qué se ha aleatorizado en este estudio? No puede decidirse si una mujer será pelirroja o no, si tendrá o no un alelo.

En realidad, los investigadores están haciendo un estudio de asociación, e introduciendo aleatorización en una variable secundaria, de adorno según Cobo; esta variable secundaria es el lugar donde ponen la inyección de anestesia, en la mandíbula derecha o en la izquierda. Se confunde a los periodistas y a los editores de la revista haciendo creer que es un estudio aleatorizado, del máximo nivel de evidencia, cuando las variables de interés son observadas: el pelo pelirrojo, la presencia de un alelo y el nivel de ansiedad.

Cobo explica que normalmente pueden aleatorizarse pocas variables dentro del mismo estudio. Los estudios médicos aleatorizan una, que es la variable sobre la cual se hace la intervención. Pero el pelo pelirrojo es imposible de aleatorizar, porque no puede intervenir en esta condición. Gonzalo Casino apunta otro detalle. El artículo se extiende al hablar de genética, cuando la publicación es una revista de endodoncia, en la que no escriben científicos sino endodoncistas. Es un estudio de baja calidad y nivel de evidencia, que tuvo mucho impacto en los medios por ser llamativo para el público y porque la nota de prensa

se publicó en Eurekalert. Después de estas conclusiones, no se siguió analizando el caso y se pasó al siguiente ejemplo.

Tercer ejemplo: ¿son los efectos clínicos de la homeopatía efectos placebo?

El siguiente ejemplo es un artículo de revisión sobre la ineficacia de la homeopatía. Se publicó en *The Lancet* en 2005 con el título *Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy*.³

Los autores hicieron una revisión de ensayos clínicos con tratamientos homeopáticos y fármacos convencionales para estimar en qué medida podían estar afectados por sesgos. La búsqueda de estudios se realizó en 19 bases de datos, y la selección de ensayos en el Registro de Ensayos Clínicos Controlados Cochrane, asumiendo que aquellos que estuvieran realizados con doble ciego y aleatorizados eran de calidad superior. Al restringir el análisis a los estudios de alta calidad, se obtuvo una *odds ratio* de 0,88 con un intervalo de confianza de 0,65 a 1,19 para la homeopatía, y una *odds ratio* de 0,58 con un intervalo de confianza de 0,39 a 0,85 para la medicina convencional. La conclusión de este estudio es que hay una evidencia muy débil de los efectos específicos de la homeopatía y una fuerte evidencia de los efectos de los fármacos convencionales. Esto es compatible con la hipótesis de que los efectos de la homeopatía apenas se distinguen de los del placebo.

El análisis de este caso en el taller fue muy útil para que los periodistas entendiéramos el significado de los conceptos de *odds ratio* y de intervalo de confianza. Cobo explicó que, entre las medidas del riesgo, la más sencilla es el riesgo relativo. Se construye midiendo la proporción de personas que experimentan cierto suceso en un grupo de estudio, la misma proporción en un grupo de referencia, y después haciendo el cociente entre ambas.

Si, por ejemplo, se mide el efecto de un tratamiento frente al placebo, y vale 0,88, esto significa que el riesgo en los pacientes tratados dis-

minuye un 12% $([0,88-1] \times 100)$ respecto a los no tratados. Es decir, los pacientes sometidos a ese tratamiento tienen un 12% menos de riesgo de sufrir el suceso que los que toman placebo. En el caso que se analiza, el intervalo de confianza para las terapias homeopáticas es de 0,65 a 1,19. Comprende desde una reducción del riesgo del 35% $([0,65-1] \times 100)$ hasta un aumento del 19% $([1,19-1] \times 100)$, todo en términos relativos, no absolutos.

Continuamos con el ejercicio. ¿Qué tipo de estudio es? Es una revisión sistemática. Hay dos claves para reconocerlo. El estudio reúne lo que hay publicado sobre un tema, es una búsqueda exhaustiva. Además, evalúa la calidad de los estudios y para ello tiene en cuenta que los ensayos sean a doble ciego (para evitar el sesgo de realización) y que tengan la adecuada aleatorización (para eliminar el sesgo de selección). Se asume que esos estudios tienen una metodología de mayor calidad y se escogen porque los pacientes son dos grupos comparables, y ni pacientes ni médicos saben qué están tomando o prescribiendo.

¿Cuál es el criterio de valoración? La revisión compara ensayos clínicos de medicina convencional de múltiples intervenciones frente a placebo, con todos los ensayos que encuentra de intervención homeopática frente a placebo. De los homeopáticos hallaron más de 100 y buscaron otros tantos equivalentes de medicina convencional, que estudiaran las mismas enfermedades y tratamientos parecidos, aunque evaluaran cosas diferentes. Luego metaanalizaron con un análisis estadístico conjunto todos los resultados de los estudios que comparaban homeopatía frente a placebo, independientemente de lo que estuvieran comparando, y obtuvieron un estimador global. Hicieron lo mismo con todos los estudios de tratamiento convencional frente a placebo equivalentes a los de homeopatía, y así llegaron a tener dos estimadores.

Una de las conclusiones del artículo es que en ambos tipos de estudios, de homeopatía y de medicina convencional, los de baja calidad mostraron un beneficio importante, mayor que los de alta calidad. Puede que, por intereses comerciales, sólo se publiquen los casos positivos. Pero también

puede haber un riesgo de sesgo porque los ensayos no estén bien aleatorizados o controlados.

No obstante, el análisis se hizo únicamente para los estudios de alta calidad. La *odds ratio* en los estudios de homeopatía, con un valor de 0,88, es no significativa, pues sólo representa una mejoría del 12% respecto al placebo, explica Alonso. El intervalo de confianza, de 0,65 a 1,19, indica que la precisión es baja. Muestra tanto importantes beneficios como perjuicios, por lo que estos estudios no son clínicamente relevantes. En este punto, Alonso aclara que cuando un intervalo de confianza incluye el 1, esto quiere decir que el tratamiento estudiado puede no tener ningún efecto positivo respecto al placebo.

Cobo apunta que esto no quiere decir que ese tratamiento sea perjudicial, sino que el efecto nulo es uno de los valores compatibles con los resultados observados para ese tratamiento; pero hay otros, unos positivos y otros negativos. «Aquí aplica aquello de que ausencia de pruebas no es prueba de ausencia», aclara Cobo. Alonso recalca que, para que pueda afirmarse que un tratamiento reduce el riesgo, el intervalo de confianza debe estar por debajo de 1. No obstante, aclara que, aun así, puede no ser clínicamente relevante por reducir muy poco el riesgo, resultar muy caro, tener efectos secundarios, etcétera. El umbral varía en función de los efectos adversos y de los inconvenientes.

Seguimos analizando el intervalo de confianza en estudios equivalentes de alta calidad en medicina convencional. Para estos casos, la *odds ratio* es 0,58, lo que significa que los tratamientos convencionales reducen el riesgo un 42% respecto al placebo. El intervalo de confianza (0,39 a 0,85) está, en el peor de los casos, por debajo de la unidad. La medicina convencional demuestra tener resultados positivos.

El intervalo de confianza cuantifica la magnitud de la incertidumbre, añade Cobo, pero no debe extraerse de él la idea de que el efecto es diferente en distintos pacientes. Una de las premisas es que el efecto es el mismo en una misma población. Asumiendo que el efecto es el mismo para todos los pacientes, el intervalo de confianza indica cuáles son los valores plausibles para este efecto común en una población.

Finalmente, se terminó este ejercicio con la conclusión de que es una revisión sistemática con un alto nivel de confianza.

Bibliografía

1. Quinn GE, Shin CH, Maguire MG, Stone RA. Myopia and ambient lighting at night. *Nature*. 1999;399:113-4.
2. Droll B, Drum M, Nusstein J, Reader A, Beck M. Anesthetic efficacy of the inferior alveolar nerve block in red-haired women. *J Endod*. 2012;38:1564-9.
3. Shang A, Huwiler-Müntener K, Nartey L, Jüni P, Dörig S, Sterne JA, et al. Are the clinical effects of homoeopathy placebo effects? Comparative study of placebo-controlled trials of homoeopathy and allopathy. *Lancet*. 2005;366:726-32.