

COMUNICACIÓN BIOMÉDICA

Para cumplir con el doble propósito de desmitificar la preparación de artículos para publicación y agudizar el juicio crítico de autores y lectores, en esta sección se incluyen trabajos diversos de la pluma de expertos en investigación científica, criterios de diseño de estudios, aplicación de métodos estadísticos, edición de textos y otros temas de importancia para la comunicación biomédica.

En este número presentamos, primeramente, un clásico de la salud pública, el discurso pronunciado por Sir Austin Bradford Hill ante la Sección de Medicina Ocupacional de la Royal Society of Medicine el 14 de enero de 1965. Los criterios expuestos por Hill para valorar la relación causal entre factores ambientales y enfermedades gozan de aceptación general en la comunidad científica. Acompaña a este trabajo la interesante comparación hecha en época reciente por Alfredo Morabia de los criterios propuestos por Hill y las reglas para determinar causación formuladas en el siglo XVIII por el filósofo escocés David Hume. El pensamiento de Hume expresa un punto de vista sorprendentemente similar al de Hill y los epidemiólogos actuales, pero opuesto al del filósofo vienés Karl Popper. Por último, el artículo de Marcia Angell revela cómo el equipo editorial de la revista The New England Journal of Medicine evalúa los estudios epidemiológicos presentados para publicación, en particular, los que tratan de asociaciones entre enfermedades y riesgos. Esos estudios observacionales plantean varios problemas de diseño y realización, y requieren una cautelosa interpretación crítica.

AMBIENTE Y ENFERMEDAD: ¿ASOCIACIÓN O CAUSACIÓN?¹

Austin Bradford Hill²

El primer objetivo de la Sección de Medicina Ocupacional recién organizada es "constituir un foro de discusión para que los médicos y los cirujanos con especial conocimiento de la relación entre enfermedades y lesiones y condiciones de tra-

¹ Publicado originalmente con el título "The environment and disease: association or causation?" (*Proceedings of the Royal Society of Medicine*. 1965;58:295-300). Se traduce con autorización de *Journal of the Royal Society of Medicine*

² Profesor Emérito de Estadística Médica, Universidad de Londres.

bajo puedan discutir sus problemas, no solo entre ellos sino también con colegas de otras especialidades, en reuniones conjuntas con otras secciones de la Sociedad". El segundo objetivo es "hacer accesibles los conocimientos sobre riesgos físicos, químicos y psicológicos del trabajo, en particular los que son raros o difíciles de reconocer".

En esta primera reunión de la Sección, y antes de que con loables intenciones comencemos a adoctrinar a colegas de otros campos, es fundamental que consideremos un problema básico de nuestra especialidad. ¿Cómo detectamos la relación entre enfermedades, lesiones y condiciones de trabajo? ¿Cómo definimos los riesgos físicos, químicos y psicológicos del puesto de trabajo, en particular cuando son riesgos raros o que no se reconocen fácilmente?

Por supuesto, hay veces en las que es fácil dar respuesta a estas preguntas a partir del cuerpo general del saber médico. Ciertos ambientes físicos, a veces extremos, son inevitablemente nocivos; una sustancia que se sabe tóxica para las personas será sospechosa cuando se encuentre en las instalaciones de la fábrica. Otras veces lo que hemos de considerar son los *efectos potenciales* de un ambiente determinado sobre los seres humanos y, de seguido, determinar esos efectos, si es que existen. Pero lo más frecuente es que no tengamos ninguna de esas orientaciones ni guías para proceder. Lo habitual es que dependamos de nuestras observaciones y del registro de ciertos sucesos para los que luego buscamos antecedentes. En otras palabras, vemos que el suceso B se asocia con la característica ambiental A. Pongamos un ejemplo concreto. Imaginemos que cierta forma de enfermedad respiratoria se asocia con la presencia de polvo en el ambiente. ¿En qué circunstancias podemos determinar la existencia de *causación* a partir de la *asociación* observada? ¿En qué debemos fundarnos para proceder de esa manera?

No tengo deseos ni capacidad para embarcarme en una discusión filosófica sobre lo que significa "causación". La "causa" de enfermedad puede ser inmediata y directa, o remota e indirecta, subyacente a la asociación observada. Pero desde el punto de vista de la medicina ocupacional, que es casi como decir de la medicina preventiva, la cuestión decisiva es si la frecuencia del suceso indeseable B resulta modificada por un cambio en la característica ambiental A. Determinar *cómo* es posible que ese cambio ejerza tal influencia puede exigir una investigación enorme. Sin embargo, para concluir "causación" y tomar medidas, no siempre tenemos que esperar sentados los resultados de dicha investigación. Puede ser necesario desenredar toda la cadena, pero a veces es suficiente con poner a la luz unos cuantos eslabones. Dependerá de las circunstancias.

Dejemos a un lado cualquier problema semántico que pueda presentarse en tal situación. Las observaciones de dos variables revelan una asociación perfectamente definida y que sobrepasa en mucho lo que podríamos atribuir al azar. ¿Qué aspectos de dicha asociación debemos tener en cuenta para decidir que la interpretación más aceptable es la existencia de una relación causal?

1) Intensidad. La intensidad de asociación sería para mí el primer criterio. Ejemplo ya clásico es el de Percival Pott, que al comparar las ocupaciones de pacientes con cáncer

de escroto con las de pacientes con otras enfermedades pudo sacar una conclusión correcta, por el *enorme* incremento del cáncer de escroto en los deshollinadores. “Todavía en los años veinte del presente siglo”, escribe Richard Doll (1964), “la mortalidad por cáncer escrotal de los deshollinadores era unas 200 veces la de los trabajadores no especialmente expuestos a alquitranes o hidrocarburos pesados, y en el siglo XVIII la diferencia relativa probablemente era mucho mayor”.

Un ejemplo mucho más moderno y general sobre el que yo mismo he reflexionado durante 15 años es el del tabaco. Los estudios prospectivos han mostrado que la mortalidad por cáncer de pulmón en fumadores de cigarrillos es 9 ó 10 veces la de los no fumadores; en grandes fumadores es 20 ó 30 veces mayor. Por otra parte, la mortalidad por trombosis coronaria en fumadores no excede o quizá no llegue al doble de la tasa correspondiente en no fumadores. Hay datos que van a favor de una relación causal, pero en este caso es mucho más fácil pensar en ciertas características vitales que pueden asociarse a ser fumador y que podrían ser la causa subyacente real o, al menos, un factor contribuyente importante, ya fuera falta de ejercicio, tipo de alimentación o algún otro factor. Pero explicar el gran exceso de cáncer de pulmón en función de cualesquiera otros factores ambientales exigiría alguna característica de modo de vida íntimamente asociada con el hecho y con la cuantía del hábito de fumar. Tal característica sería fácilmente detectable. En circunstancias en las que no podemos reconocerla o inferirla específicamente estamos razonablemente autorizados a rechazar las vagas impugnaciones del crítico de salón: “Podría existir ese tipo de factor, no es posible demostrar lo contrario”.

Ciertamente, en esa situación yo rechazaría el argumento a veces oído de que lo que importa es la diferencia absoluta entre las tasas de mortalidad de los distintos grupos y no la razón entre esas tasas. Eso depende de lo que queremos demostrar. Si lo que queremos saber es el exceso en número de muertes por cáncer pulmonar debido al tabaco (lo cual supone una hipótesis de causación), está claro que debemos usar la diferencia absoluta entre las tasas de mortalidad: 0,07 por 1 000 al año en médicos no fumadores; 0,57 en fumadores de 1 a 14 cigarrillos diarios; 1,39 en fumadores de 15 a 24 cigarrillos diarios, y 2,27 en fumadores de 25 o más. Pero de esto no se sigue en este caso o en problemas ocupacionales más específicos que esta medida ideal del efecto sobre la mortalidad sea también la mejor medida en relación a la etiología. A este respecto, la razón de mortalidad en fumadores en relación con no fumadores, que es de 8, 20 y 32 a 1, respectivamente, es mucho más informativa. Por supuesto, no hay que deducir que las diferencias reveladas por estas razones de mortalidad tengan ninguna importancia práctica. Puede que sí o puede que no, pero eso es otro asunto.

Conviene mencionar aquí el análisis ya clásico de John Snow respecto a la epidemia de cólera de 1854 (Snow 1855). La mortalidad que Snow registró en los clientes abastecidos con el agua extremadamente contaminada de la compañía Southwark & Vauxhall fue bastante baja, solo 71 muertes por 10 000 viviendas. Lo que destaca claramente es el hecho de que esta pequeña tasa es 14 veces mayor que la de 5 fallecimientos por 10 000 viviendas abastecidas por la compañía competidora Lambeth, cuya red no había sufrido contaminación con aguas fecales del alcantarillado.

Hacer hincapié en la intensidad de la asociación obliga también a mirar la otra cara de la moneda. No se debe descartar a la ligera una hipótesis de causalidad solo por el hecho de que la asociación observada no parezca muy intensa. En medicina esto se ve en muchas ocasiones. Son relativamente pocas las personas que portan el meningococo y que enferman de meningitis meningocócica. También pocas personas expuestas durante su trabajo a orina de ratas contraen consiguientemente la enfermedad de Weil (leptospirosis).

2) Consistencia. En la lista de características que deben tenerse en cuenta, yo colocaría a continuación la *consistencia* del fenómeno. ¿Se trata de una asociación repetidamente observada por diferentes personas, en diferentes lugares, circunstancias y momentos?

Este requisito puede ser de especial importancia en el caso de ciertos riesgos raros a los que se refieren los estatutos de nuestra Sección de Medicina Ocupacional. Hay muchas mentes alertas a los posibles riesgos de la industria actual, de manera que aparecerán muchos riesgos asociados con factores ambientales. Las pruebas habituales de significación estadística indicarán que algunas de estas asociaciones resultan difícilmente atribuibles al azar. Que la verdadera explicación sea una coincidencia casual o un riesgo verdadero, a menudo solo es posible decidirlo tras la repetición de las circunstancias y las observaciones.

En el ejemplo del tabaco anteriormente citado, el Comité Asesor del Director General de Sanidad del Servicio de Salud Pública de los Estados Unidos encontró la asociación del tabaco con el cáncer de pulmón en 29 estudios retrospectivos y 7 prospectivos (US Department of Health, Education & Welfare 1964). La lección es que, en líneas generales, se obtuvo la misma respuesta en varias situaciones y con técnicas relativamente diversas. En otras palabras, está justificado inferir que la asociación no es debida a algún error o falacia agazapada en cada uno de los estudios. Y, por supuesto, hay que precaverse contra esa posibilidad.

Tomemos el ejemplo citado por Heady (1958). Se interroga a pacientes admitidos al hospital para ser operados de úlcera péptica sobre ansiedades o crisis domésticas que puedan haber precipitado la enfermedad aguda. Los controles o testigos son pacientes admitidos para ser operados por una simple hernia, a los que se les hace un interrogatorio similar. Pero, como señala Heady, ambos grupos pueden no ser comparables. Una úlcera perforada obliga a acudir al hospital sin demora, incluso aunque haga pocos días que la esposa de uno se haya fugado con el vecino. Con una hernia, uno quizá prefiriera quedarse en casa algún tiempo hasta que pasara la pena (o la alegría). Por mucho que se repitiera el estudio, no habría manera de eliminar o poner de manifiesto este tipo de falacia.

Por lo tanto, nos encontramos aquí ante la paradójica situación en la que resultados distintos de otra investigación no valen para refutar la evidencia original. Incluso idénticos resultados derivados del mismo tipo de investigación no necesariamente apoyan la evidencia original. En mi opinión, resultados similares de estudios de distinto tipo —prospectivos y retrospectivos— tienen mucho peso.

Si miramos la otra cara de la moneda, hemos de decir otra vez que en ocasiones no hay repetición alguna a la que acudir o es imposible hacer un estudio similar, y a pesar de ello no dudamos en sacar determinadas conclusiones. La experiencia de las plantas refinadoras de níquel de Gales del Sur es un buen ejemplo. Cito de la Conferencia Conmemorativa de Alfred Watson que pronuncié en 1962 en el Instituto de Actuarios:

La población expuesta a riesgo, formada por trabajadores y pensionistas, era de aproximadamente un millar. Entre 1929 y 1938, 16 de ellos murieron de cáncer de pulmón y 11 de cáncer de senos paranasales. Lo esperado, según las tasas de

mortalidad específicas por edades que se registraban en aquel entonces en Inglaterra y Gales, hubiera sido una sola muerte por cáncer de pulmón (en vez de 16) y una fracción de fallecimiento por cáncer paranasal (en vez de 11). En las demás localizaciones orgánicas los certificados de defunción revelaron cánceres en 11 ocasiones, cuando lo esperado eran precisamente 10 u 11 casos. Se registraron 67 muertes debidas a todas las demás causas; en los 10 años del período considerado se esperaban 92. Por último, la división por ocupaciones de la población expuesta a riesgo mostró que el exceso de cáncer de pulmón y senos paranasales se había producido entre los obreros que trabajaban en procesos químicos.

Recientemente mi colega, el Dr. Richard Doll, ha profundizado en este tema. Doll vio que en los nueve años del período 1948 a 1956 había habido 48 muertes por cáncer de pulmón y 13 muertes por cáncer nasal. Su cálculo de las cantidades esperadas a las tasas de mortalidad habituales fue de 10 y 0,1 fallecimientos, respectivamente.

En 1923, mucho antes de que se hubiera reconocido ningún riesgo específico, hubo ciertos cambios en la planta refinadora. No se ha observado ningún caso de cáncer nasal en ninguno de los obreros que comenzaron a trabajar en la planta después de ese año, y en esos hombres no ha habido exceso alguno de cáncer de pulmón. En otras palabras, el exceso en ambas localizaciones fue una característica única en los hombres que trabajaron en la refinadora durante, *grosso modo*, los 23 primeros años de este siglo.

No se ha identificado el agente causal de estas neoplasias. Hasta hace poco no se había hecho ningún experimento con animales que diera alguna pista o algún indicio a favor de esta evidencia estadística. ¿Dudaría alguno de nosotros en aceptarla como prueba de un riesgo industrial grave? (Hill 1962).

En relación con lo expuesto, no conozco ninguna investigación similar. En este caso tenemos (o tuvimos) que razonar sobre una situación única. Tampoco fue tan difícil hacerlo.

3) Especificidad. Es evidente que el tercer aspecto que siempre debemos considerar es la especificidad. Si, como en el caso recién citado, la asociación se circunscribe a ciertos trabajadores y a localizaciones y tipos concretos de enfermedad, y no hay asociación entre el trabajo y otros tipos de fallecimiento, todo ello habla a favor de una relación causal.

Sin embargo, no se debe sobrevalorar la importancia de esta característica. En el ejemplo recién traído a colación hay una relación de causa-efecto con dos tipos diferentes de cáncer, pulmonar y nasal. La leche como portadora de infección y, en ese sentido, causa patógena, puede producir todo un cajón de sastre de enfermedades como escarlatina, difteria, tuberculosis, fiebre ondulante, resfriados, disentería y fiebre tifoidea. Antes del descubrimiento del factor subyacente —el origen bacteriano de la enfermedad— hubiera sido nocivo exigir con demasiada insistencia un efecto específico como requisito para poder echarle la culpa a la lechería.

Los estudios prospectivos sobre consumo de cigarrillos y cáncer de pulmón han sido criticados recientemente por no mostrar especificidad, es decir, por el hecho de que la mortalidad de fumadores es mayor que la de no fumadores debido a diversas causas de muerte (aunque, en realidad, los resultados de Doll y Hill, 1946, no muestran eso exactamente). Llegados a este punto, habría que volver a donde comenzamos, a la intensidad de la asociación. Si otras causas de muerte se elevan 10, 20 o incluso 50% en los fumadores mientras que el cáncer de pulmón se eleva entre 900 y 1 000%, tenemos especificidad, una especificidad que es de la magnitud de la asociación.

También hemos de tener en mente que las enfermedades pueden tener más de una causa. Es perfectamente posible padecer cáncer de escroto sin limpiar chimeneas o trabajar en una planta textil en Lancashire. Las correspondencias exactas no son frecuentes. Incluso, en mi opinión, la multicausalidad suele ser más probable que la uncausalidad. De todas formas, si conociéramos todo el proceso a fondo, quizá podríamos aislar un factor único.

En resumen, en presencia de especificidad de efecto, podemos sacar conclusiones inequívocas. Si la especificidad no es aparente, tampoco se justifica quedarse cruzado de brazos sin resolver nada.

4) Secuencia temporal. En mi opinión, el cuarto aspecto es la secuencia temporal de la asociación: cuál es el carro y cuáles son los bueyes. Esta cuestión puede ser especialmente importante para enfermedades que tardan mucho en desarrollarse. ¿Es la causa de la enfermedad una dieta particular o son los estadios primarios de la enfermedad los que llevan a adoptar unos hábitos dietéticos peculiares? Una profesión determinada o cierto ambiente de trabajo, ¿promueve la infección por el bacilo tuberculoso? ¿O son las personas que eligen ese tipo de trabajo más susceptibles de contraer la tuberculosis en cualquier ambiente, o incluso han contraído ya la enfermedad? Este problema temporal puede no ser muy frecuente pero requiere consideración, sobre todo cuando entran en juego factores de selección en el lugar de trabajo o en la industria.

5) Gradiente biológico. En quinto lugar, si la asociación revela un gradiente biológico (o sea, si muestra una curva dosis-respuesta) tendremos que tomarla más en serio. Por ejemplo, el hecho de que la tasa de mortalidad por cáncer de pulmón aumente linealmente con el número de cigarrillos fumados diariamente apoya en gran medida el argumento mucho más simple de que los fumadores tienen una mortalidad mayor que los no fumadores. La comparación resultaría más débil, aunque no necesariamente eliminada, si estuviera basada, digamos, en una tasa de mortalidad mucho más alta en quienes fuman poco y menor en quienes fuman mucho más. En ese caso habría que imaginar alguna relación mucho más compleja para satisfacer la hipótesis de causación. El notable efecto dosis-respuesta admite una explicación muy simple y, obviamente, aclara mucho las cosas.

Lo mismo sería válido para cualquier polvo industrial presuntamente nocivo. Cuanto más polvo en el ambiente, mayor incidencia de enfermedad será previsible. A menudo, la dificultad radica en garantizar alguna medida cuantitativa ambiental que sea satisfactoria para valorar la relación dosis-respuesta. En cualquier caso, siempre debemos buscar ese tipo de parámetro.

6) Verosimilitud. Es conveniente que la causa que suponemos sea biológicamente verosímil, pero estoy convencido de que no podemos ser muy exigentes en este terreno. Lo que es biológicamente verosímil depende de los conocimientos biológicos del momento.

Cito de nuevo de mi Conferencia Conmemorativa de Alfred Watson (Hill 1962), en la que indiqué que

“ . . . en el siglo XVIII no había conocimiento biológico alguno para apoyar (o refutar) la observación de Pott del exceso de cáncer en los deshollinadores. La ignorancia decimonónica llevó a un afamado ensayista, crítico de las “falacias” y de las asociaciones “absurdas” de la estadística, a concluir que “no podría haber nada más ridículo para el extranjero que pasara la noche en el camarote de proa de un barco de emigrantes que achacar el tifus allí contraído a los bichos que pudiera haber en los cuerpos de los enfermos”. Y en tiempos más cercanos, ya en el siglo XX, no había conocimiento biológico alguno que apoyara los datos [de nocividad teratógena] que apuntaban a la rubéola”.

En resumen, la asociación que observamos puede ser nueva para la ciencia o la medicina y no debemos descartarla a la ligera solo porque parezca rara. Como Sherlock Holmes aconsejaba al Dr. Watson, “cuando se ha descartado lo imposible, lo que resta, *aunque sea improbable*, debe ser lo cierto”.

7) Coherencia. La interpretación de las observaciones como relación causa-efecto no debe ser contradictoria con los hechos conocidos de la historia natural y la biología de la enfermedad. En expresión del Comité Asesor del Director General de Sanidad de los Estados Unidos, debe ser coherente.

Así, en la discusión del cáncer de pulmón el Comité consideró que su asociación con el consumo de cigarrillos es coherente con el incremento progresivo de las dos variables durante la pasada generación y con las diferencias de mortalidad en los dos sexos. Estas son características que podrían aplicarse muy bien a problemas ocupacionales. La conocida diferencia de mortalidad por cáncer de pulmón entre zonas rurales y urbanas no impide mantener la coherencia. Con la restricción del efecto al pulmón ocurre lo mismo.

Personalmente, considero que la histopatología del epitelio bronquial de los fumadores y el aislamiento en el humo de tabaco de factores generadores de cáncer cutáneo en animales de laboratorio son datos muy coherentes con todo lo demás. Este tipo de pruebas de laboratorio puede apoyar enormemente una hipótesis y, claro está, puede identificar el verdadero agente causal. Pero su falta no anula el valor de las observaciones epidemiológicas en seres humanos. Es indudable que el arsénico puede generar cáncer cutáneo en seres humanos, pero nunca ha sido posible mostrar tal efecto en ningún otro animal. En términos generales, las observaciones epidemiológicas de John Snow sobre la transmisión del cólera por medio del agua de la bomba de Broad Street nunca habrían sido cuestionadas si Robert Koch hubiera estado allí para aislar el vibrión colérico de los pañales de los niños, del propio pozo o del caballero de Brighton en estado delicado de salud. El hecho de que el trabajo de Koch tuviera que aguardar aún otros 30 años no debilitó los argumentos epidemiológicos, pero los hizo más difíciles de establecer frente a las críticas —justas e injustas— del momento.

8) Experimento. A veces es posible acudir a pruebas experimentales o semiexperimentales. Por ejemplo, cuando se toma alguna acción preventiva por haberse observado alguna asociación, ¿se comprueba algún efecto de las medidas tomadas? Se reduce el polvo del taller, se cambia de aceites lubricantes, las personas dejan de fumar. ¿Cambia la frecuencia de los eventos asociados? Este puede ser el hecho más demostrativo de la hipótesis de causación.

9) Analogía. En algunas circunstancias puede ser apropiado juzgar por analogía. Ahora que conocemos los efectos de la talidomida y la rubéola, seguramente estaremos dispuestos a aceptar pruebas más débiles respecto a otro medicamento u otra enfermedad vírica durante la gestación.

He aquí nueve perspectivas de lo que debemos estudiar en la asociación antes de pregonar una relación causal. No obstante, no creo que se pueda especificar, como alguien ha sugerido, un conjunto de reglas estrictas, de fácil aplicación y de obligado uso para determinar si existe una relación causa-efecto. Ninguna de las nueve perspectivas que he mencionado puede servir como prueba indiscutible de la hipótesis de causación y ninguna puede ser exigida como *conditio sine qua non*. A lo que pueden contribuir estos criterios es a clarificar la cuestión fundamental de si alguna otra manera de explicar el conjunto de hechos observados es igual o más probable que la posible relación causa-efecto.

Pruebas de significación³

Ninguna de las pruebas formales de significación puede responder a estas cuestiones. Estas pruebas pueden y deben indicarnos el efecto que podría tener el azar y la probable magnitud de dicho efecto. Pero no contribuyen ni un milímetro más a demostrar nuestra hipótesis.

Uno de los estudios de salud ocupacional que hice hace 40 años para la Junta de Investigaciones Sanitarias del Consejo de Investigaciones Médicas tuvo que ver con los trabajadores de las hilanderías de algodón de Lancashire (Hill 1930). La cuestión que tuve que responder, mediante el uso de los archivos del Seguro Nacional de Salud, fue la siguiente: los trabajadores de las salas de carda que atienden las máquinas procesadoras de algodón en bruto, ¿experimentan trastornos distintos a otros operarios de las mismas fábricas textiles relativamente no expuestos al polvo y a las fibras habituales en la sala de carda? La respuesta fue un "sí" rotundo. Entre los 30 y los 60 años de edad los trabajadores de la sala de carda sufrían más del triple de enfermedades respiratorias que los demás trabajadores, mientras que respecto a causas no res-

³ *Tests of significance* en el original. Esta expresión estadística a veces se ha traducido como "contrastes" o "docimasia de significación" (Milton & Tsokos: *Estadística para biología y ciencias de la salud*, Madrid: Interamericana; 1987, trad de A. Pérez de Vargas; Cramer: *Métodos matemáticos de estadística*, Madrid: Aguilar; 1953, trad. de E. Cansado). En nuestro idioma, el término "prueba" puede referirse a un método "con el que se pretende mostrar y hacer patente la verdad o falsedad de algo", o indicar un "indicio, señal o muestra que se da de una cosa". Estos dos significados distintos corresponden en inglés a *test* y *proof* (o *evidence*). Los términos "contraste" y "docimasia" no se han generalizado como traducción de *tests of significance*, que suele traducirse como "pruebas de significación" o, incluso, con un anglicismo antifonético, como "tests de significación". Aquí se sigue la traducción más habitual que, por desgracia, dada la ambivalencia de "prueba", genera cierta ambigüedad. (N. del t.)

piratorias su experiencia no difería de los demás. Esta diferencia marcada en cuanto a causas respiratorias salió a la luz no por períodos de enfermedad excesivamente largos, sino a partir de un absentismo excesivo y repetido de los trabajadores de la sala de carda.

Todo esto se olvidó hace ya mucho, pero hoy me interesa lo siguiente. Expuse mis resultados en 36 tablas, para varones y mujeres por separado y para media docena de grupos de edad. Había un montón de sumas. Sin embargo, en ninguna parte pensé que fuera necesaria una prueba de significación. La evidencia era tan obvia, las diferencias entre grupos eran tan importantes, el contraste entre causas respiratorias y no respiratorias era tan específico que ninguna prueba formal podía añadir absolutamente nada al razonamiento. Entonces, ¿para qué usarlas?

¿Actuaríamos hoy de la misma manera? Me permito dudar. Entre las dos guerras mundiales se hizo mucho hincapié entre clínicos e investigadores respecto a la importancia de no pasar por alto los efectos del azar sobre las observaciones. Quizá demasiado a menudo se habían deducido principios generales de una tanda de dos hombres y un perro de laboratorio; o se había concluido el tratamiento de elección a partir de una diferencia entre dos rimeros de pacientes, que lo más probable es que no tuviera significado alguno. Por lo tanto, fue un correctivo útil que los estadísticos enseñaran las pruebas de significación e hicieran hincapié en su necesidad, aunque solo fuera para que sirvieran como precaución previa a la deducción de una conclusión, en el momento de pasar de lo particular a lo general.

Me pregunto si el péndulo no ha llegado hasta el extremo contrario, no solo en lo referente a los atentos alumnos, sino también en cuanto a los estadísticos mismos. ¿No es una tontería negarse a sacar conclusiones porque no consta el error estándar? Afortunadamente, creo que no hemos ido tan lejos como nuestros amigos de los Estados Unidos, donde, según se dice, algunos editores de revistas devuelven un original porque no consta que se hayan aplicado pruebas de significación. Lo cierto es que muchas veces esas pruebas son totalmente inútiles, porque la diferencia es obvia hasta lo grotesco; o porque es insignificante o, porque, significativa o no, carece completamente de importancia práctica. Peor aún es que el brillo de la tabla de valores t distrae la atención de los problemas de recuento. Solo una fracción —además desconocida— del personal de la fábrica se ofreció voluntaria para un procedimiento o entrevista, o 20% de los pacientes sometidos a cierta terapéutica se perdieron en el seguimiento, o no se toma contacto con 30% de una muestra aleatoriamente seleccionada. Las muestras que se nos describen son a veces como la de aquel hombre que en una historia de Jonathan Swift “quería vender su casa y llevaba un ladrillo en el bolsillo para enseñarlo como muestra y animar a posibles compradores”. Si las fórmulas mágicas aparecen, el autor, el editor y los lectores se quedan tan tranquilos.

Por supuesto que exagero. Pero a menudo sospecho que malgastamos el tiempo, nos aferramos a una sombra y perdemos la sustancia, y debilitamos nuestra capacidad para interpretar los datos y tomar decisiones razonables, sea cual sea el valor P . Y, demasiadas veces, de “no hay diferencias significativas” deducimos que “no hay diferencias”. Como el fuego, la prueba de χ^2 es un siervo excelente, pero un mal amo.

La necesidad de acción

Finalmente, en la vida real, al pasar de la idea de asociación a la de causación hemos de considerar las consecuencias. No habríamos de hacerlo en aras de razones científicas. Las pruebas han de ser juzgadas por sus méritos y el juicio (en ese sentido) debe ser completamente independiente de lo que se derive de él o de a quién

pueda afectar. Pero, en un sentido más práctico, de seguro que hemos de preguntarnos qué se deriva de nuestra decisión. En medicina ocupacional nuestro objetivo suele ser tomar medidas. Si hay una causa operante y un efecto deletéreo, estaremos deseosos de intervenir para eliminar o reducir la muerte o la enfermedad.

Esa ambición es de lo más loable, pero lleva inevitablemente a usar distintos raseros al emitir un veredicto. Con pruebas relativamente escasas podemos restringir el uso de un medicamento para las náuseas matutinas de la embarazada. Si nos equivocamos y deducimos una relación causal donde solo había una asociación, no habremos hecho mucho daño. La buena señora y la industria farmacéutica seguro que sobrevivirán.

Con pruebas apropiadas podríamos tomar medidas en algo que parezca un riesgo ocupacional. Por ejemplo, en un ambiente limitado y sin demasiada injusticia si estamos equivocados, podríamos sustituir un lubricante tal vez carcinógeno por otro no carcinógeno. Pero necesitaremos pruebas de mucho peso para hacer que la gente queme en sus casas un combustible que no les guste, o que dejen de fumar los cigarrillos o de comer las grasas y dulces que les encantan. Cuando digo pruebas de peso, he de subrayar de nuevo que no se trata de poner el punto sobre todas las íes y batirse con cada crítico antes de actuar.

Toda labor científica, sean estudios observacionales o experimentales, es incompleta. El trabajo científico siempre es susceptible de alteración o modificación por los conocimientos nuevos. Eso no nos da licencia para desconocer los conocimientos que ya existen, o para posponer la acción claramente necesaria en un momento dado.

¿No puede ser, preguntaba Robert Browning, que el mundo se acabe esta noche? Sí, pero, basándonos en las pruebas disponibles, casi todos nos prepararemos para llegar al trabajo mañana a las 8:30.

Referencias

Doll R (1964) In: *Medical Surveys and Clinical Trials*. Ed. L J Witts. 2nd ed. London: p 333.

Doll R & Hill AB (1964) *Brit med. J.* i, 1399, 1460.

Heady JA (1958) *Med. World. Lond.* 89,305.

Hill AB (1930) *Sickness amongst Operatives in Lancashire Spinning Mills*. Industrial Health Research Board Report No. 59. HMSO, London. (1962) *J. Inst. Actu.* 88,178.

Snow J (1855) *On the Mode of Communication of Cholera*, 2nd ed. London (Reprinted 1936. New York).

US Department of Health, Education & Welfare (1964) *Smoking and Health*. Public Health Service Publication No. 1103. Washington.