

La improbabilidad de acertar

Alfonso Carbajo

La pandemia de Covid-19 que todavía seguimos padeciendo nos ha venido a recordar la fragilidad de nuestra condición. En un plano más mundano, ha dado a muchos comentaristas la oportunidad de exhibir su familiaridad con la jerga de los profesionales de la sanidad, usando vocablos como «sensibilidad» y «especificidad» para designar la fiabilidad de los tests diagnósticos; desgraciadamente para ellos, también nos ha dado la oportunidad a nosotros de comprobar que bajo el ornamento verbal hay más confusión que sustancia.

Así, un periodista de El País se lamentaba el dos de abril de los resultados obtenidos con los tests rápidos adquiridos poco antes por el gobierno: «los ensayos preliminares son descorazonadores» porque «no detectan positivos», ya que «tienen una sensibilidad del 30%, cuando debería ser superior al 80%».

El reportero ignora el salto de la cantidad a la calidad porque al juzgar la sensibilidad de un test, la diferencia entre uno de 30% y otro de 80%, no es reducible a 50 puntos porcentuales, por una razón análoga a la que nos impide decir que entre un cuerpo en el que el termómetro marca 37 grados y otro en el que mide 20 grados existe una mera diferencia de 17 grados: se trata del abismo entre la vida y la muerte. Un test con una sensibilidad del 80% es relativamente bueno, aunque no excelente: de cada 100 enfermos, detecta correctamente como enfermos (positivos) a 80, pero declara erróneamente sanos (negativo) a 20 (que podrán seguir contagiando). Ahora bien, un test con una sensibilidad de un 30% es una aberración porque es peor que si no existiera.

Imagínese lo absurdo de la situación. El gobierno no solo ha dejado inermes ante la pandemia a miles de sanitarios y de policías. Después de haber canalizado las compras de material sanitario a través de empresas de su confianza —unas en situación concursal, otras sin actividad relevante, ninguna especializada en el tema— y de tantos viajes a China y tantas comisiones, resulta que habríamos detectado mejor a los infectados mediante el procedimiento gratuito de tirar una moneda al aire que acierta, en promedio, la mitad de las veces.

Confiar en la intuición para valorar situaciones de riesgo o de incertidumbre conduce con mucha frecuencia al disparate, y esta torpeza de nuestro instinto se pone especialmente de manifiesto al tratar de extraer consecuencias de los resultados de un test diagnóstico.

Supongamos que ha aparecido una enfermedad cuya incidencia es de un caso por cada 1000 habitantes, asintomática en la fase inicial, pero detectable con un test diagnóstico de extraordinaria precisión. El test tiene una sensibilidad del 100% (su aplicación a todo sujeto afectado dará siempre un resultado positivo: la probabilidad de un falso negativo es cero). Por el lado de la especificidad, sin embargo, no alcanza la perfección

absoluta, quedándose en el 99%. O sea, el test registra un 1% de falsos positivos, y esa es la probabilidad de que a una persona sana la clasifique como infectada.

Paseando por la calle en la mejor disposición se encuentra usted con una unidad móvil que administra el test gratuitamente. Le aseguran que es indoloro y que no produce reacción. Decide hacer la prueba y el resultado del test es positivo, una mala noticia, ciertamente. ¿Pero cuál es la probabilidad de que usted padezca efectivamente la enfermedad? Aventure un número.

Si, teniendo en cuenta toda la evidencia disponible, ha estimado una probabilidad de enfermedad inferior al 10%, muchas felicidades porque es usted una de las raras personas capaces de evaluar correctamente la probabilidad de ocurrencia de sucesos que puedan acaecer en situaciones complejas. Pero si su estimación es superior a la correcta —digamos que ha calculado una probabilidad de 20, 50 o incluso 90%, por ejemplo—, tampoco debe apurarse porque está usted en buena compañía.

En un artículo publicado en *New England Journal of Medicine* (1978), *Interpretation by Physicians of Clinical Laboratory Results*, Ward Casscells daba cuenta de los resultados de una encuesta realizada entre el personal de la facultad de medicina de Harvard. La cuestión planteada era muy similar al supuesto descrito más arriba, si bien la precisión del test era inferior al del hipotético nuestro: la sensibilidad y la especificidad eran ambas el 95% (con la misma incidencia del 1/1000). Una muestra de 64 sujetos —de los que 20 eran médicos titulares, 20 médicos en prácticas y 24 estudiantes de la facultad— arrojó los siguientes resultados: el 45% de los encuestados estimaron un valor de 95%, siendo la media de las respuestas un 56%. Menos de la quinta parte acertó la respuesta correcta: 1,9%.

No se trata de un caso aislado. Los fallos en la valoración de riesgos son frecuentes entre los profesionales de la medicina, como puede apreciarse en el trabajo de Bailey Kuklin titulado *Probability Misestimates in Medical Care*, publicado en 2006 en *Arkansas Law Review*, que estudia una multitud de errores cometidos en la evaluación de riesgos en todo tipo de prácticas hospitalarias, desde las intervenciones quirúrgicas a las mamografías.

Por otra parte, la torpeza en la estimación probabilística no es privativa de los médicos. En el mundo anglosajón se llama falacia del fiscal a la confusión entre la probabilidad condicionada de A, dado B, y la probabilidad (condicionada) de B, dado A. Porque los fiscales, buscando una sentencia condenatoria, suelen concentrar su esfuerzo en probar que la probabilidad de la evidencia presentada, condicionada al supuesto de inocencia, es muy baja, cuando lo que deben demostrar es que la probabilidad de inocencia del acusado, dada la evidencia, es baja, cosa muy distinta.

Estoy siendo injusto con los fiscales porque errores de inferencia estadística los comenten todos, los fiscales, los acusadores privados, los abogados defensores, los peritos y hasta los jueces, y el problema viene de antiguo porque, ya en 1837, el famoso ingeniero Denis Poisson, atraído por el tema, publicó sus *Recherches sur la probabilité des jugements en matieres criminelles et matiere civile*. Pero posiblemente el episodio en el que el cálculo erróneo de probabilidades ha desempeñado un papel fundamental en el desarrollo de un proceso kafkiano con consecuencias fatales para un acusado

inocente es el conocido como el caso de Sally Clark.

La abogada Sally Clark tuvo un hijo en 1996 que murió repentinamente a las 10 semanas, víctima —se estimó— del síndrome de muerte infantil instantánea (SIDS). La criatura se había vacunado recientemente, y los padres solicitaron en el hospital una investigación de la vacuna como posible causa de muerte, examen que, por errores burocráticos, nunca se realizó.

Desconsolados, decidieron que la mejor forma de superar el trauma sería tener otro hijo que, efectivamente, vino al mundo en 1998. Desgraciadamente, a las ocho semanas, recién vacunado, Sally lo descubrió muerto en su cuna. El forense encontró indicios de violencia y cambió, inexplicablemente, el informe de la primera autopsia. Los Clark fueron detenidos, aunque el marido quedó en libertad poco después, al formalizarse cargos por doble asesinato contra Sally. Dedicado a la defensa de su consorte, abandonó su práctica jurídica de fusiones y adquisiciones de empresas, concentrándose en las fases del procedimiento penal y en el cuidado del tercer hijo de ambos, nacido en 1999, justo antes de que Sally entrara en prisión, sentenciada a dos cadenas perpetuas. Para cubrir los gastos procesales tuvo que vender la vivienda familiar.

Desde el principio, tanto en primera como en segunda instancia, el proceso estuvo plagado de irregularidades (ignorancia de testimonios relevantes, contradicciones entre los expertos de la acusación, supresión de evidencia documental esencial para la defensa) pero el destino de la acusada lo selló Sir Roy Meadow, experto estelar de la acusación, catedrático de Pediatría y autor del best-seller *The ABC of Child Abuse*. Sir Roy creía que el 40% de las muertes infantiles atribuidas al SIDS, eran, en realidad, consecuencia de malos tratos, y convenció al jurado popular de la culpabilidad de Sally con un argumento pintoresco. Dadas las circunstancias de los casos, la única alternativa al homicidio como causa de las dos muertes era la muerte de cuna, el SIDS. Ahora bien, esta causa tenía una probabilidad muy pequeña en Gran Bretaña, solo una por cada 8500 muertes infantiles. Como se trataba de dos muertes, había que multiplicar esa probabilidad por sí misma, según Sir Roy, resultando una probabilidad de 1 dividido por 73 millones, una cifra tan minúscula que hacía inaceptable la posibilidad de inocencia. Lo de los 73 millones, número sensiblemente superior a la población de Inglaterra, impresionó al jurado popular que entendió que la inocencia de la acusada era más difícil de creer que extraer la única bola roja entre 73 millones de bolas. El argumento también convenció a los jueces en primera instancia y al tribunal que entendió la apelación y confirmó la sentencia en 1999: dos cadenas perpetuas.

Sir Roy ensarta varios disparates en su declaración. En primer lugar, la muerte de cuna no es la única causa natural de muerte repentina de un bebé. Hay también factores genéticos, intoxicaciones, e incluso reacciones a vacunas, por ejemplo. Esto aumenta la probabilidad de la muerte natural respecto a la probabilidad del asesinato. El segundo error consiste en suponer que las muertes de los dos hijos son sucesos independientes —única forma de justificar la multiplicación de las probabilidades—, cuando es sabido que, en general, la probabilidad de contraer cualquier enfermedad aumenta al descubrirse que otros miembros de la familia también la padecen. Por último, incurre en la confusión conceptual fundamental consistente en invertir los términos de la probabilidad condicionada relevante. Lo que al jurado le compete determinar para

declarar la culpabilidad sin atisbo de duda (*beyond the shadow of a doubt*), es la probabilidad (alta o baja) de la inocencia del acusado, dada la evidencia disponible, mientras que Sir Roy se empeñó en demostrar la (bajísima) probabilidad de la evidencia disponible, dada la inocencia.

El padre y el marido de Sally se encargaron de difundir las aberraciones del caso, interesando a juristas, médicos y matemáticos, hasta llegar a la Royal Statistical Society que solicitó al Lord Canciller una revisión del proceso. En su desarrollo a lo largo de 2002 quedaron de manifiesto la debilidad de las pruebas de la acusación y la inconsistencia de la tesis de Sir Roy. Declarada inocente, Sally quedó en libertad en enero de 2003. Poco después el Colegio de Pediatría expulsaba de la corporación a Sir Roy Meadow. Siempre pleiteante, interpuso un recurso y consiguió ser readmitido. Menos suerte tuvo su víctima. Presa de una depresión profunda desde las muertes de sus hijos, las peripecias de los juicios, la experiencia carcelaria y la hostilidad popular desatada contra ella por la prensa amarilla que la presentaba en titulares como la «madre asesina» no hicieron más que agravar su condición, y en 2007 Sally Clark moría en su casa a los 42 años, cuatro después de ser puesta en libertad.

De este episodio se pueden extraer varias lecciones de aplicación práctica a nuestros problemas de estos días. La primera es la que expresaba meridianamente, en su slogan de ventas, un agente de seguros de Perbes: nadie está libre de un accidente. Y tanto. Piénsese que lo que un comentarista describió como «la perversión judicial más grave de la historia legal de Gran Bretaña», tuvo lugar allí, en Inglaterra, y no en Afganistán, y no en la edad media sino hace 17 años. Como para estar preparado.

La segunda lección tiene que ver con la apelación supersticiosa de algunos a «la ciencia» y a «la autoridad de los expertos». El doctor Meadow es una prueba de que el venerado hoy como experto puede ser desenmascarado mañana como charlatán. La ciencia moderna surge al negar la autoridad y desarrollar el lema cartesiano *De omnibus dubitandum*.

Por último, las disposiciones de los jueces durante los dos primeros juicios nos revelan en cierto modo su sentir íntimo acerca de la calidad científica de diferentes disciplinas. Para ellos la medicina era más respetable que la probabilidad o la estadística, como lo prueba el que permitieran al doctor Meadow pontificar sobre cuestiones estadísticas pese a que comparecía en calidad de médico. ¿Habrían mostrado la misma tolerancia con la referencia más tangencial de un experto estadístico a una cuestión biológica elemental, sin ordenarle que se limitara a opinar sobre su campo de especialidad?

Esta actitud básica de los jueces ante los métodos estadísticos es representativa del conjunto de la población, y esto explica la frecuencia con la que los profanos se equivocan al tomar decisiones en condiciones de incertidumbre. En estas cuestiones, la intuición desnuda es una guía engañosa, como demuestra Daniel Kahneman en *Thinking Fast and Slow* (2005).

No debe extrañarnos, por tanto, que con la lluvia diaria de datos que proporciona la pandemia, abundan las meteduras de pata de políticos y periodistas que, a la postre, no tienen una formación mejor ni peor que el resto de los ciudadanos. En este sentido, no se les puede exigir que dominen todos los temas. Pero sería saludable que siguieran el

consejo que un juez dio a un testigo pericial que reconoció no entender los mecanismos de una prueba: «Si no la sabía debió haber guardado silencio y, en todo caso, haber confesado a tiempo su ignorancia». ¿Pero qué políticos o qué periodistas se resignarán a estar callados?