

Pero... ¿quién encarceló a Sally?

RAÚL IBÁÑEZ TORRES

La probabilidad y la estadística son potentes herramientas matemáticas con importantes aplicaciones en nuestra vida, pero cuya mala utilización, así como una mala interpretación de su significado, puede tener graves consecuencias. Así mismo, con frecuencia son utilizadas para manipularnos y modificar nuestras opiniones.

Con el objetivo de ilustrar las anteriores afirmaciones, en este artículo se van a analizar algunos ejemplos reales, y en algunos casos dramáticos. Falsos positivos en medicina, inocentes que acaban en prisión, estudios científicos paradójicos, manipulación de informaciones periódicas o el estudio estadístico de la vacuna de la polio.

Palabras clave: Estadística, Probabilidad condicional, Falso positivo, Paradoja de Simpson, Ensayo clínico.

But... who put Sally in prison?

Probability and Statistics are powerful mathematical tools with important implications in daily life, but misuse of them has negative consequences. Moreover, usually they are used to manipulate and to modify our opinions.

In order to illustrate this, we are going to analyze some real examples. False positives in medicine, innocents who end up in prison, paradoxical scientific researches, manipulation of information in newspapers, or the medical experiment of the Salk vaccine for Polio.

Keywords: Statistics, Conditional probability, False positive, Simpson's paradox, Medical experiment.

La probabilidad y la estadística son potentes herramientas matemáticas con importantes aplicaciones en nuestra vida, pero cuya mala utilización, así como una mala interpretación de su significado, puede tener graves consecuencias. Así mismo, con frecuencia son utilizadas para manipularnos y modificar nuestras opiniones.

Con el objetivo de ilustrar las anteriores afirmaciones, en este artículo, que recoge la conferencia *Pero... ¿quién encarceló a Sally? Las matemáticas en el banquillo de los acusados*, impartida en el CI-BEM-Congreso Iberoamericano de Educación Matemática (Madrid, 2017), se van a analizar algunos ejemplos reales, y en algunos casos dramáticos. Falsos positivos en medicina, inocentes que acaban en prisión, estudios científicos paradójicos, manipulación de informaciones periódicas o el estudio estadístico de la vacuna de la polio.

Falsos positivos en medicina

Para explicar qué son los falsos positivos y la importancia de los mismos, no vamos a mostrar exactamente un caso real, sino un ejemplo clásico de la divulgación matemática extraído del exce-

lente libro *El hombre anumérico* (Tusquets, 1990) del matemático y divulgador estadounidense John Allen Paulos. Se plantea la siguiente cuestión:

Supongamos que existe una prueba médica para detectar si una persona padece una cierta enfermedad (por ejemplo, cáncer o SIDA), cuya fiabilidad es del 98 %, y que dicha enfermedad la sufren 1 de cada 200 personas. Un día nos hacemos esta prueba y nos da positiva... ¿hasta qué punto nos debemos de preocupar? ¿qué probabilidad existe de que tengamos dicha enfermedad?

El problema es ciertamente importante, puesto que de la comprensión del mismo dependen tanto aspectos relacionados con la comunicación entre el equipo médico y el paciente, el primero debe transmitir al paciente una gravedad en consonancia con la probabilidad mencionada, como aspectos médicos, es decir, la gestión médica de la situación.

En la literatura científica existen estudios en los que se intenta ver el grado de comprensión de la anterior cuestión por parte del personal médico. Por ejemplo, en el artículo *Interpretation by physicians of clinical laboratory results*, sus autores preguntaron a un grupo de 60 personas, entre estudiantes y profesores de la Escuela de Medicina de la Universidad de Harvard, una cuestión similar a la planteada y la mayoría de las personas pensaba que la respuesta era, más o menos, la

fiabilidad de la prueba, en el caso anterior, del 98 %. Algo similar comenta el psicólogo alemán Gerd Gigerenzer en su libro *Calculated risks: how to know when numbers deceive you*.

Este es un ejemplo de lo que se conoce como probabilidad condicionada. No nos estamos preguntando cuál es la probabilidad de que una persona cualquiera tenga dicha enfermedad, sino la probabilidad de que esté enferma sabiendo que ha dado positivo en el análisis (esta es la condición). A continuación, calcularemos dicha probabilidad, pero sin adentrarnos aquí en la teoría de la probabilidad condicionada, ni en el teorema de Bayes que permite calcularla, sino que realizaremos un razonamiento más intuitivo.

Revisemos la información del problema. El primer dato es el número de personas que padecen dicha enfermedad, 1 de cada 200 personas. Cada enfermedad tiene un porcentaje específico de personas que la sufren. Por ejemplo, se estima que en España aproximadamente 1 de cada 300 personas tiene la enfermedad del SIDA. Y, en Europa, se suele considerar que las «enfermedades raras» son aquellas que hay menos de 1 enfermo por cada 2000 habitantes.

El segundo dato es la fiabilidad de la prueba. Que esta sea del 98 %, es decir, que haya una probabilidad del 98 % de que «acierte» el análisis, significa que si una persona padece la enfermedad tiene una probabilidad del 98 % de que su prueba dé positiva y un 2 % de que dé negativa, a pesar de que realmente está enferma. Y, al contrario, si está sana, tiene una probabilidad del 2 % de que el test sea positivo y un 98 % de que dé negativo.

Calculemos ahora la probabilidad condicionada del problema. Imaginemos que se realizan 10 000 pruebas médicas a otras tantas personas. Dentro de ese grupo, como la enfermedad la padecen 1 de cada 200 personas, habrá unas 50 personas que padezcan la enfermedad y 9 950 sanas.

Ahora, si tenemos en cuenta la fiabilidad de la prueba, que es del 98 %, entonces:

- a) De las 50 personas que padecen la enfermedad, la prueba les dará positiva a unas 49 (que es el 98 % de 50) y dará negativa a 1 (el 2 %).



Figura 1. Fotografía de Tom Mallinson.
Fuente: Wikimedia Commons

- b) De las 9950 personas sanas, la prueba les dará positiva a unas 199 (que son el 2% de 9950), al resto les dará negativa (el 98%).

	Enferma	Sana	Total
Prueba +	49	199	248
Prueba -	1	9751	9752
Total	0	9950	10000

Tabla 1

En conclusión, dentro de ese grupo de 10000 personas, habrá alrededor de $49 + 199 = 248$ pruebas de detección de la enfermedad que han dado positivas. La mayoría de ellas, 199, serán de falsos positivos —la prueba es positiva, pero la persona no padece la enfermedad— y solo 49 serán positivos reales. Luego:

¡¡La probabilidad de tener realmente la enfermedad, habiendo dado positivo el análisis, es de $49/248$, aproximadamente una probabilidad del 20%!!

Por tanto, la probabilidad de que sea un falso positivo es del 80%.

Y eso para una enfermedad que tiene una tasa alta de afectados, un enfermo por cada doscientas personas, pero si es una enfermedad con una incidencia menor ese porcentaje baja aún más. Por ejemplo, para una enfermedad tal que hay un caso por cada mil habitantes, si el test sigue teniendo una fiabilidad del 98%, entonces habiendo dado positivo, solamente existe una probabilidad del 4% de estar enfermo. Luego un 96% de que sea un falso positivo.

Los falsos positivos no solamente afectan a los análisis médicos, sino que aparecen en diferentes situaciones de nuestra vida cotidiana, que van desde los tests de embarazo o los análisis de drogas, hasta los experimentos científicos.

El caso de Sally Clark, cuando el mal uso de la estadística te lleva a la cárcel

El siguiente es un ejemplo dramático de cómo el mal uso de la probabilidad y la estadística, en particular, en un juicio, puede tener muy graves

consecuencias, como que una persona inocente acabe en la cárcel.

En diciembre de 1996, la abogada británica Sally Clark, y su marido Steve, perdieron a su primer hijo cuando este tenía once semanas de vida. Se declaró que la muerte de su hijo se debía al «síndrome de la muerte súbita del lactante» (SMSL). El diagnóstico de muerte por SMSL se hace cuando se produce una muerte repentina e inesperada de un bebé menor de un año de edad y en la autopsia no se encuentra ninguna causa que explique la misma.

Al año siguiente, el matrimonio Clark volvió a tener un hijo. Este murió ocho semanas después de su nacimiento, en enero de 1998, y de nuevo se declaró que la muerte era debida al SMSL. En ambos casos, la madre se encontraba sola en casa y los cuerpos de los niños presentaban traumatismos, que podían explicarse como consecuencia del intento de reanimarles por parte de la madre. Un mes más tarde, Sally Clark fue detenida y acusada de asfixiar a sus dos hijos.

Durante el juicio, la acusación llamó a declarar a un profesor de pediatría de la Universidad de Leeds, sir Roy Meadow. Este declaró que «la probabilidad de que un bebé muera del SMSL es de 1 entre 8543, luego si multiplicas esta por sí misma, obtienes que la probabilidad de que dos bebés mueran del SMSL es de 1 entre 73 millones», y que una doble muerte por SMSL no se producirá más que una vez cada cien años. A lo que el fiscal añadió que «esta [la probabilidad de dos muertes por SMSL] es extremadamente pequeña, lo que hace muy improbable que los bebés murieran de forma natural e implica poderosamente el asesinato».

La fiscalía no presentó más evidencias de la culpabilidad de Sally Clark que el argumento estadístico anterior, de hecho, no había ninguna evidencia de que fuese una mala madre, que hubiese desatendido a sus hijos o que fuera violenta. Fue declarada culpable del asesinato de sus dos hijos e ingresó en prisión en noviembre de 1999.

Sin embargo, varios fueron los fallos estadísticos en el argumento del pediatra y la fiscalía.

1. Para calcular esa probabilidad de «1 entre 73 millones» de que dos bebés murieran por el síndrome de la muerte súbita del lactante, el pro-

fesor Meadow multiplicó la probabilidad de que un bebé muriese de ese síndrome, 1 entre 8 543, por sí misma. Pero eso implica que dos muertes por SMSL en una misma familia son sucesos independientes (como tirar dos veces una moneda al aire), sin embargo, pueden existir factores genéticos y medioambientales que predispongan a algunas familias a que se produzcan este tipo de muertes, por lo que un segundo caso de SMSL sería en tales familias más probable que en otras familias.

El matemático británico Ray Hill, utilizando datos del CESDI, The Confidential Enquiry for Stillbirths and Deaths in Infancy, del año 2000, llegó a la conclusión de que dada una primera muerte por SMSL en una familia, la probabilidad de una segunda muerte se incrementaba entre 10 y 22 veces.

El segundo error tenía que ver con el dato ofrecido sobre la incidencia del síndrome de la muerte súbita del lactante. El profesor de pediatría tuvo en cuenta el dato genérico, que la probabilidad de que un bebé muera de SMSL es de 1 entre 8 543. Sin embargo, los niños tienen más probabilidades de morir por SMSL que las niñas, de hecho, esa probabilidad es de 1 entre 1 300, y precisamente los Clark tuvieron dos niños.

Por lo tanto, podríamos rehacer las cuentas del profesor Meadow. La probabilidad de que el primer hijo muriera de muerte súbita era de 1 entre 1 300 (1/1 300), mientras que para el segundo, atendiendo al análisis del profesor Hill, estaría en una horquilla de 1 entre 60 y 1 entre 130. En conclusión, la probabilidad de que dos niños (bebés varones) hubiesen muerto en una misma familia de SMSL podría ser tan pequeña como $(1/1\,300) \times (1/130) = (1/169\,000)$, es decir, 1 entre 169 000 (que es el 0,0006 %).

Esa probabilidad sigue siendo muy pequeña, pero si tenemos en cuenta la gran cantidad de bebés que nacen al año en Gran Bretaña, más de 700 000 bebés, se van a producir más casos de familias con dos muertes por SMSL, como así ocurrió. Algunas de ellas terminaron en la cárcel, como Donna Anthony —encarcelada en 1998 por el asesinato de sus dos bebés, un niño y una niña— o Angela Cannings —encarcelada en 2002 por el asesinato de sus dos hijos—, en cuyos jui-

cios también intervino el pediatra sir Roy Meadow.

Más aún, el fiscal y el jurado interpretaron el valor de la probabilidad de que dos bebés de una misma familia murieran por SMSL como la probabilidad de que Sally Clark fuese inocente. Como esa probabilidad era muy pequeña, era bastante improbable que esa mujer fuese inocente. Este razonamiento es similar a considerar que una persona que ha acertado la combinación ganadora de la lotería primitiva ha hecho necesariamente trampas, puesto que la probabilidad de que toque el premio gordo es muy pequeña, del 0,000007 %.

Sin embargo, no era la probabilidad de que dos bebés de una familia murieran por el síndrome de la muerte súbita del lactante lo que importaba en el juicio, sino la probabilidad de que si dos bebés (en este caso niños) de una misma familia mueren, lo hagan por SMSL.

¿Cómo saber si esa probabilidad es grande o pequeña? Solo había dos posibles causas para explicar las muertes de los dos niños en esa familia, SMSL o asesinato. Por lo tanto, lo que necesitaba el jurado era comparar la probabilidad de que los dos niños murieran de muerte súbita con la probabilidad de que murieran asesinados. El matemático Ray Hill estimó que era 9 veces

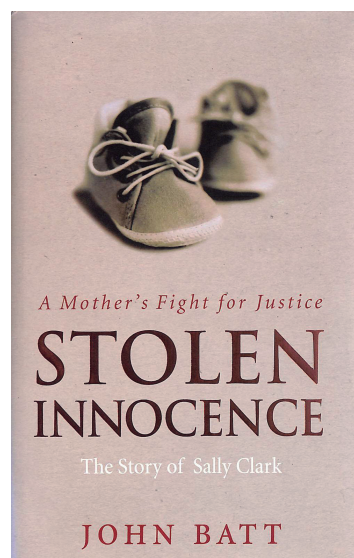


Figura 2. El abogado John Batt, miembro del equipo legal de Sally Clark, escribió el libro *Stolen Innocence* (Ebury Press, 2005), contando la historia de Sally Clark

más probable que dos niños, de una misma familia, fuesen víctimas del SMSL, que víctimas de asesinato.

Los Clark apelaron la sentencia, para lo cual contrataron a dos profesores de estadística, puesto que fue el mal uso de la misma lo que llevó a Sally a la cárcel. Sin embargo, perdieron la apelación. A pesar del revés, continuaron luchando. Entonces, descubrieron que el patólogo que había trabajado para la acusación había ocultado que su segundo hijo había sufrido una infección bacteriana en el momento de su muerte, y que podría ser la causa de la misma, por lo que habría sido una muerte natural. En consecuencia, se anuló su condena y, tras dos años y medio en prisión, Sally Clark fue puesta en libertad. Por desgracia, murió en 2007 por una intoxicación etílica.

A raíz de su caso, se revisaron los casos de otras mujeres en situaciones similares, como por ejemplo las dos citadas anteriormente, Donna Anthony y Angela Cannings, quienes fueron puestas en libertad.

La solución a un problema como este, que es la misma que para el caso de los falsos positivos, pasa por que las personas implicadas en estos juicios, abogados y jueces, tengan un mayor conocimiento de estos temas, pero también, y esto es esencial, por la contratación, siempre que sea necesario, de personas expertas en este tipo de temas, con formación matemática, y más concretamente, con formación en estadística y probabilidad.

La paradoja de Simpson

La conocida como paradoja de Simpson es otro ejemplo de la importancia que tiene el conocimiento de las herramientas matemáticas y la comprensión de la información que emana de un estudio estadístico. Esta paradoja pone de manifiesto que es posible que se produzca un comportamiento diferente en la información estadística de dos, o más grupos, cuando los datos de estos se combinan, o el revés, cuando los datos de la población de estudio se fragmentan

en grupos. Esta situación se suele presentar en estudios estadísticos sociales y médicos, por lo que una lectura errónea de los resultados puede tener consecuencias muy negativas, lo cual obliga a un análisis detallado de la situación cuando esta se produce.

Mostremos un ejemplo. Imaginemos que el mismo día se publican en dos periódicos los siguientes titulares, «El parque automovilístico de Bilbao es más nuevo que el de Madrid» en *El periódico de Bilbao* y «El parque automovilístico de Bilbao es más viejo que el de Madrid», en el *Diario de Madrid*.

Si vamos al interior de *El periódico de Bilbao* para leer la noticia correspondiente, se cuenta que se ha realizado un estudio de la antigüedad del parque automovilístico de Bilbao y Madrid, dividiendo estas ciudades en dos zonas, los barrios del centro de la ciudad y el resto de barrios. El resultado del estudio se puede ver en la tabla 2.

Es decir, en el centro de Bilbao hay más coches nuevos (67 %) que en el centro de Madrid (64 %). Mientras que hay más coches nuevos en el resto de barrios de Bilbao (45 %), que en el resto de barrios de Madrid (42 %) (tabla 3).

Sin embargo, cuando vamos a la correspondiente noticia en el *Diario de Madrid*, se muestra la siguiente tabla con la antigüedad del parque automovilístico de ambas ciudades, pero en conjunto, sin separar el estudio por barrios (tabla 4).

	Centro Bilbao	Centro Madrid
Coches nuevos	67 %	64 %
Coches viejos	33 %	36 %

Tabla 2

	Resto B. Bilbao	Resto B. Madrid
Coches nuevos	45 %	42 %
Coches viejos	55 %	58 %

Tabla 3

	Bilbao	Madrid
Coches nuevos	55 %	57 %
Coches viejos	45 %	43 %

Tabla 4

Según los datos de este periódico, el parque automovilístico de Madrid es más nuevo que el de Bilbao. Esto nos puede llevar a pensar que una de las dos informaciones, o las dos, son falsas, puesto que llevan a conclusiones contradictorias. Sin embargo, veamos que esto no es así.

Miremos de nuevo a la información del estudio, pero no en porcentajes sino en número de coches (por supuesto, esta información es ficticia ya que me he inventado los números para el ejemplo). Expresaremos las cantidades en miles de coches (tablas 5 y 6).

Sumando las cantidades, se obtienen los resultados globales (tabla 7).

Como vemos, da la impresión de que la información por partes y junta es contradictoria, a este fenómeno es a lo que se llama la «paradoja de Simpson».

La paradoja de Simpson fue mencionada a principios del siglo XX por el estadístico británico George Udny Yule (1871-1951), por lo que se le suele llamar también «efecto Yule-Simpson», pero suele tomar su nombre del también estadístico británico Edward H. Simpson (1922), quien lo describió en detalle en 1951 en relación con ciertas pruebas médicas. Simpson fue uno de los matemáticos que estuvo trabajando como criptoanalista en los barracones de Bletchley Park, tratando de descifrar los mensajes que los alemanes mandaban codificados con la máquina Enigma.

	Centro Bilbao	Centro Madrid
Coches nuevos	60 000 (67 %)	90 000 (64 %)
Coches viejos	30 000 (33 %)	50 000 (36 %)

Tabla 5

	Resto B. Bilbao	Resto B. Madrid
Coches nuevos	50 000 (45 %)	30 000 (42 %)
Coches viejos	60 000 (55 %)	40 000 (58 %)

Tabla 6

	Bilbao	Madrid
Coches nuevos	110 000 (55 %)	120 000 (57 %)
Coches viejos	90 000 (45 %)	90 000 (43 %)

Tabla 7

Veamos un caso real relacionado con pruebas médicas, publicado en el artículo «Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy» (C. R. Charig, D. R. Webb, S. R. Payne, J. E. Wickham, *Br. Med. J. — Clin. Res. Ed.*—, 1986). En el estudio se trataba de comparar el comportamiento de dos tratamientos, que llamaremos A y B, para las piedras renales y se separaban los casos leves (pequeñas piedras) de los casos más graves (piedras grandes). Los resultados se pueden ver en la tabla 8.

Vemos que cuando separamos el estudio por grupos, dependiendo de su gravedad (piedras pequeñas y piedras grandes) el tratamiento más efectivo es el tratamiento A, mientras que cuando miramos los resultados con todos los casos (leves y graves) el tratamiento más efectivo es el B.

Curiosamente la razón matemática de esta aparente paradoja (que no es tal) es muy sencilla, tiene un origen aritmético. Cuando tomamos una proporción matemática de la forma a/b (como cualquiera de los porcentajes de los que estamos comentando) y otra proporción c/d , la suma de estas $(a/b + c/d)$ no es igual a $(a+c)/(b+d)$, como bien estudiamos en la escuela, y por eso se produce la paradoja.

$$\frac{a}{b} + \frac{c}{d} \neq \frac{a+c}{b+d}$$

Pero analicemos un poco más en profundidad el motivo por el que se produce la paradoja en este caso. Miremos de nuevo al esquema de los datos y porcentajes. Si nos fijamos bien descubrimos que la paradoja de Simpson se produce por dos hechos significativos:

- a) Los tamaños de los grupos son muy diferentes. Los médicos han recomendado más el tratamiento B para los casos leves de piedras pequeñas (270 frente a 87), mien-

	Tratamiento A	Tratamiento B
Piedras pequeñas	93 % (81/87) [grupo 1]	87 % (234/270) [grupo 2]
Piedras grandes	73 % (192/263) [grupo 3]	69 % (55/80) [grupo 4]
Total	78 % (273/350)	83 % (289/350)

Tabla 8

tras que han recomendado más el tratamiento A para los casos graves de piedras grandes (263 frente a 80), por lo que el resultado total está dominado por los grupos grandes (en el esquema el grupo 2 —piedras pequeñas y tratamiento B— y el grupo 3 —piedras grandes y tratamiento A—), y menos por los menos numerosos.

- b) En el estudio anterior el éxito porcentual total que se observa está más basado en la gravedad de los casos, que en el tratamiento seguido. El porcentaje de éxitos en los pacientes graves (con piedras grandes), utilizando el tratamiento A (o el B, que es peor aún), es menor que el éxito en los casos leves, independientemente de que sean tratados con A o B.

La vacuna de Salk para la poliomielitis

Vamos a terminar este artículo con un ejemplo más positivo, el ensayo clínico realizado en 1954 para analizar la efectividad de la vacuna de Salk para la poliomielitis, y que ha sido uno de los mayores de la historia, el mayor hasta los años 80, en el cual participaron más de un millón de niños. Este estudio clínico nos sirve de ejemplo para ver qué cuestiones deben ser tenidas en cuenta a la hora de realizar una estadística médica para estudiar la efectividad de un medicamento o tratamiento, y que, en muchas ocasiones, demasiadas, no se tienen.

La poliomielitis, también conocida como polio, es una enfermedad infecciosa, producida por el virus *poliovirus*, y que afecta al sistema nervioso. No es una enfermedad común, como pueda ser el caso de la gripe, pero afecta principalmente a los niños (con edades entre 4 y 14 años) y aunque en la mayoría de los casos es asintomática, cuando afecta al sistema nervioso suele tener consecuencias muy graves, puede producir la muerte, pero también parálisis, atrofia muscular, deformidades o problemas del sistema respiratorio, que pueden llevar a terminar con un pulmón artificial para

poder vivir. Además, el comportamiento de la enfermedad, o del virus que la genera, es muy variable, algunos años su incidencia es mayor que otros, o afecta a unas zonas y a otras no, lo cual dificulta en gran medida su análisis.

El presidente norteamericano Franklin D. Roosevelt (1882-1945) había sufrido esta enfermedad cuando era un joven político, lo cual le ocasionó una parálisis. Fueron los esfuerzos de Roosevelt y su socio, el abogado Basil O'Connor, por combatir esta enfermedad los que dieron las condiciones necesarias para que fuera posible uno de los mayores estudios clínicos de la historia, el estudio estadístico de la vacuna de Salk contra la polio, con más de un millón de niños.

En los años 30, del siglo XX, se habían desarrollado algunas vacunas contra la polio, con diferentes antígenos, e incluso dos de ellas se empezaron a suministrar durante un tiempo. Sin embargo, las vacunas no resultaron efectivas, y al menos una de ellas fue la causante de algunos casos de parálisis, por lo que se dejaron de utilizar. Este fracaso pesó en el ánimo de los investigadores, que suspendieron su trabajo durante algunos años, hasta que en la década de los 40, se retomaron las investigaciones, desarrollando vacunas con virus vivos de cepas más inofensivas y también con virus neutralizados.

El virólogo Jonas Edward Salk (1914-1995) trabajó con el virus tratado con formaldehído, es decir, se eliminó su capacidad infecciosa al acabar con él, pero se mantenía su capacidad antigénica, que desencadenaba la creación de anticuerpos en el individuo vacunado. Después de un gran trabajo científico llegó la hora de evaluar su efectividad, y sus potenciales efectos adversos. Se probó primero con unos pocos niños, comprobándose que había sido efectiva con ellos. Sin embargo, esos casos no eran científicamente concluyentes, solo la hacían una buena candidata a ser considerada una vacuna contra la poliomielitis. Llegaba entonces el momento de analizar su efectividad real con un ensayo clínico a gran escala.

Antes de empezar con el ensayo clínico en sí mismo, merece la pena reflexionar un poco sobre qué tipo de ensayo se tenía que diseñar para obtener la mejor información posible. Una posibilidad podía ser realizar una vacunación de todos

los niños (en este caso, de Estados Unidos) a través de las escuelas y ver cuál era el resultado de la misma, ver si de un año a otro había habido una variación significativa en el caso de casos. Este método de actuación, que se había llevado a cabo con otras vacunas, tenía varios problemas de tipo estadístico (además de alguna consideración médica).

- a) *Variabilidad en el tiempo de la incidencia del virus.* El virus era muy irregular y la incidencia de un año a otro variaba mucho, con lo cual esa disminución en el número de casos podía no ser fruto de la vacunación, sino de la irregularidad de la enfermedad. Así, si nos fijamos en la gráfica del número de casos de polio en EE.UU. desde 1930 hasta 1956, veremos que es muy variable. De hecho, entre el año 1931 y 1932 disminuyó enormemente el número de casos, lo cual habría generado, de haberse realizado la prueba en 1931, un falso éxito. Al igual que en los años 1935, 1937, 1944, 1946, 1949, 1952, 1954 y 1955 (figura 4).

Otra alternativa podría haber sido administrar la vacuna en uno, o varios, estados y comparar los resultados con otro, y

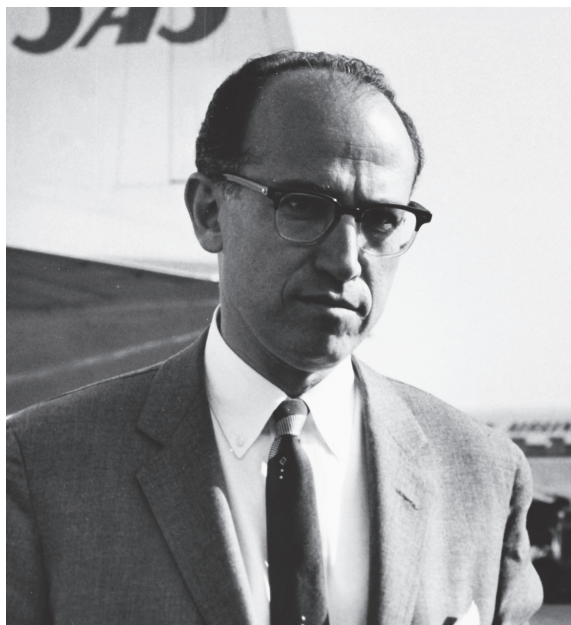
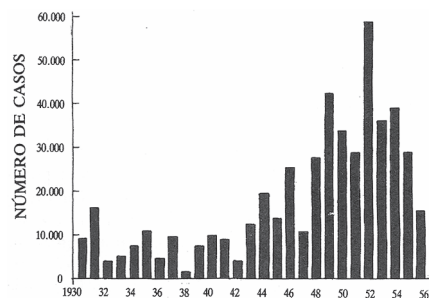


Figura 3. Jonas Edward Salk



* ESTIMADOS. CIFRAS COMPLETAS HASTA EL 8 DE DICIEMBRE.

Figura 4. Poliomielitis en Estados Unidos en los años 1930-1956

otros, estados en los que no se hubiese suministrado la vacuna.

- b) *Variabilidad geográfica de la incidencia del virus.* Por desgracia, la enfermedad también tenía una gran variabilidad espacial, podía darse en un porcentaje alto en un estado y no en el vecino, lo que hacía que esta posibilidad, de nuevo, no fuese válida. Por ejemplo, en 1956 mientras que en Chicago hubo una epidemia de polio, en Nueva York había una baja incidencia de esta enfermedad.

Por lo tanto, no era posible la vacunación de todos los niños de Estados Unidos, o solo de algunos de los estados, para obtener conclusiones firmes sobre la eficacia de la vacuna, sino que había que realizar un estudio estadístico diseñado científicamente, con un grupo de niños vacunados, otro grupo de control y otras características importantes a tener en cuenta.

Se realizó un estudio estadístico de doble ciego, es decir, ni el sujeto ni los que realizan el estudio conocen que sustancia se ha suministrado a cada individuo. Los experimentos estadísticos médicos diseñados así son más complicados, y caros, de realizar, sin embargo, desde el punto de vista científico son más correctos, más fiables.

En ocasiones nos encontramos en los medios de comunicación estadísticas que han sido realizadas con 100, 1 000 o 10 000 individuos, como mucho. Por lo tanto, una cuestión interesante era cuántas personas se necesitaban para el estudio estadístico de la vacuna de la poliomielitis, para que este fuera fiable.

Por ejemplo, nos puede parecer que una muestra de 40 000 niños podría haber sido más que suficiente para extraer conclusiones interesantes sobre la efectividad de la vacuna de Salk. Veamos si es así. La incidencia de la poliomielitis en Estados Unidos en los años 50, del siglo XX, era de 50 casos por cada 100 000 personas. En un estudio estadístico con 40 000 individuos, con 20 000 en el grupo de control y 20 000 en el grupo de vacunados, esa tasa de incidencia (recordemos, 50 casos por cada 100 000 personas) quería decir que a priori habría del orden de 10 casos de polio en el grupo de control, por lo que si hubiese 5 casos solo de polio en el grupo de vacunados tendríamos una tasa de éxito del 50 %. Pero una disminución en tan solo 5 personas puede deberse a causas aleatorias, y no a la vacuna, ya que como hemos comentado esta enfermedad era muy variable. En consecuencia, el resultado no sería para nada significativo.

Por otra parte, con 200 000 personas en el grupo de control y otras 200 000 personas en el grupo de los vacunados, la incidencia de la enfermedad en el grupo de control podría ser del orden de 100 casos (todo esto sin tener en cuenta el efecto placebo), lo que significaba que con 50 casos en el grupo de los vacunados se tendría un éxito, aparentemente, del 50 %. Con estas cifras se obtienen ya resultados convincentes. El ensayo clínico, que se realizó con más de un millón de niños involucrados (aunque no todos con un estudio doble ciego), fue todo un éxito.

En 1964 se autorizó otra vacuna contra la polio, del virólogo polaco nacionalizado estadounidense Albert Bruce Sabin (1906-1993), que se suministraba por vía oral, y que sustituyó en gran medida a la vacuna de Salk.

Desde que en 1988 la *Organización Mundial de la Salud* lanzara una campaña para erradicar

la poliomielitis, se ha conseguido reducir en un 99 % el número de casos de polio. En 1994 el continente americano fue declarado libre de polio, en 2000 la región del pacífico occidental, incluida China, en el 2002 Europa, y así se ha erradicado de muchos lugares, aunque aún hay países donde permanece, como Pakistán, Afganistán, Irak, Siria, Nigeria, Guinea Ecuatorial, Camerún y Etiopía.

Referencias bibliográficas

- CASSELLS, W., A. SCHOENBERGER y T. B. GRABOYS (1978), «Interpretation by physicians of clinical laboratory results», *N. Engl. J. Med.*, n.º 229, 18, 999-1001.
- GIGERENZER, G. (2002), *Calculated risks: how to know when numbers deceive you*, Simon and Schuster, Nueva York.
- IBÁÑEZ, R. (2015a), «Ébola, polio y estadísticas médicas», *Cuaderno de Cultura Científica* (blog), Bilbao.
- (2015b), «Falsos positivos o la importancia de comprender la información», *Cuaderno de Cultura Científica* (blog), Bilbao.
- (2015c), «La probabilidad en el banquillo de los acusados», *Cuaderno de Cultura Científica* (blog), Bilbao.
- (2016), «La probabilidad a juicio», en *El secreto de los números*, Universidad de Alicante, Alicante, 259-275.
- MEIER, P. (1992), «El mayor experimento de la historia en el campo de la sanidad pública: la gran prueba de la vacuna de Salk contra la poliomielitis (1954)», *La estadística, una guía de lo desconocido*, Alianza editorial, Madrid, 39-53.
- PAULOS, J. A. (1990), *El hombre anumérico: el analfabetismo matemático y sus consecuencias*, Tusquets, Barcelona.
- SCHNEPS, L., y C. COLMEZ (2013), *Math on Trial: How Numbers Get Used and Abused in the Courtroom*, Basic Books, Nueva York.