

Ensayos clínicos a gran escala: un sofisma con el pretexto de la ciencia*

J Penston

Consultant Physician Gastroenterologist. Scunthorpe General Hospital (Inglaterra).

Resumen

En las últimas décadas, los grandes ensayos clínicos aleatorizados se han convertido en el método de referencia de la investigación clínica, siendo la fuente de información más fiable en la cual basar decisiones terapéuticas y situándose en el lugar más destacado en la jerarquía de la evidencia. En esta conferencia se cuestiona el lugar que actualmente ocupan los ensayos clínicos realizados con elevado número de pacientes.

El resultado de cualquier gran ensayo clínico presenta pequeñas diferencias entre los grupos ensayados. Esto es suficiente para concluir que estos resultados tienen un valor discutible. Cuanto más se analizan los grandes ensayos clínicos, más cuestiones se presentan en cuanto a la validez de su metodología: las debilidades potenciales del proceso de aleatorización y de la técnica de enmascaramiento. También se ha prestado poca atención a las muchas ocasiones en las que estos ensayos no han podido satisfacer las condiciones necesarias de la inferencia causal fiable.

Todo ello tiene implicaciones importantes para los pacientes individuales. Los estudios sugieren que cuando son completamente informados de las ventajas reales, no son tan proclives a aceptar el tratamiento. Tales decisiones no encuentran el favor de la clase médica, cuyos argumentos tienden a centrarse en las ventajas supuestas del uso de los medicamentos en la población en general. Más aún, debido a los problemas de validez externa de muchos de estos ensayos, incluso estos argumentos son injustificables. En ningún caso, el tratamiento de un paciente individual se puede justificar sobre la base de las ventajas previstas en la población en general.

La consecuencia de esto ha sido un fallo general en apreciar las limitaciones de los datos aportados por los grandes ensayos clínicos.

Igual de preocupante es la ausencia de cualquier medio para confirmar o refutar los resultados de los grandes ensayos clínicos. Frecuentemente en los grandes ensayos clínicos se da un interés adquirido por un resultado determinado. Las compañías farmacéuticas que están detrás de los ensayos pueden sesgar intencionadamente la investigación, y este sesgo puede pasar desapercibido.

La conclusión es que los grandes ensayos clínicos no son el material para los principales avances en medicina. Es probable que sus resultados sometan a tratamientos prolongados a pacientes que no obtienen beneficios de los mismos y que además consumen recursos de los sistemas de salud para beneficio de la industria farmacéutica.

Palabras clave: Ensayo clínico aleatorizado. Investigación médica. Paciente

Summary

In recent decades, large randomized clinical trials have become the gold standard for clinical research, serving as the source of the most reliable information on which therapeutic decisions are based and taking the most prominent position in the evidence hierarchy. This conference questions the place occupied by clinical trials involving large numbers of patients.

The outcome of any large clinical trial demonstrates small differences among the groups studied. This circumstance is enough to lead to the conclusion that the results are of debatable value. The more large clinical trials are analyzed, the greater the number of questions that arise with regard to the validity of the methodology: the potential weaknesses of the randomization process and of the masking technique. In addition, little attention has been paid to the many occasions on which these trials have been unable to satisfy the necessary conditions of reliable causal inference.

All these aspects have important implications for individual patients. The studies suggest that when they are fully informed of the real advantages, they are not as inclined to accept the treatment. Such decisions are not approved of by medical professionals, whose arguments tend to focus on the supposed advantages of the use of medication among the general population. Moreover, owing to the problems of external validity of many of these trials, even these arguments are unjustifiable. In no case can the treatment of an individual patient be justified on the basis of the foreseeable advantages for the general population.

The consequence of this has been a general failure to perceive the limitations of the data provided by large clinical trials.

Equally disturbing is the absence of any means of confirming or refuting the results of large clinical trials. In these studies, there is frequently an acquired interest in a given outcome. The drug companies that are behind the trials can intentionally introduce a bias in the investigation, and this bias can go undetected.

The conclusion is that large clinical trials are not the substance of which the major advances in medicine are made. Their results will probably subject patients to prolonged treatments from which they obtain no benefits and that, moreover, consume resources pertaining to the health care systems for the benefit of the pharmaceutical industry.

Keywords: Randomized clinical trial. Medical research. Patient.

*Conferencia del IX Congreso de la SEFAP. Pamplona, 22 de octubre de 2004 (traducción).

Voy a hablarles de los ensayos clínicos aleatorizados (ECA) a gran escala, la visión de un hereje. Todos saben que la herejía es aquello que hace alguien que cuestiona la visión ortodoxa de la Iglesia o alguien más particularmente que cuestiona sus enseñanzas. Yo me considero culpable de ambos pecados.

En primer lugar, quisiera decirles que los ECA a gran escala son la ortodoxia actual dentro de la investigación médica. Se consideran una especie de religión entre algunos grupos de profesionales. Todos estarán cansados ya de oír hablar del «*standard gold*». Se han publicado más de medio millón de ECA a gran escala en los últimos 50 años y tengo la sensación de que casi todos han incluido este término.

Nosotros consideramos que estos ECA son el método de referencia en la investigación médica. En particular los ECA a gran escala, los megaensayos, se considera que aportan las pruebas más fiables y de mejor calidad sobre las que podemos basar nuestros tratamientos.

Cuando abrimos el buzón nos encontramos con directrices clínicas que parecen entregárenos cada semana y cuando las leemos, vemos en la parte superior de la tabla de evidencia a los ECA a gran escala junto a metanálisis, muy por debajo aparecen los ECA de menor tamaño, todavía más abajo los estudios observacionales como los de cohortes y los casos y controles, luego los no controlados, serie de casos e informes de casos.

Creo que esta jerarquía que vemos una y otra vez en las directrices clínicas nos demuestra hasta qué punto los ECA son la visión ortodoxa de la comunidad investigadora. Pero también se siente una gran reverencia ante estos estudios, por ejemplo, en el quincuagésimo aniversario del primer ECA, el *British Medical Journal* dedicó todo un número a celebrar este desarrollo, y un par de años después, los editores del *Lancet* y también del *British Medical Journal* escribieron conjuntamente sobre el tema. En esta revista se puede leer que Gran Bretaña le ha dado al mundo Shakespeare, la física de Newton, la teoría de la evolución, la democracia parlamentaria y el ECA. O estos editores no piensan demasiado bien acerca de Shakespeare, Newton y Darwin o por el contrario tienen una visión exageradamente buena de lo importante que resultan los ECA. En realidad debo decir que muchos de estos editores y otros son quienes han creado esta nueva religión. Tristemente, hay mucha gente en la investigación médica que sigue, sin más, los pasos marcados por esos editores.

Quien critique los ECA y su metodología o la metodología de los ECA a gran escala, será considerado un hereje. En Gran Bretaña, nos gusta pensar que el primer ensayo clínico (EC) fue el que realizó en 1753 James Lean, capitán de la marina real, quien publicó un estudio sobre el escorbuto. Trató a los marineros con escorbuto con una variedad de tratamientos, uno de los cuales eran cítricos y obviamente descubrió que éstos, que contenían la vitamina C, curaban el escor-

buto, y esto es lo que nosotros pensamos que es el punto de partida de los EC. Si queremos ser sinceros, diremos que quien realmente lo hizo fue Johannes Fibrihsked, médico danés que en 1898 publicó la metodología básica de los EC. Lo que hizo fue asignar a pacientes de forma alternativa diferentes tratamientos o ningún tratamiento, intentando crear dos grupos iguales; pero todos sabrán que el primer ECA fue el del estudio surgido en la era moderna acerca del tratamiento con estreptomycin para la tuberculosis pulmonar, publicado en octubre de 1948. Uno de los principales «*arquitectos*» de aquel primer ensayo fue Agustín Bartfortail, quien escribió esta cita: «*el objetivo de la aleatorización consistía en garantizar de antemano que siempre que fuera posible los grupos de control y el tratado fueran iguales en todos los aspectos relevantes*». Cuando hablaba de los aspectos relevantes, hacía referencia al hecho de que podían influir en el resultado.

Inicialmente, la aleatorización parece ser un proceso bastante sencillo, no obstante hay muchas fases en las que se pueden producir errores. En primer lugar, debemos garantizar que realmente nos aporte grupos equilibrados. Nos debemos plantear preguntas como, por ejemplo: ¿Este proceso de aleatorización es realmente aleatorizado? ¿Realmente la ocultación de la aleatorización ha sido adecuada, o vamos a encontrarnos con un desequilibrio en la selección? También nos deberíamos preguntar si este proceso ha producido un desequilibrio basal simplemente como resultado del azar. Después debemos cuestionarnos si la aleatorización se ha mantenido a lo largo de todo el estudio o no. Esto habitualmente se hace con la técnica del doble ciego, que debo decirles que a menudo no tiene éxito y nos lleva a que los diferentes grupos se gestionen de manera distinta. Por último, debemos evitar que se produzca una interrupción en la aleatorización durante la fase analítica. Debemos garantizar que todos los pacientes aleatorizados al principio del estudio formen parte del análisis por intención de tratar. Todos estos aspectos tienen que ver con lo que significa la validez interna (VI) de un EC. Una vez que completamos el ensayo y analizamos los datos, identificamos una diferencia en los resultados. Tras lo cual, inferimos que esta diferencia en el resultado se debe al tratamiento, es decir, el tratamiento es la causa de la diferencia obtenida o alcanzada en el resultado. Pero esto solamente será cierto si hemos sido capaces de excluir otras posibles explicaciones. En primer lugar, debemos garantizar que no haya ninguna distribución desigual de las variables relevantes. Como ya he dicho con anterioridad, esto se debe hacer utilizando un proceso de aleatorización adecuado manteniéndolo a lo largo de todo el estudio. Posteriormente, debemos excluir el sesgo como explicación, esto habitualmente se hace con la técnica del doble ciego. Por último, también debemos excluir el azar mismo como explicación de la diferencia en el tratamiento, y esto hay que hacerlo a través de un análisis estadístico.

Decimos que si todo esto se ha tenido en cuenta, entonces el estudio tendrá VI. Ésa es la teoría, pero en la práctica es muy diferente. Se han llevado a cabo muchos estudios que han analizado la calidad de los ECA a lo largo de los últimos 20 años que han demostrado una y otra vez que tienen un nivel de calidad bastante bajo que no parece estar mejorando con el tiempo.

Quisiera mencionar uno de los estudios, el más reciente que he encontrado se publicó en *JAMA* (Huwiler-Muntener K, et al. *JAMA* 2002;287;2801-4), donde los autores estudiaban muchos ECA y descubrieron que: la asignación equilibrada de los pacientes solamente se conseguía en una cuarta parte de éstos, el ciego no era adecuado en un 20% de los pacientes y solamente la mitad de los estudios incluían un análisis por intención de tratar. Por tanto, es como para estar preocupados y los autores que tenemos esta preocupación proponemos sin embargo el ECA, a pesar de lo desanimados que estamos con lo que está ocurriendo. En mi opinión, la clave es que las condiciones de la VI en los ECA con frecuencia no se cumplen. Eso es lo que debería realmente ser el núcleo del tema. Hemos hablado de la VI, pero ¿qué pasa con la validez externa (VE)? Se cree que los resultados de estos ECA se podrían aplicar o generalizar a la población más amplia de pacientes que presenten la misma enfermedad, pero esto solamente será así si la población del estudio, es decir los pacientes reclutados para el ensayo, representan a la población general de pacientes con esa enfermedad.

Pero el problema que tenemos es que la población del estudio no se selecciona de forma aleatoria. Todos sabrán que hay una larga lista de criterios de inclusión y exclusión en cada ensayo y éstos, sin duda, siempre llevan a diferenciarse entre los pacientes reclutados en el ensayo y los excluidos del mismo. Un ejemplo de tres ECA a gran escala (tabla I), entre la mitad o dos terceras partes de los pacientes que podrían haber formado parte de los estudios fueron excluidos; esto obviamente tiene una serie de implicaciones importantes. En los mismos, se concluyó con mucha claridad que la mortalidad de los pacientes incluidos era diferente de la de los que eran excluidos.

Pero me gustaría centrarme en un estudio de dos hospitales generales de un distrito de Inglaterra en el que revisaron ensayos publicados anteriormente. Descubrieron que el 70% de los pacientes elegibles habían sido excluidos de aquellos estudios. Después, estudiaron con más detenimiento el resultado alcanzado por los ensayos en aquellos pacientes, y observaron que el índice de mortalidad en los pacientes incluidos en los ensayos era de un 8%, mientras que el índice de mortalidad en pacientes excluidos de los estudios alcanzaba un índice del 18%. Este análisis se limita sólo a los pacientes a quienes se les aplicaron tratamientos para la trombólisis, por lo tanto nos demuestra con claridad que los criterios de selección de ECA influyen en el resultado.

Tabla I.

ECA A GRAN ESCALA: VALIDEZ EXTERNA

Estudio	Número de pacientes	
	Pacientes excluidos	Pacientes escogidos
ASSET	13.318	62%
GISSI-2	38.086	67%
GISSI-3	43.047	55%

Hay otro aspecto de la VE que no debemos olvidar, es el efecto producido por la participación en el propio ensayo. Con esto quiero decir que existe un nivel de cuidado diferente entre el que se da a los pacientes de los ECA y el de la práctica clínica habitual. Habitualmente el nivel de asistencia médica es superior en los pacientes incluidos en los ECA, y esto puede afectar a la VE del estudio.

Como conclusión podríamos decir que las diferencias entre los pacientes del estudio y la población general, junto con el efecto provocado por la participación en el estudio, nos plantean que siempre debemos cuestionarnos la VE.

Cuando comenzamos con los ECA a gran escala nos encontramos con una situación muy prometedora, el cielo es azul y hay muy pocas nubes, pero cuando empezamos a estudiarlos con detenimiento vemos que surgen los primeros problemas. Hay que plantearse una serie de preguntas muy serias con respecto a la VI y VE, pero esto es solamente el punto de partida.

¿Por qué elegimos hacer ECA a gran escala? La respuesta es porque esperamos que haya una pequeña diferencia en el tratamiento. Si va a haber una pequeña diferencia en el tratamiento, se supone que vamos a necesitar un gran número de pacientes para poder conseguir una significación estadística. Esto es algo muy claro. Y ¿por qué esperamos que haya una diferencia en el tratamiento tan pequeña? Esto se debe al hecho de que sólo algunos de ellos desarrollarán el evento sin el tratamiento y dado que algunos de esos pacientes se beneficiarán del tratamiento, la siguiente pregunta que nos debemos plantear es: ¿por qué solamente serán algunos de ellos y no la mayoría o todos los pacientes? La respuesta es porque tenemos un conocimiento limitado con respecto a cuáles son los pacientes que van a desarrollar el resultado buscado sin el tratamiento y porque también desconocemos las relaciones que existen entre la enfermedad y el mecanismo del tratamiento. Sabemos muy poco, y esto queda reflejado en muchos ensayos por el número de pacientes que tenemos que reclutar. Si estudiamos cualquier ECA a gran escala nos damos cuenta de que sabemos muy poco sobre la población a estudio. Conocemos la demografía, pero no las variables relevantes que tienen que ver con el evento y tampoco cuáles son los pacientes que van a desarrollar el resultado y cuál

Tabla II.

Mayoría superflua		
ASSET	Lancet 1990;335;1175-8	90%
LATE	Lancet 1993;342;759-66	90%
GISSI-3	Lancet 1994;343;1115-22	83%
LIPID	NEJM 1998;339;1349-57	92%
CARE	Circulation 1999;99;216-23	96%
HOPE	NEJM 2000;342;145-53	82%
PROGRESS	Lancet 2001;358;1033-41	86%
PROSPER	Lancet 2002;360;1623-30	84%

les no. Al progresar el ensayo, analizamos el resultado y entonces empezamos a reconocer el grupo placebo y el grupo activo donde podemos identificar las cifras de pacientes que han desarrollado o alcanzado el evento. Esto nos permite llegar a ciertas deducciones con respecto a la población original del estudio. Podemos afirmar que la población general está formada por dos grupos, los que van a desarrollar el evento y los que no lo hacen, a estos últimos les denominamos mayoría superflua.

Si estudiamos los diferentes ECA de esta serie, (tabla II) más o menos entre un 85% y un 95% de los pacientes reclutados para estos estudios constituían la mayoría superflua.

Sé que es una realidad muy extraña la que nos encontramos cuando la mayoría de los pacientes reclutados para un ensayo clínico no aportan ningún dato del fármaco porque no van a desarrollar el evento. Es casi como esto: imaginen que diseñamos un experimento para probar una nueva técnica, «la preservación de las manzanas en situación de almacenamiento o en tránsito». Seríamos muy sabios si evitáramos la situación en la que el 90% de las manzanas estudiadas fueran imitaciones de cera. Ésa es exactamente la situación en la que nos encontramos con los ensayos a gran escala. Y ¿por qué nos encontramos con este problema? En parte porque a lo largo de las últimas décadas nos hemos sentido totalmente hipnotizados por las cifras, los números de gran tamaño. Cuando oímos que hay un ensayo en el que se implica a miles o decenas de miles de pacientes, casi de forma automática pensamos que el ensayo debe ser de buena calidad, que debe ser importante, que debe ser relevante. Las empresas farmacéuticas lo saben, lo saben con toda claridad. Escuchen a sus visitantes médicos cuando se enorgullecen de sus últimos ensayos, 50.000 pacientes, más de 350 países. El mensaje está claro, ese gran tamaño del ensayo es una virtud, eso es lo que nos están diciendo, pero en realidad lo que ocurre es justamente lo contrario, cuanto más grande es el ensayo, más debilidades incluye. También es un síntoma de una falta subyacente de conocimiento con respecto al tema que se está estudiando. El gran tamaño de los ensayos o estudios también refleja el pequeño tamaño que se consigue en el efecto del tratamiento; estudiémoslo con más detenimiento. Hay muchos ensayos de gran escala en los que la reducción absoluta del

Tabla III.

REDUCCIÓN ENGAÑOSA DE RIESGO RELATIVO

Estudio	Mortalidad	
	Reducción relativa (%)	Reducción absoluta (%)
ASSET	26	2,6
ISIS-2	25	2,8
GUSTO	14	1,0
GISSI-3	11	0,8
HOPE	16	1,8
PROSPER	15	2,1

riesgo se sitúa entre 1-3%. Todos los ECA a gran escala tienen una diferencia de tratamiento absoluto muy pequeña, esto es un axioma, es cierto al 100%. Ése es el motivo por el que se hacen ensayos a gran escala. Pero ésta no es la sensación que uno tiene cuando lee las revistas médicas o ve la televisión, siempre se habla de un 25% de reducción de los ataques al corazón, un 10% de infartos de miocardio que se evitan, son cifras muy grandes y esto se debe a que constantemente se intenta disfrazar esa pequeña diferencia en el tratamiento.

En primer lugar, tenemos que centrarnos en lo que es la reducción del riesgo relativo (RR), esto ocurre en todos los ensayos y especialmente en los resúmenes de los artículos. La mayor parte de los médicos sólo leen los resúmenes, los resultados no deseados por los investigadores se pueden ocultar realmente en el texto sabiendo que hay muy pocas personas que los van a leer y, finalmente, también está la manipulación de los datos en las conclusiones.

La tabla III demuestra hasta qué punto la reducción del riesgo relativo puede ser algo que nos equivoque, que nos confunda. Tenemos aquí la reducción relativa, la reducción absoluta y el índice de mortalidad. En todos los casos la relativa tiene un porcentaje muy superior, habitualmente multiplicado por un factor de 10 o más. Yo creo que la reducción relativa del riesgo es una estadística que realmente se introduce sin más para satisfacer a los investigadores y a los patrocinadores, inflando las diferencias de tratamiento. Hay otra forma de evitar hablar acerca de estas pequeñas diferencias absolutas en el tratamiento, y es centrándonos en los resultados. Constantemente nos centramos en los resultados de la población en lugar de en los pacientes individuales. Esto nos permite acentuar las grandes ventajas, es decir, esos 10.000 muertos que nos hemos ahorrado etc., etc., y aparentemente puede justificar la prescripción a gran escala de estos fármacos.

Pero este argumento tiene fallos, en primer lugar porque la VE es siempre cuestionable y en segundo lugar porque, en cualquier caso, no se puede utilizar para justificar los tratamientos de un paciente individual. Los efectos sobre la población no pueden ser utilizados para hacerle tomar a un paciente un trata-

miento específico. ¿Qué significan estos resultados para el paciente individual? Recordemos el tamaño del que estamos hablando, diferencias absolutas del 1, 2 quizás 3%, lo que significa que el porcentaje de pacientes que no van a obtener ningún beneficio en absoluto de estos fármacos se encuentra entre el 96 y el 99%. ¿Qué piensan los pacientes acerca de estas diferencias? No es una pregunta fácil de responder, porque no hay demasiados estudios al respecto. Sin embargo, los que se han llevado a cabo sugieren que si se informa totalmente a los pacientes, si se les da tiempo para considerar la información sobre lo que se les administra y en particular, si se les habla de las diferencias absolutas del tratamiento, los pacientes no se mostrarán especialmente dispuestos a aceptar esos tratamientos. El mejor estudio sobre este tema procede de Darlington, una ciudad en el noreste de Inglaterra (*Trewby PN, et al. Clin Med 2002;2;527-33*). En este estudio se vieron implicados 307 pacientes, dos terceras partes habían estado ingresados en la unidad coronaria o habían recibido algún tipo de fármaco cardiovascular. El otro tercio no tenía ningún tipo de enfermedad cardíaca y no estaba recibiendo ningún tipo de fármaco cardiovascular. Este estudio consistía en dar a cada uno de estos individuos mucha información acerca de una nueva medicación hipotética para las enfermedades cardíacas. Se les informaba a través de textos y gráficos y se les daba tiempo para estudiar los datos antes de que se les entrevistara. Los resultados de estos estudios fueron realmente sorprendentes: la reducción absoluta del riesgo más baja que aceptarían los pacientes para aceptar un tratamiento a largo plazo era de entre el 20 y el 30%, en otras palabras, decían que querían un beneficio multiplicado por 10, porque si no no aceptarían formar parte del ensayo a gran escala. El 75% de los pacientes decían que no aceptarían el fármaco si el beneficio absoluto fuera inferior al 5% en un periodo de tiempo de 5 años.

Parece ser que existe un conflicto entre lo que los pacientes piensan y lo que se establece en las directrices actuales, y creo que este conflicto surge porque los autores de las directrices no comprenden el significado de estas pequeñas diferencias para los pacientes individuales.

Veamos si somos capaces de iluminar un poquito más ese significado, de aclararlo un poco más. Me gustaría que pensarán en este ensayo hipotético, 20.000 pacientes, aleatorizados, para fármaco o placebo, siendo el punto final la mortalidad a lo largo de 5 años. El fármaco reduce la mortalidad del 8 al 7%. Y ahora nos planteamos: ¿cuál es el significado de esa reducción del 1% en la tasa de mortalidad? Para intentar responder a esta pregunta imaginemos a un paciente a quien se le dice el diagnóstico y entonces le dice el médico: la mortalidad a los 5 años con esta enfermedad es del 7%. ¿Qué significa eso?, se lo podemos explicar al paciente de otra forma, 7 de cada 100, o 1 de cada 14, pero lo que realmente nos preocupa es el significado en término de cuán relevante o significati-

vo resulta esto para la vida del paciente, y ¿cómo podemos evaluar ese significado? Podemos hacerlo a través de la respuesta del paciente: ¿habría habido diferencia si se le hubiera dicho a usted que la mortalidad era del 8% en vez del 7%? Píenselo durante un minuto, yo creo que es inconcebible que nos hubieran respondido de manera diferente.

Lo que quiero decir con esto es que si aceptamos esta idea y no hay diferencia entre el significado del 7 y el 8% para los pacientes individuales, entonces nos preguntamos: ¿Cuál es el significado de esa reducción del 8 al 7%? Si se les plantearan preguntas sobre esto ustedes alzarían la mano y dirían: sí pero lo que realmente importa es la diferencia, la reducción en el resultado.

Otro ensayo hipotético, en este caso con cientos de miles de pacientes, en el que el fármaco reduce la mortalidad del 0,8 al 0,7%; hay una diferencia en el tratamiento del 0,1% y ¿eso es importante o justifica el tratamiento?

Dudo que haya muchas personas aquí presentes hoy que dijeran que esto es realmente importante o que realmente justifique el tratamiento. Por lo tanto, resulta razonable argumentar que este significado en términos de la importancia para el paciente no se puede basar únicamente en que sea una reducción del resultado.

El tamaño importa, pero entonces esto nos vuelve a plantear lo que es el significado de esa pequeña diferencia. Yo creo que esas diferencias tan pequeñas en términos de los pacientes individuales no tienen significado, no importan. Se nos están juntando cada vez más nubes y esas nubes son cada vez más oscuras, más grises con respecto a los ensayos a gran escala. ¿Cómo hemos llegado a esta situación tan terrible? Yo creo que se debe a que a lo largo de los últimos 20 años, o incluso más, la investigación médica se ha alejado de las ciencias y se ha metido cada vez más en el terreno de la estadística.

Volvamos a lo más básico, Francis Bacon, que nació en la Inglaterra isabelina, era abogado, político que llegó al Parlamento y principalmente fue filósofo, filósofo de las ciencias. Además de sus maravillosas obras también escribió muchos ensayos que son famosos por sus citas, y quiero darles una de esas citas: «*Las mujeres son las amantes de los hombres cuando éstos son jóvenes, las compañeras en la edad intermedia y las enfermeras cuando son mayores*». Esto no es lo que llamaríamos políticamente correcto en ningún sentido, pero en este caso Francis Bacon ha sido reconocido como el padre del método científico.

La ciencia no comienza en un vacío, sino con un conocimiento básico y con una teoría, lo que nos permite contar con definiciones estrictas de los objetos que se ven implicados y también nos permite utilizar mediciones precisas. Y ese conocimiento básico, esa teoría básica, también nos permite identificar la mayor parte, si no todas, las variables relevantes con antelación. El resultado de todo ello es que podemos formar

clases homogéneas, grupos de objetos e individuos que se parecen entre sí mucho en términos del resultado. Una vez llegamos a este grupo o clase homogénea podemos llevar a cabo un experimento, tal y como aconsejaba Francis Bacon.

El resultado de las ciencias es que desarrollamos lo que llamamos una generalización universal. Las generalizaciones universales se aplican a todos los ejemplos de un fenómeno, y tienen una serie de aplicaciones prácticas muy interesantes.

Las ciencias nos permiten construir estructuras enormes como ésta, el puente Hanbridge que tiene 3 millas de longitud y resulta visible desde la punta del hospital donde yo trabajo. También nos permiten asistir a conferencias en Pamplona y a la vez conectar la televisión y ver a Michael Owen jugando y metiendo un gol para el Real Madrid. Las ciencias nos permiten también viajar de un lugar del mundo a otro en cuestión de unas pocas horas.

Piensen en cuantas generalizaciones tienen que ser del todo fiables antes de que podamos incluso empezar a construir un avión, por no decir lo que eliminamos o lo que tenemos en cuenta para poder despegar y aterrizar con seguridad. Esto es ciencia en su sentido más puro, de esto es de lo que les estoy hablando.

Pero, ¿cómo es que las ciencias nos pueden ofrecer unas generalizaciones tan fiables, unas raíces tan fiables? Para responder a esto debemos remontarnos en la historia de nuevo, no tan lejos como a la Inglaterra isabelina sino a la Ilustración, al año 1700 de David Hume, el más grande de los filósofos británicos. No voy a dedicarle demasiado tiempo a la filosofía de Hume, pero sí quiero mencionar un aspecto importante: la teoría de la causalidad. Lo que nos dijo Hume fue que hacemos observaciones, vemos ejemplos de A seguida de B y una vez realizamos varias observaciones llegamos a deducir una generalización. A causa B, todas las A causan B, esto puede ser una observación simple pero también se aplica a la formación de lo que llamamos experimentos repetidos, no marca realmente ninguna diferencia, pero el punto clave y crucial es el siguiente: Hume dijo que la generalización tiene una precisión que dependerá de dos rasgos importantes: la similitud o parecido, y la regularidad. Cuando hablaba de ese parecido o similitud quería decir que en todos los casos de A nos vamos a encontrar con parecidos en todos los aspectos importantes o relevantes con el resultado. Y cuando hablaba de regularidad se refería a que todas las observaciones mostrarán ejemplos de A con B y ningún ejemplo de A sin B, es decir, nos encontraremos con una asociación regular entre A y B.

No piensen que esto es observar, sin más no, no, esto se puede aplicar también a los experimentos. En esta cita que no hace falta detallar él enfatizaba lo siguiente: «*Si la regularidad, o parecido no existen, cualquier inferencia o deducción causal que establezcamos a partir de los datos se verá en gran medida debilitada*». El parecido y la regularidad serán los cimientos para tener una deducción o inferencia cau-

sal fiable, que a su vez será responsable de que seamos capaces de producir generalizaciones universales. Y es por ello que tenemos aplicaciones prácticas en ejemplos individuales.

¿Y cómo se relaciona esto con estos ECA a gran escala? Como ya he dicho con anterioridad, éstos comienzan a partir de un conocimiento base limitado. Lo único que podemos hacer es reducir la heterogeneidad de la población.

Nosotros creemos que la respuesta a todo esto va a ser la aleatorización, y lo único que ella hace es que la ignorancia se divida en proporciones iguales. Entonces llamamos al estadístico y le decimos: identifica las pequeñas diferencias y dínos si esas diferencias resultan estadísticamente importantes. El resultado será una generalización vaga y en la que no podamos tener fiabilidad. No hay ningún tipo de parecido y no hay ningún tipo de regularidad, no puede haber parecido porque admitimos que partimos de una población mixta. No hay regularidad porque no estudiamos los ejemplos individuales, todos los datos se analizan de forma colectiva, no hay ningún eslabón claro entre la causa y el efecto. Y lo que quiero transmitirles en general, por lo tanto, es lo siguiente: Las ciencias y los ensayos a gran escala son maneras diferentes de investigar o estudiar fenómenos dentro del mundo natural, cuentan con metodologías diferentes y con conceptos distintos de causalidad. Uno de ellos produce generalizaciones fiables con aplicaciones prácticas muy amplias y el otro no lo consigue. Y si todo lo que acabo de contarles es cierto ¿cómo sabemos si los resultados de cualquier ECA a gran escala son válidos? Si nos planteamos esta pregunta con respecto a las ciencias, la respuesta es muy obvia, decimos: estudia un ejemplo individual, te podemos demostrar cómo funciona este fenómeno y si hay dudas al final de ese proceso, siempre podemos decir: repite el experimento original.

La replicación, la repetición es uno de los rasgos cardinales de las ciencias y es posible debido a la precisión de las definiciones, y de las mediciones. Debido a ello podemos recrear esas condiciones originales, volver a evaluarlas y confirmar o rechazar el hallazgo del final.

Pero, ¿qué pasa con los ECA a gran escala? Los ejemplos individuales son claramente relevantes: ¿Cómo podrían llegar a tener ningún tipo de relevancia con respecto a los resultados de los estudios? La experiencia clínica sin embargo nos demuestra que no tenemos suficientes pacientes para poder confirmarla y en cualquier caso tampoco se controlaría. ¿Y qué pasa con respecto a la repetición o replicación? Aquí tenemos un problema y es que no podemos recrear las condiciones originales.

Tomen cualquier ensayo a gran escala y estudien la definición de la población del estudio y verán que no se puede recrear. Se podrían preguntar entonces lo siguiente: ¿Seguro que hay algún EC a gran escala que pueda confirmar a otro? Yo creo que no. Tomemos un ejemplo: el estudio A es un ensayo a gran escala que

produce resultados positivos, el estudio B produce resultados negativos.

Imagínense y recuerden qué es lo que ocurre, que no rechazamos el estudio A, no decimos que sea falso, hay diferencias entre ambos estudios, no se pueden comparar. Sin embargo, en realidad, si un estudio no rechaza al otro el primero tampoco puede confirmar al segundo, y nos quedamos con esto. La única conclusión a la que podemos llegar: los resultados de los ECA a gran escala no son capaces de llevarnos a ninguna verificación o falsificación independiente. Ni confirman ni rechazan, si creen ustedes o aceptan los resultados de cualquier estudio a gran escala creo que lo harán simplemente por una cuestión de fe, por un acto de fe. Eso no es ciencia.

Esto me lleva a otro tema y es el siguiente, si no podemos demostrar que los resultados son ciertos ¿cómo podemos saber que no son el resultado de un fraude? Si hay alguien aquí que no crea que el fraude es algo común, le sugiero que estudie Internet o las revistas profesionales porque hay muchísimos ejemplos de fraude en la medicina.

Si quieren estudiarlo de una manera más formal pueden hacerlo a través de encuestas y auditorías, y volverán a encontrar que hay fraude, pero un fraude además bastante común. En Suecia, en Estados Unidos y en Noruega; miren donde miren y busquen fraude en la medicina lo encontrarán, hasta cierto nivel u otro nivel, y creo que esto es especialmente preocupante cuando hablamos de los ECA a gran escala.

La industria farmacéutica siempre pretende obtener enormes beneficios con los fármacos que estudian los megaensayos. Se ve implicada en los ensayos, en su diseño, en la recogida de los datos, en el análisis, incluso en la redacción de los resultados, y tienen eslabones financieros muy íntimos y muy cercanos con los investigadores. Hay motivo y oportunidad, y debemos garantizar que los resultados no estén abiertos a desafíos.

Un cínico podría decir que nos encontramos ante tierras fértiles para el fraude, por lo menos deberíamos preocuparnos con respecto a que existe la posibilidad de que se produzca fraude. Después de todo lo que he dicho sería razonable preguntarse: ¿por qué se alegan cosas tan fehacientemente con respecto a los ensayos clínicos? ¿Por qué no grita todo el mundo en contra de ellos? Es porque existen diferentes grupos que están empecinados en defenderlos porque tienen intereses ocultos.

La educación médica, los alumnos que estudian medicina, los internistas, todos están adoctrinados. Nunca se les pide que cuestionen la base, los cimientos. El *establishment* de la investigación médica, su reputación a menudo se basa en los ECA a gran escala y en muchos casos también tienen algún tipo de implicación financiera personal. Los gobiernos, en concreto en el Reino Unido, estamos ahora en un proceso de reforma del Servicio Nacional de Salud. Muchas de estas reformas se basan en la medicina basada en la eviden-

cia que dependen de los ECA a gran escala. Pero lo que es todavía más importante es que los gobiernos defienden estos estudios porque la industria farmacéutica es un sector muy importante dentro de la economía; sin duda en Gran Bretaña lo es.

Al final del todo nos encontramos con que la que principalmente defiende estos ensayos a gran escala es la industria farmacéutica. Si no los hubiera no se podría demostrar ningún tipo de efecto de estos fármacos. Es la única manera que tienen de convencer a las instituciones y a las autoridades de que sus productos tienen algún tipo de efecto. De no existir, la industria farmacéutica o la empresa debería inventarlos.

Los ECA a gran escala tienen que ver con enfermedades comunes, eso es cierto por definición. Esto significa que están sujetos a manipulación. Esto es cierto en las publicaciones pero es incluso todavía más cierto en los materiales de promoción, y finalmente la mayoría redundante de la que hemos hablado antes aumenta en gran medida el mercado de estos fármacos.

Lo que quiero decir con todo esto es que los ECA a gran escala obtienen enormes beneficios, millones y millones de dólares para la industria farmacéutica.

Hay un aspecto adicional en la defensa de estos estudios, que también debemos tener en cuenta y es lo que llamamos el argumento de «*No hay alternativa*», cualquiera de ustedes que haya escrito de alguna forma crítica con respecto a los ensayos clínicos sabrá de qué estoy hablando. Los revisores de las revistas dicen una y otra vez: «*es que no hay alternativa a los ensayos a gran escala*», por lo tanto hay que aguantarse y aceptarlos. Pero este argumento también es falso.

En primer lugar, no responde a los desafíos planteados contra los ensayos a gran escala. En ese sentido es lógicamente irrelevante. Este argumento, en segundo lugar, da por supuesto que los estudios tienen un valor demostrado; pero esto es un círculo vicioso. Es si se demuestra o si es valioso el argumento que estamos presentando y si finalmente los argumentos no son ciertos, es decir, hay una alternativa a los ensayos a gran escala. No es una alternativa sencilla pero es una alternativa y es volver a las ciencias, escuchar lo que dijo Hume con respecto a la regularidad y al parecido o similitud.

Comenzamos con la población de los estudios e identificando las variables superfluas, reducimos gradualmente el tamaño de esa mayoría redundante. Aumentando a la vez la homogeneidad de los estudios. De esa manera nunca será perfecto, no hay nada en las ciencias humanas que sea perfecto, pero por lo menos estaremos yendo en la dirección correcta. Entonces, una vez tengamos una población de estudio más homogénea, evaluaremos el fármaco. Identificando también en este caso las variables relevantes según vamos avanzando hasta llegar a esta situación.

Tendríamos un ensayo que en el grupo de placebo, la mayor parte de los pacientes, casi todos ellos, desarrollarán el evento y en el grupo del fármaco la mayor

parte de los pacientes no desarrollarán el episodio y aquí tendremos algo real, algo sustancial, algo que realmente vemos en la práctica clínica.

Pero obviamente reducir la mayoría superflua tendrá también sus propios críticos. Al reducir esta mayoría también bajarán los beneficios de las industrias farmacéuticas drásticamente. Si se elimina la mayoría redundante entonces también se eliminará el 90% de los beneficios y esto no lo van a aceptar con demasiada facilidad las empresas y compañías implicadas. Es muy poco probable que apoyen este cambio sin lucha y sin resistencia.

Pero antes o después yo creo que tendremos que enfrentarnos a estos problemas de los ECA a gran escala, intentando encontrar una metodología diferente. Ya les he demostrado que la VI nunca se puede dar por sentada, hay que cuestionarla y existen preguntas al respecto.

La VE siempre se puede desafiar o cuestionar. La mayoría superflua es un recordatorio constante de los fallos subyacentes a estas técnicas, a estos métodos. Tenemos datos complejos y opacos, es muy fácil manipular los datos tanto a nivel de publicación como a nivel de material de promoción, y obviamente esto nos permite que los intereses ocultos cambien los datos a su favor.

Tenemos unos efectos de los tratamientos muy pequeños, los pacientes no están convencidos con respecto a los beneficios y probablemente no acepten los tratamientos si se les da toda la información, y yo creo que estos resultados en términos de pacientes individuales no tienen ningún tipo de significación clínica.

También deberíamos mencionar que esta gran mayoría superflua significa que la mayor parte de los pacientes que reciben los fármacos podrían haber tenido el mismo resultado o muy parecido sin el fármaco. Por lo tanto, estamos malgastando a gran escala los recursos de los sistemas sanitarios públicos.

Pero por debajo de todo ello tenemos el problema de los cimientos, fundamentos de los ECA a gran escala. Esto no es ciencia, las inferencias causales estadísticas son muy cuestionables. Hay otra cosa importante, y es que no podemos, de forma independiente, verificar los resultados. Obviamente nunca podemos estar seguros de que no haya fraude en ellos.

Aquellos de ustedes que acepten mis ideas se irán pensando que los ensayos a gran escala deben cambiarse, pero sospecho que para muchos de ustedes, adoctrinados con ellos y sus opiniones, no aceptarán lo que acabo de decir, en cuyo caso ya saben lo que tienen que hacer con los herejes. ■