



Metascience: reflections on the reproducibility of research in pharmacology

Title in Spanish: *Metaciencia: reflexiones sobre la reproducibilidad de la investigación en farmacología*

Santiago Cuéllar Rodríguez¹

¹Académico Correspondiente de la Real Academia Nacional de Farmacia, Madrid

*Corresponding Author: santiago.cuellar.rodriguez@gmail.com An Real Acad Farm Vol. 85, Nº 1 (2019), pp. 4-11

Received: January 21, 2019 Accepted: February 25, 2019

Language of Manuscript: Spanish

1. INTRODUCCIÓN

El Diccionario de la Real Academia Española define la ciencia como el conjunto de conocimientos obtenidos mediante la observación y el razonamiento, sistemáticamente estructurados y de los que se deducen principios y leyes generales.

Sin embargo, en esta definición se echa de menos la referencia a la forma y el método de cómo se obtienen esos conocimientos y su desarrollo ulterior. En efecto, la ciencia implica tanto *adquisición sistemática* de conocimiento como su *acumulación organizada* y su *desarrollo evolutivo*. Asimismo, cabe agregar un tercer elemento constitucional: la ciencia es una *referencia del conocimiento* o, al menos, de un cierto tipo de conocimiento, ya que *no* todo el conocimiento posible es susceptible de ser abordado por la ciencia, debido a las restricciones que impone su propia metodología; la ciencia no puede – ontológicamente – afrontar la globalidad de la realidad, porque su visión se basa en modelos que, aunque puedan ser muy complejos o muy útiles, simplifican – reducen – la realidad e impone, al menos, algún grado de confirmación empírica directa o indirecta.

En definitiva, la ciencia es – debe ser – *pública*; es decir, accesible para cualquier persona suficientemente cualificada en la materia en cuestión, sin ninguna otra exigencia que entender el método científico estándar. También debe ser *concreta*, comunicable en términos claramente definidos – no ambiguos – y compartidos por la comunidad científica internacional; igualmente, debe estar *sujeta a una arquitectura simbólica y lógica específica y compartida* (matemática, química, etc.), independiente de la lengua particular de cada científico. Finalmente, la ciencia debe ser *objetiva*, es decir, ajena a ideas particulares no argumentadas ni contrastadas (1).

Formulado en términos muy elementales, el *método científico* es definido tradicionalmente como un procedimiento estandarizado que utiliza básicamente cuatro pasos consecutivos – susceptibles de reiteración, lo

que permite optimizar progresivamente el *proceso* – que pueden identificarse con los siguientes:

- *Observación de fenómenos* o, en ocasiones, en lugar de la observación – o además de ella – una idea deducida teóricamente a partir de otras.
- *Formulación de una hipótesis*, como un modelo que intenta explicar el fenómeno observado o justificar la consistencia lógica de la idea teórica.
- Realización de un *experimento* que permita confirmar la validez y el rango predictivo de la hipótesis formulada, utilizando unos procedimientos previamente contrastados que garanticen la máxima objetividad y representatividad del experimento, en línea directa con la hipótesis formulada.
- *Formulación de una ley*, si el experimento confirma con un margen suficiente de fiabilidad estadística el modelo propuesto por la hipótesis. Esto debe hacerse mediante un sistema simbólico y estructural apropiado – matemático químico, etc. – que permita integrarla en un ámbito científico y tecnológico ampliamente compartido, es decir, *traslacional*.

No obstante, estas cuatro características básicas – formuladas en estos términos – son excesivamente genéricas y, además, no recogen o trivializan otros aspectos fundamentales, particularmente los relacionados con la *experimentalidad*, u obligatoriedad de la experimentación, la *formalización* de la hipótesis y de los resultados, la *transparencia*, la *publicidad* y la *aceptabilidad*.

La *experimentalidad* se refiere a la continuidad entre el proceso de observación, la formulación de la hipótesis y el diseño y la ejecución del experimento, para comprobar – o refutar – la hipótesis propuesta. Un elemento fundamental de la experimentalidad es la *reproducibilidad*; es decir, la cualidad de que el experimento produzca los mismos resultados cuando se repita en las condiciones previamente establecidas. Un estudio científico que no sea reproducible

deja de tener carácter científico *ipso facto*; podrá ser conocimiento hipotético, pero no será conocimiento científico.

Con la *formalización* – incluyendo la *matemática* y la *química* – se busca incluir los elementos observacionales dentro de un marco definido y aceptado mayoritariamente de leyes y teorías – es decir, un *paradigma* –, que se expresa con un lenguaje inequívoco, exento de particularismos culturales o sociales. Además, las observaciones, hipótesis, experimentos y lenguajes científicos deberían ser *transparentes y públicos*; es decir, deben ser claramente definidos y difundidos en medios de acceso generalizado.

Otro de los pilares fundamentales que soportan el valor de la prueba de la evidencia científica – es decir, su *aceptabilidad* generalizada – es la *objetividad*, manifestada como la ausencia de prejuicios, sesgos de opinión o intereses espurios; sin embargo, es obvio que una objetividad absoluta es inalcanzable, por lo que lo más próximo que hemos acertado a inventar es la *intersubjetividad*, la coparticipación y el consenso mayoritario de la comunidad científica a la hora de reconocer un procedimiento, un dato o una relación de causalidad; bien es cierto que bajo el término de *comunidad científica* se incluye a una oscura mezcla de personajes con intereses políticos, sociales y económicos diversos.

Finalmente, a todas estas características de la ciencia debería añadirse la de la *provisionalidad*; es decir, la argumentación teórica y la demostración experimental de todo procedimiento científico deben estar sujetas a una revisión crítica continua. Nada en la ciencia, ni en ninguna otra forma de conocimiento racional, debería aceptarse como definitivo.

2. METACIENCIA, REPRODUCIBILIDAD Y REPETIBILIDAD

Podemos definir la *metaciencia* como la práctica de la metodología científica para estudiar el método científico y la filosofía de la ciencia; en otras palabras, la ciencia estudiándose a sí misma. Este término ha ido cobrando impulso en las últimas décadas, particularmente en el contexto de lo que ha sido denominada, discutiblemente, como la "crisis de reproducibilidad" de la ciencia.

Por tanto, la metaciencia pretende caracterizar la validez de un ámbito de investigación específico, como por ejemplo la fiabilidad de los estudios en biomedicina y su aplicabilidad-representatividad para la autorización oficial de nuevos medicamentos e indicaciones terapéuticas, así como para la elaboración y reforma de guías de actuación clínica; igualmente, pretende optimizar a través de una crítica rigurosa y sistemática las prácticas editoriales (especializadas y generalistas), la transparencia de la financiación de los estudios y de los intereses de sus autores, y la idoneidad de los métodos estadísticos empleados en cada caso.

La *reproducibilidad experimental (empírica)* es, como se indicó anteriormente, la cualidad de que un experimento

produzca los mismos resultados – o muy similares, dentro de un estrecho margen de confianza – cuando se repite en las condiciones y con los métodos descritos para dicho experimento. Es un concepto claramente relacionado con los de *repetibilidad* y *testabilidad*. Por repetibilidad entendemos la propiedad de obtener resultados consistentes al *replicar* – repetir – un experimento científico pero utilizando un conjunto distinto de datos a los empleados en el estudio original, pero coherentes con estos, siguiendo la misma metodología experimental. La *experimentalidad* de un estudio científico debe implicar necesariamente que la hipótesis pueda ser *refutable* o *falsable*, utilizando la terminología de Popper (2), ya que si no puede probarse que una hipótesis puede ser falsa, tampoco podría probarse que es cierta en términos científicos. La *falsabilidad* significa que los contraejemplos – datos experimentales contrarios a la hipótesis establecida – son lógicamente posibles y su viabilidad práctica de ser observados es real.

Pese a que dichos conceptos son claros, en muchas ocasiones las publicaciones científicas refieren protocolos, métodos, procesos de cálculo y conjuntos de datos que son *opacos*, lo que *impide reproducir* hallazgos aparentemente relevantes. Así ocurre en dos de cada tres artículos destacados de psicología, uno de cada tres de economía experimental y más del 75 % de los que identifican dianas terapéuticas contra el cáncer (3).

La reproducibilidad de un estudio es, repetimos, una pieza fundamental del método científico y sin ella, se corre el riesgo de enfrentarnos a una hipertrofiada bibliografía repleta de información falsa o, al menos, de cuestionable fiabilidad, que nunca se llega a corregir o depurar. Por eso, es tan importante investigar la propia investigación científica, a través de la *metaciencia*, para diseñar o corregir los métodos que permiten realizar, evaluar, revisar y difundir correctamente la ciencia.

Es fácil comprobar que en los artículos originales publicados en revistas científicas, el apartado destinado a describir la metodología empleada en el experimento cada vez tiene un papel menos relevante, reduciéndose en muchos casos a la mínima expresión y empleando una terminología imprecisa; además es común el recurso a métodos y herramientas de investigación no validados previamente de forma empírica o se emplean procedimientos inapropiados de inferencia estadística; en este sentido, las pruebas de significación de hipótesis nula con informes de valores *p* están integradas en millones de artículos, a pesar de ser un método inferencial deficiente para la mayoría de estos estudios (4).

3. ¿UNA METODOLOGÍA CIENTÍFICA APROPIADA?

John Ioannidis estableció (5) hace algunos años un conjunto de seis corolarios en torno al diseño de estudios biosanitarios que, a la vista del desarrollo que está teniendo este ámbito de la investigación científica, adquieren un valor renovado y ayudan a alertar sobre la superinflación de estudios irrelevantes o incluso abiertamente falsos a que se está sometiendo la

bibliografía científica correspondiente. De forma sintética, son:

- Cuanto menor sea el *tamaño de la muestra estudiada*, menos probable será que los resultados de la investigación sean ciertos. Es decir, los resultados de la investigación son más probables con estudios amplios, como los ensayos clínicos controlados y aleatorizados, en los que se llega a incluir varios miles de sujetos, que con estudios con pequeño tamaño muestral.

- Cuanto más pequeño sea el *tamaño del efecto* – la magnitud de la medida del cambio esperado como resultado de una intervención experimental – menos probable será que los resultados de la investigación sean ciertos. Por ello, los hallazgos más relevantes de la investigación son más probables en los estudios donde se esperan grandes efectos, como el impacto del hábito de fumar en el cáncer o en las enfermedades cardiovasculares, que en aquellos donde los efectos postulados son pequeños, como los factores de riesgo genético para enfermedades multigénicas. Por este motivo, si los tamaños del efecto real son muy pequeños – por ejemplo, con riesgos relativos menores a 1,05 – en un estudio, es probable que éste esté plagado de falsas afirmaciones, como es habitual encontrar en trabajos que estudian los determinantes genéticos o nutricionales de enfermedades complejas.

- Cuanto mayor sea el número y menor la *selección de relaciones probadas* en un campo científico, menos probable será que los resultados de la investigación sean ciertos. Por eso, los hallazgos de la investigación son más probables en los estudios con diseños confirmatorios, como los grandes ensayos clínicos controlados aleatorios de fase III, o los metanálisis de los mismos, que en los experimentos que generan hipótesis.

- Cuanto mayor sea la *flexibilidad* en los diseños, definiciones, resultados y modos analíticos en un campo científico, menos probable será que los resultados de la investigación sean ciertos. La flexibilidad aumenta el riesgo – o el potencial – para transformar resultados negativos en positivos, mientras que los hallazgos veraces son más comunes cuando los resultados o *variables* son inequívocos y universalmente aceptados y contrastados – es decir, *robustos*, como por ejemplo la supervivencia global en un periodo o la tasa de muerte por cualquier motivo – en lugar de diseños que incluyen escalas de valoración subjetiva, como las empleadas frecuentemente en neurología y psiquiatría.

- Cuanto mayores sean los *intereses y prejuicios financieros y de otro tipo* en un campo científico, menos probable será que los resultados de la investigación sean veraces. Los *conflictos de intereses* son muy comunes en la investigación biomédica y, en general, se informan de forma inadecuada y escasa, aunque no necesariamente tiene raíces financieras; a veces, los científicos pueden tener prejuicios en torno a una teoría científica o un compromiso con sus propios hallazgos, o están relacionados con la promoción académica o profesional de

los investigadores. Asimismo, algunos científicos de prestigio pueden suprimir, a través de los procesos de *revisión por pares*, la aparición y difusión de hallazgos que refuten sus propios hallazgos, condenando así su campo a perpetuar el falso dogma. En este sentido, la evidencia empírica sobre la *opinión de los expertos* muestra que ésta es muy poco fiable (6).

- Cuanto más *activo* sea un campo científico (con más equipos científicos involucrados), menos probable será que los resultados de la investigación sean ciertos. Este corolario aparentemente paradójico se debe a que, el *valor predictivo positivo* (VPP) de los hallazgos aislados disminuye cuando muchos equipos de investigadores participan en el mismo campo. Esto puede explicar por qué ocasionalmente vemos grandes emociones seguidas rápidamente por graves decepciones en campos que atraen una gran atención pública. Cuando muchos equipos de investigación trabajan en el mismo campo, dando lugar a una producción masiva de datos experimentales y ya que la rapidez es esencial para vencer a la competencia, cada equipo puede priorizar la búsqueda y difusión de sus resultados positivos más impresionantes, mientras que los resultados negativos quedan relegados, salvo que otro equipo haya encontrado una asociación positiva en la misma cuestión. El *fenómeno de Proteus* es el término acuñado para describir este fenómeno de alternancia rápida de afirmaciones y refutaciones opuestas (7).

4. REPRODUCIBILIDAD Y REPRESENTATIVIDAD (SIGNIFICACIÓN) ESTADÍSTICA

La base de datos *PubMed (US National Library of Medicine; National Institutes of Health)* comprende más de 29 millones de referencias de literatura científica biomédica. En el año 2018 se registraron 1.312.059, frente a “tan solo” 531.744 en el 2000, lo que implica que el número de registros anuales se ha multiplicado por 2,5 y esto se refiere únicamente a la literatura biomédica publicada mayoritariamente en inglés y en revistas de cierto nivel de impacto bibliográfico.

¿Toda esta información es relevante científica y clínicamente? Desde luego, no faltan expertos que consideran que toda esta información es en su mayoría ruido, por más científico que resuene su eco, ya que buena parte de esta producción es clínicamente irrelevante, cuando no preliminar, redundante o simplemente sesgada o falsa (8).

Sin embargo, la mayoría de los profesionales de la salud desconocen esta auténtica superinflación de evidencia clínica de baja calidad y cómo ella contribuye sustancialmente a un deterioro de la investigación biomédica y de la medicina clínica (9). Gran parte de la investigación médica publicada no es fiable o es, al menos, incierta, ni ofrece beneficios a los pacientes o no es útil para quienes toman decisiones sanitarias o financieras, sin que la mayoría de los profesionales sanitarios sean conscientes de este problema e incluso ni siquiera disponen, en su mayoría, de las habilidades necesarias para evaluar la fiabilidad y utilidad de la evidencia médica

publicada. Obviamente, los pacientes y las familias también carecen a menudo de evidencia médica relevante, precisa y específica en el momento de la toma de decisiones sobre su propia salud.

Atendiendo a estas limitaciones, parece razonable insistir en que los profesionales de la salud sean más sensibles a las limitaciones de la evidencia científica publicada, capacitarlos para hacer una evaluación crítica y mejorar sus habilidades de comunicación, sin olvidar a pacientes, periodistas, responsables políticos y sectores implicados o interesados en la sanidad.

El *metanálisis* es una útil herramienta estadística que permite agrupar y sintetizar datos procedentes de varios estudios con contenido, metodología y temática similares, recopilando estimaciones de un cierto efecto (expresado en un índice de *tamaño del efecto* de cada estudio), facilitando la valoración del efecto. La utilización en biomedicina del metanálisis permite valorar estos efectos en contexto: si el tamaño del efecto es consistente en el metanálisis, su precisión es mayor que con un solo estudio y, si el tamaño del efecto varía, esa variación puede ser descrita y explicada.

El valor intrínseco de la metodología metanalítica ha disparado la publicación de revisiones sistemáticas y metanálisis durante los últimos años en prácticamente todas las disciplinas biosanitarias. En este sentido, *PubMed* registraba 2.336 metanálisis publicados en 2006, 11.355 en 2016 y son más de 15.000 los referenciados en 2018, mientras que las revisiones sistemáticas son más del doble de esa cantidad. Sin embargo, solo una mínima parte de estos artículos aportan alguna novedad en sus respectivos temas ni resultan útiles, y muchos de ellos ni siquiera están bien hechos. En definitiva, la superposición, la redundancia y la duplicación han ido proliferando de forma exponencial; de hecho, la revisión sistemática de 73 metanálisis publicados en 2010 mostró que dos tercios tenían al menos otro metanálisis superpuesto que cubría las mismas comparaciones, indicaciones, ajustes y resultados, o de algunos subconjuntos de los mismos, llegando el grado de superposición y redundancia a ser impresionante en algunos temas (10).

La *falta de transparencia y de accesibilidad a los datos originales* es otro aspecto muy relevante, como ha sido puesto de manifiesto por la revisión de una muestra aleatoria de 149 artículos biomédicos publicados entre 2015 y 2017, en los que se investigó la proporción de fuentes de financiación pública y/o privada, y los posibles conflictos de intereses, compartiendo protocolos y datos sin procesar (11). Según este trabajo, un buen número de estos 149 estudios no revelaba información sobre financiación (31 %) o conflictos de interés (31 %). Entre los 104 artículos con datos empíricos en los cuales sería pertinente el intercambio de protocolos o de datos, solo el 18 % discutieron datos disponibles públicamente y solo uno de ellos (1,0 %) incluyó un enlace a un protocolo de estudio completo. Entre los 97 artículos en los que podría ser pertinente la replicación en estudios con datos diferentes, apenas hubo cinco intentos de replicación (5,2

%). En opinión de los autores de este informe, a pesar de que ha habido mejoras en los últimos años en ciertos indicadores clave de reproducibilidad y transparencia, todavía existen oportunidades para mejorar las prácticas de investigación reproducibles en toda la literatura biomédica.

Otro aspecto particularmente relevante es saber si los *procedimientos de análisis estadístico* utilizados en la investigación biomédica, especialmente en el ámbito clínico, son los más adecuados para garantizar la reproducibilidad de los resultados. Un ejemplo paradigmático es la denominada *prueba de significación de la hipótesis nula* (H_0 ; *NHST*, *null hypothesis significance test*) propuesta por Fisher.

Esta prueba considera dos probabilidades (p), la *crítica* y la *calculada*, y un tipo de error (*Error Tipo I*). La p crítica, denominada también *alfa* (α), hace referencia a la probabilidad asociada a la decisión de rechazar la H_0 cuando es verdadera, es decir, al Error Tipo I. La p crítica se establece en base a un juicio subjetivo respecto a las consecuencias de cometer un Error Tipo I y el valor máximo aceptado en la comunidad científica suele ser de 0,05. Por su parte, la p calculada o nivel de significación p del resultado empírico expresa la probabilidad de obtener un estadístico muestral igual o mayor al obtenido, dado un determinado tamaño muestral y asumiendo que la muestra ha sido extraída de una población en la que la H_0 es exactamente verdadera. Esta probabilidad depende de los valores de los parámetros reales de la población de la que se ha extraído la muestra y del tamaño muestral. Si H_0 es verdadera en la población, los estadísticos muestrales serán menos probables a medida que aumenta el tamaño de la muestra y la p calculada será menor. En este contexto, cuando la p calculada es igual o menor que la p crítica se rechaza la hipótesis nula (H_0), lo que significa que los resultados obtenidos no son consecuencia del error muestral.

Con todo, la *prueba de significación de hipótesis nula* tiene varias deficiencias que probablemente contribuyen a la crisis de replicación que afecta a la ciencia biomédica en general. Tales deficiencias llevan a sugerir que esta prueba no debería ser la práctica estadística dominante y predeterminada de todas las investigaciones biomédicas y psicológicas, sino que se debería considerar un método inferencial específico para cada tipo de investigación (12). En cualquier caso, siempre que los investigadores utilicen la *NHST*, deberían justificar su uso y publicar los cálculos previos al estudio y los tamaños del efecto, incluidos los hallazgos negativos.

5. REPRODUCIBILIDAD DE LA INVESTIGACIÓN CLÍNICA Y AUTORIZACIÓN OFICIAL DE MEDICAMENTOS

La *Food & Drug Administration (FDA)* y la *European Medicines Agency (EMA)* son las agencias reguladoras oficiales que, en Estados Unidos y en la Unión Europea respectivamente, tienen la responsabilidad de recibir, estudiar, evaluar y dictaminar la procedencia de las solicitudes de autorización de comercialización de nuevos

medicamentos y, en su caso, de nuevas indicaciones para medicamentos previamente autorizados, para su aprobación definitiva por las autoridades correspondientes que, en el caso de la Unión Europea, es la Comisión Europea.

Por consiguiente, atendiendo al enorme potencial económico que implica este tipo de decisiones, el principal *leit motiv* de la mayor parte de la investigación clínica que se realiza a escala mundial, es de la obtención de evidencia científica que permita sustentar adecuadamente una demanda de registro oficial por una agencia reguladora de un nuevo medicamento o de una nueva indicación terapéutica en el ámbito geográfico correspondiente.

Obviamente, en este ámbito competencial, la *reproducibilidad* se convierte en una condición no solo exigible en términos de metodología y ética científica, sino también en una cuestión con numerosas y muy relevantes ramificaciones de índole legal y económica. Por ello, se han llevado a cabo varios trabajos dirigidos a estudiar la calidad de la investigación clínica relativa en relación con la autorización oficial de nuevos medicamentos y/o nuevas indicaciones terapéuticas. Nosotros hemos seleccionado tres amplios estudios recientes que ilustran la importancia de esta cuestión y cómo se implementa en términos reales.

El primero de estos estudios (13) tuvo como objetivo conocer en qué medida el uso de medicamentos con indicaciones no autorizadas (*off-label*) – muy prevalente aunque obviamente controvertido y solo legal en condiciones muy estrictas – produce resultados clínicos inferiores a los medicamentos autorizados oficialmente para una misma indicación. Para ello, los autores realizaron una búsqueda sistemática en *PubMed*, *MEDLINE*, *PubMed Health* y la *Cochrane Library* hasta mayo de 2015, que incluían metanálisis de ensayos clínicos aleatorizados (ECA) que comparaban los medicamentos con o sin autorización para una misma indicación en un mismo tipo de población.

En total, se seleccionaron 25 comparaciones de tratamiento que incluían 153 ECA y 24.592 pacientes. En seis de las 25 comparaciones (24 %), los medicamentos con la indicación no autorizada (*off-label*) fueron significativamente superiores (cinco de 25) o inferiores (uno de 25) a los tratamientos aprobados, aunque hubo una *heterogeneidad* estadística sustancial en las comparaciones ($I^2=43$ %; que representa la proporción de la variación entre estudios respecto de la variación total, donde un valor del 0 % es considerado excelente, hasta el 25 % es baja, hasta un 50 % es moderada y por encima de este porcentaje es elevada). No obstante, en general, los medicamentos que no tenían autorizada la indicación produjeron resultados más favorables que los tratamientos aprobados (cociente de probabilidad u *odds ratio*, $OR=0,72$; $IC_{95} \%$ 0,54 a 0,95). Por su parte, el análisis de los resultados de las revisiones sistemáticas comparativas ($n=22$) mostró menos heterogeneidad y ninguna diferencia estadísticamente significativa en general ($OR=0,85$; $IC_{95} \%$ 0,67 a 1,06; heterogeneidad $I^2=0$ %). En definitiva, los autores concluyeron que el estatus de aprobación oficial no

indica de manera fiable qué medicamentos producen efectos más favorables en los ensayos clínicos aleatorizados que comparan la indicación *off-label* con la autorizada.

El siguiente estudio (14) analizó con qué frecuencia la *European Medicines Agency* (EMA) ha recomendado a la Comisión Europea la autorización de medicamentos en base a estudios no aleatorizados y si existe una asociación entre los efectos del tratamiento y la preferencia de EMA por realizar más pruebas en ensayos clínicos aleatorizados (ECA). Para ello, revisó todas las autorizaciones iniciales de comercialización en la base de datos de EMA de medicamentos para humanos entre 1995 y 2015.

De un total de 723 expedientes relativos a la autorización de nuevos medicamentos, 51 (7,1 %) fueron autorizados en base a estudios clínicos no aleatorizados. Estos 51 medicamentos fueron autorizados para 71 indicaciones. En los 51 pares de indicación-medicamento sin necesidad de aportar más ECA, las estimaciones del efecto fueron grandes ($OR=12,0$; $IC_{95} \%$ 8,1 a 17,9) en comparación con las estimaciones del efecto en los 20 pares de indicación-medicamento para los cuales se requirieron nuevos ECA ($OR=4,3$; $IC_{95} \%$ 2,8 a 6,6), con una diferencia estadísticamente significativa entre los efectos ($p=0,0005$).

En resumen, se emplearon datos procedentes de ensayos clínicos no aleatorizados para el 7 % de las autorizaciones de medicamentos de la EMA, con la particularidad de que los tamaños de efectos más grandes se asociaron con una mayor probabilidad de aprobación basándose solo en datos no aleatorizados, aunque no se encontró un umbral definido del efecto del tratamiento para la aprobación del fármaco sin datos aleatorizados.

Por último, otro reciente estudio (15) ha evaluado si los ensayos clínicos realizados en determinados países podrían distorsionar los resultados, en el sentido de magnificar el efecto terapéutico del fármaco ensayado. En concreto, se realizó un estudio recopilando todos los metanálisis publicados sobre gabapentina y pregabalina en las principales bases de datos bibliográficos internacionales, explorando específicamente el elevado nivel de significación en los metanálisis y si se analizaron los estudios atípicos con resultados extremos, es decir, con diferencias superiores a 0,8 desviaciones estándar (*s*, *s* o *SD: standard deviation*) con relación al valor medio del metanálisis.

Todos los 10 metanálisis evaluados mostraron resultados favorables estadísticamente significativos para ambos fármacos. Sin embargo, se observaron estimaciones de heterogeneidad I^2 que excedían el 90 %, claramente excesiva, en 4 metanálisis del empleo gabapentina o pregabalina en dolor posoperatorio. En estos 4 metanálisis, 77 estudios clínicos tenían estimaciones que diferían $>0,8$ s del valor medio; de ellos, 39 (51 %) presentaron resultados extremadamente favorables y 33 (43 %) provenían de países poco desarrollados y sin tradición de investigación clínica, 22 (29 %) no informaron sobre la financiación del ensayo clínico y 20 (30 %) no reportaron

conflictos de interés de los autores. Por el contrario, 27 de los 38 (71 %) estudios con resultados desfavorables provinieron de países desarrollados y con una larga tradición en investigación clínica.

6. ¿MEDICINA PERSONALIZADA VS. MEDICINA BASADA EN LA EVIDENCIA?

Estamos abocados a hacer una *medicina personalizada* o *de precisión*, gracias a los avances convergentes en el conocimiento y el desarrollo de tecnologías como la biología de sistemas, la secuenciación genómica, el análisis exploratorio de alto rendimiento, el progresivo descubrimiento de biomarcadores predictivos muy selectivos y asequibles para una implantación generalizada, y la disponibilidad de datos y ciencia computacional que permite la elaboración de modelos integrados de enfermedades, así como el registro electrónico generalizado de datos sanitarios, la disponibilidad de tecnologías inteligentes (sistemas expertos) y la llamada *medicina P4 (P4M): Participativa, Personalizada, Predictiva y Preventiva*. Es decir, estamos accediendo de forma progresiva a una medicina guiada por biomarcadores en los niveles de los sistemas, que toma en cuenta los avances metodológicos y los descubrimientos de los perfiles fisiopatológicos integrales de las enfermedades complejas poligénicas, multifactoriales y degenerativas, con el objetivo de optimizar la eficacia de la prevención y el tratamiento de enfermedades, considerando las características específicas de cada individuo (genéticas, bioquímicas, fenotípicas, psicosociales, así como su entorno y estilo de vida) para intervenciones específicas a través de la implementación de la mencionada *P4M* (16).

Sin embargo, la implementación clínica de la medicina de precisión plantea numerosos desafíos. En primer lugar, la medicina de precisión debe contrastarse con la práctica contrastada y ampliamente utilizada de la *medicina basada en la evidencia*, que se basa en metanálisis o estudios centrados en grupos, de los cuales se derivan recomendaciones promediadas. Obviamente, este enfoque de "talla única para todos" puede proporcionar soluciones inadecuadas para valores atípicos, los cuales están lejos de ser una rareza ya que todas las personas caen dentro de esta categoría para algún rasgo. Precisamente, tales valores atípicos se pueden manejar mejor usando medicina de precisión.

Por ello, es necesario establecer un puente entre la medicina de precisión y la medicina basada en la evidencia, aunque esto requerirá el intercambio responsable de datos a nivel mundial, así como programas de capacitación actualizados periódicamente. De esta manera, la *medicina de precisión basada en la evidencia* cambiará progresivamente de la terapia a la prevención, lo que llevará finalmente a una atención sanitaria centrada en el ciudadano (17).

Con el curioso título de "*La medicina basada en la evidencia ha sido secuestrada: mantén el rumbo y lanza a los piratas por la borda*", John Ioannidis publicó (18) una

certera reflexión acerca de las críticas que se han formulado contra la *medicina basada en la evidencia*, tales como centrarse en los beneficios e ignorar los eventos adversos; estar interesado en valores promedios e ignorar la gran variabilidad en los riesgos individuales y la capacidad de respuesta; ignorar la interacción clínico-paciente y el juicio clínico; conducir a algún tipo de reduccionismo; y caer presa de la corrupción por conflictos de intereses.

Para Ioannidis, ninguna de estas deficiencias es necesariamente inherente a la medicina basada en la evidencia. De hecho, ésta ha contribuido en gran manera a minimizar esas mismas deficiencias en la investigación clínica y en la atención médica. Sin embargo, la medicina basada en la evidencia está pagando el precio de su éxito: después de haber sido ampliamente reconocida, ha acabado siendo manipulada o simplemente mal usada para apoyar intereses espurios. A veces, los conflictos de intereses llegan a tal nivel que cabe preguntarse si este aparente secuestro de la medicina basada en la evidencia será reversible. Sin embargo, no puede olvidarse que la medicina basada en la evidencia es un valioso conjunto de herramientas conceptuales y, por ello, *vale la pena intentar eliminar los sesgos provocados por los piratas que han secuestrado el barco*, tal como lo expresa Ioannidis.

7. ¿ESTAMOS ANTE UNA CRISIS DE REPRODUCIBILIDAD DE LA CIENCIA BIOSANITARIA?

Sin duda alguna, nos encontramos actualmente en el seno de un intenso debate en torno a la importancia de la reproducibilidad como un indicador de la calidad y fiabilidad de la investigación científica, no solo en el ámbito biosanitario, debido a lo que muchos consideran como un aumento de las prácticas fraudulentas, de la falta de rigor y, en general, de la merma de la fiabilidad en la producción de conocimiento. En este sentido, según una encuesta realizada en 2016 por la revista *Nature* (19), alrededor del 90 % de los encuestados estuvo de acuerdo en que hay una crisis "leve" o "significativa", y entre el 40 % y el 70 % estuvo de acuerdo en que la notificación selectiva, el fraude y las presiones para publicar "siempre" o "a menudo" contribuyen a que la investigación no sea reproducible. Es obvio que los resultados de esta encuesta, que no fue aleatoria, pueden no representar a toda la población de científicos, pero reflejan una creciente corriente de opinión en el ámbito científico.

¿Está la ciencia biosanitaria en crisis? Esta es la pregunta que cada vez más personas – científicos o no – se están haciendo. No obstante, hablar de *crisis de la ciencia* sea probablemente exagerado, aunque es cierto que actualmente nos enfrentamos con algunos nuevos desafíos a añadir a los tradicionales, particularmente en los que refiere a la *reproducibilidad*, con resultados que son fabricados, falsificados, sesgados, con poca potencia, seleccionados y, en definitiva, irreproducibles.

Sin embargo, es preciso tener en cuenta que el propio concepto de reproducibilidad puede ser interpretado de

múltiples maneras e incluso cumplir diversas funciones epistémicas, según la investigación de que se trate. Dada esa multiplicidad y diversidad, hay autores (20) que señalan que la búsqueda acrítica de la reproducibilidad como un valor epistémico global puede resultar engañosa y potencialmente perjudicial para el avance científico e incluso hay casos en que la mejora de la reproducibilidad no promueve una investigación de alta calidad.

Así pues, afirmar que "la ciencia está en crisis" carece de fundamento empírico y, como mantienen algunos especialistas (21), es incluso contraproducente ya que en lugar de inspirar a las jóvenes generaciones a hacer más y mejor ciencia, podría fomentar en ellas el cinismo y la indiferencia; en lugar de invitar a un mayor respeto e inversión en la investigación, se corre el riesgo de desacreditar el valor de la evidencia y alimentar posiciones anticientíficas, tal como hemos podido comprobar con los movimientos anti-vacunas, entre otros diversos.

En realidad, las quejas sobre una disminución en la calidad de la investigación se repiten a lo largo de la historia de la ciencia, desde sus mismos inicios. No obstante, hay un par de elementos que caracterizan situación actual y que permiten un cierto grado de optimismo. El primero es que la validez real de estas preocupaciones se está evaluando científicamente mediante la *metaciencia*. El segundo es el creciente poder de las tecnologías de la información y la comunicación (TIC), que están transformando las prácticas científicas en todos los campos, al igual que están transformando todos los demás aspectos de la vida humana. Estas tecnologías facultan y facilitan que la investigación sea más precisa, poderosa, abierta, democrática, transparente y autocrítica que nunca. Al mismo tiempo, esta revolución tecnológica crea nuevas expectativas y nuevos desafíos a los que los *metacientíficos* se están enfrentando.

Por lo tanto, la situación actual de la ciencia biomédica podría describirse mejor como "nuevas oportunidades y desafíos" o incluso como una "revolución", en lugar de una "crisis", no solo porque es una forma más convincente e inspiradora, sino que también porque, sencillamente, está mejor respaldada por la evidencia empírica (21).

8. REFERENCIAS

1. Cuéllar Rodríguez S. Some thoughts on science. *An Real Acad Farm.* 2015; 81(3): 230-8. <http://www.analesranf.com/index.php/aranf/article/download/1606/1685>
2. Popper KR. Conjeturas y refutaciones. El desarrollo del conocimiento científico. ISBN 9788475091464 (1983).
3. Ioannidis JPA. Rethink Funding. *Sci Am.* 2018; 319(4): 52-5. Doi: 10.1038/scientificamerican1018-52.
4. Ioannidis JPA. Massive citations to misleading methods and research tools: Matthew effect, quotation error and citation copying. *Eur J Epidemiol.* 2018; 33(11): 1021-3. Doi: 10.1007/s10654-018-0449-x.
5. Ioannidis JPA. Why most published research findings are false. *PLoS Med.* 2005; 2(8): e124.
6. Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. *JAMA.* 1992; 268: 240-8.
7. Ioannidis JP, Trikalinos TA. Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol.* 2005; 58: 543-9.
8. Casino G. El ruido y la crisis. Sobre la insostenibilidad del actual modelo de comunicación médica. En: *Escepticismo. Una mirada escéptica sobre la salud y la información.* Cuadernos de la Fundación Dr. Antonio Esteve nº 34, p. 21 (2015). ISBN: 9788494257186.
9. Ioannidis JPA, Stuart ME, Brownlee S, Strite SA. How to survive the medical misinformation mess. *Eur J Clin Invest.* 2017; 47(11): 795-802. Doi: 10.1111/eci.12834.
10. Siontis KC, Ioannidis JPA. Replication, Duplication, and Waste in a Quarter Million Systematic Reviews and Meta-Analyses. *Circ Cardiovasc Qual Outcomes.* 2018; 11(12): e005212. Doi: 10.1161/CIRCOUTCOMES.118.005212.
11. Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015-2017. *PLoS Biol.* 2018; 16(11): e2006930. Doi: 10.1371/journal.pbio.2006930.
12. Szucs D, Ioannidis JPA. When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment. *Front Hum Neurosci.* 2017; 11: 390. Doi: 10.3389/fnhum.2017.00390.
13. Ladanie A, Ioannidis JPA, Stafford RS, Ewald H, Bucher HC, Hemkens LG. Off-label treatments were not consistently better or worse than approved drug treatments in randomized trials. *J Clin Epidemiol.* 2018; 94: 35-45. Doi: 10.1016/j.jclinepi.2017.11.006.
14. Djulbegovic B, Glasziou P, Klocksieben FA, Reljic T, VanDenBergh M, Mhaskar R, et al. Larger effect sizes in nonrandomized studies are associated with higher rates of EMA licensing approval. *J Clin Epidemiol.* 2018; 98: 24-32. doi: 10.1016/j.jclinepi.2018.01.011.
15. Desai K, Carroll I, Asch S, Hernandez-Boussard T, Ioannidis JPA. Extremely large outlier treatment effects may be a footprint of bias in trials from less developed countries: randomized trials of gabapentinoids. *J Clin Epidemiol.* 2018 Oct 23. pii: S0895-4356(18)30291-9. doi: 10.1016/j.jclinepi.2018.10.012.
16. Cuéllar Rodríguez S. El futuro ya está aquí. *Panorama Actual Med.* 2018; 42 (419): 1131-2.
17. Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Med.* 2016; 8(1): 134.

18. Ioannidis JPA. Hijacked evidence-based medicine: stay the course and throw the pirates overboard. *J Clin Epidemiol.* 2017; 84: 11-3. Doi: 10.1016/j.jclinepi.2017.02.001.
19. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature.* 2016; 533(7604): 452-4. doi: 10.1038/533452a.
20. Leonelli S. Rethinking Reproducibility as a Criterion for Research Quality, in Luca Fiorito, Scott Scheall, Carlos Eduardo Suprinyak (ed.): *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise (Research in the History of Economic Thought and Methodology, Volume 36B)*. Emerald Publishing Limited, pp.129-46. (2018).
21. Fanelli D. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proc Natl Acad Sci U S A.* 2018; 115(11): 2628-31. doi: 10.1073/pnas.1708272114.