



## ¿QUÉ NOS PASÓ?

### MEDICINA BASADA EN LA EVIDENCIA, EPIDEMIOLOGÍA CLÍNICA, ESTADÍSTICA, METODOLOGÍA DE LA INVESTIGACIÓN

La investigación clínica no está funcionando bien: la mayoría de los hallazgos publicados son falsos y, además, inútiles<sup>1,2</sup>.

Yo pertenezco a una generación de médicos que vivió un cambio drástico en la forma de ver y ejercer la profesión: fuimos formados en una época en la que las decisiones médicas se tomaban con base en la experiencia de los profesores (nadie se preguntaba si las conductas eran correctas o no) y nos tocó ejercer en otra época, donde las decisiones deberían fundamentarse en estudios epidemiológicos del tipo de los ensayos clínicos. Dicho de otro modo, pertenezco a la generación de médicos que vio nacer la Medicina Basada en la Evidencia (MBE).

Cuando esta historia comenzó la MBE parecía una especie de movimiento intelectual que muchos abrazamos con un entusiasmo loco: la medicina dejaría de ser un “arte” para convertirse en una ciencia. No más “*en mi experiencia...*”, bienvenidos los datos y la estadística. Esta nueva forma de entender el ejercicio médico tenía un superhéroe que se llamaba David Sackett (1934 – 2015), un internista educado en Harvard que fue de los primeros en hablar del tema, fundó el Departamento de Epidemiología Clínica de la Universidad de McMaster y escribió en los 90s (con sus compañeros de gesta) los primeros artículos guía sobre la apreciación crítica de la literatura médica<sup>3</sup>.

El camino que tuvo que recorrer la MBE no fue fácil, entre otras cosas porque los profesores más rancios la entendieron como una vía que los nuevos médicos queríamos usar para sacarlos del juego: o abrazaban la ciencia o se quitaban del medio. Como dije arriba, algunos médicos colombianos medio rebeldes e intelectualmente inquietos desafiábamos la autoridad y no compartíamos eso de que el paciente tenía una caverna en un pulmón porque un señor encontraba una “pectoriloquia áfona” que nadie más identificaba (¿no ve que era áfona?), sino

que queríamos evidencias objetivas de lo que los profesores decían. Este sacrilegio fue mejor recibido en Bogotá, donde amigos y colegas fundaron el Departamento de Epidemiología Clínica de la Universidad Javeriana, cuya influencia en la educación médica colombiana es innegable. Cambiar paradigmas en Medellín es cosa seria hoy e impensable hace 25 años, por lo que los que allí nos formamos recorrimos un camino más solitario y menos aceptado. Cosas de la vida. Afortunadamente se cruzaron en mi camino personajes maravillosos como el Dr. Oscar Lema Tapias, el Dr. Arturo Morillo Quiñones, el Dr. Juan Luis Londoño F. y muchos otros en los que pude apoyarme para acoger esta nueva forma de ver la medicina. Después se me apareció la virgen y vine a dar a Cali y a la Fundación Valle del Lili, a trabajar como cardiólogo intervencionista y a difundir este mensaje... pero eso, es otra historia.

Todo parecía ser simple: la MBE mejoraría nuestra profesión porque las decisiones clínicas se tomarían con base en la mejor evidencia disponible. Si para un determinado problema la literatura ofrecía evidencia de calidad, el médico debería usarla. Ese era todo el misterio.

Pero las cosas no han salido como creímos que lo harían. Como dije al principio, la mayoría de los hallazgos publicados son falsos y, además, inútiles. ¿Qué nos pasó? De eso se trata este artículo.

El 21 de junio del 2018 el New England Journal of Medicine hizo un tipo de publicación que nunca pensé que llegaría a ver<sup>4</sup>. Se trata de un refrito de los datos del PREDIMED, un estudio español que pretendía evaluar el efecto que la dieta mediterránea podría tener sobre el riesgo cardiovascular en pacientes que, de entrada, eran de alto riesgo. Desde que se publicó la primera versión del protocolo en 2003<sup>5</sup> pensé que el estudio tenía pocas probabilidades de salir adelante por múltiples razones: 1. Los pacientes de alto riesgo cardiovascular tienen signos de una compleja enfermedad metabólica que comprendemos poco, que tiene un sustrato inflamatorio y un importante componente genético. El solo pensar que podemos conjurar la fisiopatología de una enfermedad sistémica tan seria agregándole un poco de aceite de oliva a la dieta es subestimar el problema de forma descarada. En otras palabras, la probabilidad pretest de que la hipótesis sea cierta era, de entrada, extremadamente baja. 2. Aún en los pacientes de alto riesgo cardiovascular la incidencia de eventos es muy baja, por lo que los investigadores deberían usar una muestra enorme, prácticamente imposible de controlar, en especial si pensamos en intervenir su estilo de vida y su dieta. 3. Probablemente utilizarían un desenlace combinado, que es un recurso cuya única pretensión de fondo es hacer el estudio factible en términos de tamaño de muestra, es decir, es una forma amañada y artificial de mejorar el poder del estudio, que lo vuelve factible, pero afecta considerablemente la confiabilidad de los resultados. 4. Aunque la población estudiada es de alto riesgo, la prensa y los interesados querrán extrapolar los resultados a toda la humanidad, vendiéndole la falsa ilusión de que están haciendo algo real para evitar un infarto. Muchos médicos a los que les fascina ser entrevistados en los medios de comunicación acogerían felices los resultados y la medicina, en lugar de avanzar, retrocederá. Ya veo a Camila Zuluaga y a otros periodistas por el estilo tragando aceite de oliva y nueces porque “somos lo que comemos” 5. Los sesgos que este estudio tendría son de todos los tipos y de una magnitud tal que, si los autores fueran estrictos, el proyecto no debería ni pasar del protocolo. Nadie me pondría atención sobre estas

preocupaciones porque, al fin y al cabo, ¿qué importancia puede tener un simple cardiólogo de provincia, retirado del ejercicio clínico y que a todo le pone problema?

Pues sí señor, el estudio se realizó y sus resultados se publicaron en el NEJM en 2013<sup>6</sup>. Resulta que los investigadores detectaron “*desviaciones del protocolo*” y se retractaron de la publicación casi de inmediato. Las tales “*desviaciones del protocolo*” fueron en realidad muy graves e incluyeron el hecho de que algunos pacientes recibieron una dieta sin ser aleatorizados, en uno de los sitios se usaron mal las tablas de aleatorización, etc. Aun así, como el resultado sugirió que la dieta mediterránea prevenía eventos en esta población, el efecto mediático fue el esperado y la dieta mediterránea está siendo recomendada ampliamente, incluso hoy en día. Pero el tema de fondo es simple: los datos de este estudio no permiten negar la hipótesis nula y por lo tanto no prueban de ninguna manera la utilidad de la dieta mediterránea. Pues bien, los investigadores que retiraron el estudio se pusieron a revisar (con más cuidado) la base de datos y volvieron a publicar el estudio el 21 de junio de este año<sup>4</sup>. Aunque parezca increíble los editores del NEJM permitieron semejante despropósito, pues dice literalmente en la revista que se publican “*estimaciones de efectos de la dieta mediterránea, revisadas, y basadas en un análisis que **no se fundamenta exclusivamente en la suposición de que todos los participantes fueron asignados aleatoriamente***”. Francamente, no se ni cómo redactaron esta frase tan rimbombante para decir que la aleatorización salió mal.

Algo está mal. Muy mal.

Pues bien, el 25 de junio del 2018, solo cuatro días después de la publicación del NEJM, Eric Topol (editor en jefe de Medscape, un muy reconocido cardiólogo cuya trayectoria vale la pena estudiar<sup>7</sup>) le hizo una entrevista al doctor John Ioannidis<sup>8</sup>, que resulta ser el antihéroe de toda la historia que vengo contándoles. Un señor personaje (a mí me parece simplemente un genio) al que hay que estudiar y seguir.

John P. A. Ioannidis es un internista norteamericano que creció en Grecia, donde obtuvo varios reconocimientos académicos, como el premio nacional de la Sociedad Griega de Matemáticas y algunos más. Se formó como médico en Atenas, hizo medicina interna en Harvard y enfermedades infecciosas en Tufts. Se trasladó a Stanford en 2010 donde, entre muchas otras cosas, dirige un centro de estudios llamado METRICS<sup>9</sup> (Meta-Research Innovation Center at Stanford - @METRICStanford en Twitter). Ioannidis es un experto en estadística y metodología de la investigación que construyó una carrera sobre la crítica del trabajo de otros investigadores, con lo que se inició una nueva disciplina que se llama meta investigación, que no es otra cosa que estudiar los estudios de investigación hechos por otros con el fin de evaluar su calidad. Ahora está buscando formas de mejorar el proceso científico. Ioannidis es también escritor y poeta, ha publicado más de 1.000 artículos científicos y viaja por todo el mundo dictando conferencias, gracias al enorme respeto que le tienen las grandes figuras de la medicina mundial. De manera más reciente ha incursionado en la evaluación de las investigaciones que se hacen en otros campos como la economía<sup>10</sup>, con resultados que tienen sorprendida a la comunidad científica.

Ioannidis saltó a la fama al publicar en 2005 un ensayo en PLOS Medicine titulado “*Por qué la mayoría de los hallazgos publicados en investigación son falsos*”<sup>1</sup>. Este ensayo tiene más de 2.5 millones de visitas en el portal de PLOS (de lejos el artículo más leído y citado de la Public Library of Science) y ha convertido a Ioannidis en el terror de los malos investigadores y de la mala ciencia. Este investigador se ha ganado el respeto y la admiración de la comunidad científica mundial defendiendo dos premisas: “la mayoría de los hallazgos publicados son falsos<sup>1</sup> y, además, inútiles<sup>2</sup>” y proponiendo soluciones<sup>8,11</sup>.

La forma como Ioannidis presenta su caso es lo que resumo a continuación, aunque le recomiendo a los interesados estudiar en detalle las publicaciones originales que, en mi humilde opinión, son fundamentales si los médicos queremos replantear la forma como entendemos, hacemos y consumimos investigación. Me permito utilizar sin aspavientos lo que Ioannidis publica pues los artículos en los que me baso están en PLOS Med (la revista médica de la Public Library Of Science) bajo una licencia “*Creative Commons*” que permite el uso, distribución y reproducción del contenido, **por cualquier medio**, siempre y cuando se cite al autor original, cosa que no me canso de hacer a lo largo de estas páginas<sup>1-2</sup>. He hecho todo el esfuerzo para traducir la información, no solo del inglés al español, sino también del lenguaje matemático y técnico que usa Ioannidis, a una gramática más convencional, que les facilite la comprensión a todos. Lo que presento no es una traducción literal, pues lo descrito tiene incorporados párrafos de mi autoría que tienen como objetivo enfatizar, donde me ha parecido adecuado o necesario, el mensaje de Ioannidis. He compilado también las referencias que me han parecido más relevantes y en cualquier caso, insisto: para una lectura más precisa y profunda los remito directamente al autor.

Estoy convencido de que poco ha salido como creímos que lo haría en los 90s, cuando dábamos los primeros pasos en Medicina Basada en la Evidencia. Se trataba de una idea para integrar la habilidad clínica de cada médico con la mejor evidencia externa disponible y hoy tenemos médicos bajo una tremenda presión del mercado, vemos que demasiadas discusiones en los departamentos clínicos se tratan acerca de dinero, tanto, que yo creo que hoy el término MBE ha sido reemplazado por Medicina Basada en las Finanzas (MBF). Por años he sentido mucha desilusión con lo que he visto en investigación médica, sin embargo, Ioannidis ha logrado poner todo esto en blanco y negro y sus planteamientos han hecho que mire el tema con renovada esperanza.

Empecemos entonces por la publicación del 2005<sup>1</sup>, donde el autor nos explica por qué los resultados de la mayoría de las investigaciones son falsos. El tema podríamos resumirlo más o menos así:

## **LA MAYORÍA DE LOS HALLAZGOS POSITIVOS SON FALSOS**

Lo primero sobre lo que debemos reflexionar es el hecho de que, con mucha frecuencia, los médicos vemos que las afirmaciones hechas por un estudio son refutadas por otro posterior y viceversa (hay abundante literatura al respecto). Este asunto llega a volverse tan confuso que al clínico se le vuelve casi imposible encontrar una respuesta clara a sus preguntas de todos los

días, porque la literatura estará llena de estudios que soportan y refutan prácticamente cualquier hipótesis. Ante tanta contradicción deberíamos colegir entonces que la gran mayoría de los hallazgos publicados son falsos. Ioannidis trata de modelar una estructura que permita entender mejor los hallazgos positivos que resultan falsos, argumentando inicialmente que este problema es consecuencia de la infundada estrategia de emitir conclusiones con base en un único estudio, respaldado simplemente por una prueba estadísticamente significativa, definida por un valor de  $p < 0.05$ . Entendemos entonces como hallazgo positivo de una investigación cualquier relación (intervención efectiva, factor de riesgo, etc.) que tenga significancia estadística formal. Todos sabemos que no solo se requiere significancia estadística para proclamar la confirmación de una hipótesis. De hecho, la probabilidad de que un hallazgo positivo de una investigación sea realmente cierto depende de tres cosas fundamentales: **1.** La probabilidad previa (antes de hacer el estudio) de que una hipótesis sea cierta (a esto lo conocemos como probabilidad pretest), **2.** El poder estadístico del estudio y sus sesgos y **3.** El nivel de significancia estadística. Pues bien, en cualquier campo de investigación se pueden construir tanto hipótesis ciertas como falsas acerca de la presencia de una determinada relación. Siendo así, Ioannidis propone definir como  $R$  a la proporción del número de relaciones verdaderas sobre el número total de relaciones investigadas en ese campo.  $R$  es entonces una característica propia de cada área de investigación y varía mucho dependiendo de si el campo de investigación se enfoca en relaciones altamente probables o se concentra en solo una (o unas pocas) de las millones de hipótesis que podrían postularse. Siendo así, la probabilidad pretest de que una relación sea cierta es  $R/(R+1)$ , la probabilidad de que un estudio encuentre una relación verdadera refleja su poder ( $1-\beta$  [1-el error estadístico tipo II]), mientras que la probabilidad de afirmar la existencia de una relación cuando esta no existe en la realidad representa el error estadístico tipo I ( $\alpha$ ). Si definimos a  $c$  como el número de relaciones que están siendo estudiadas en un determinado campo, Ioannidis pudo elaborar una tabla de  $2 \times 2$  que correlaciona lo que se encontró en una investigación (hallazgos positivos) con lo que existe en la realidad:

RELACIÓN ENCONTRADA EN LA INVESTIGACIÓN	RELACIÓN QUE EXISTE EN LA REALIDAD (GOLD STANDARD)		
	SI	NO	Total
SI	$c(1 - \beta)R/(R + 1)$	$c\alpha/(R + 1)$	$c(R + \alpha - \beta R)/(R + 1)$
NO	$c\beta R/(R + 1)$	$c(1 - \alpha)/(R + 1)$	$c(1 - \alpha + \beta R)/(R + 1)$
Total	$cR/(R + 1)$	$c/(R + 1)$	$c$

DOI: 10.1371/journal.pmed.0020124.t001

A partir de este punto podemos entender que, después de que una relación se reclama como presente o verdadera porque tuvo significancia estadística formal, la **probabilidad post test** de que realmente sea cierta se conoce como el valor predictivo positivo (VPP), que no es otra cosa que el complementario de la probabilidad de que un reporte positivo sea en realidad falso. De acuerdo con la tabla anterior se puede establecer que el  $VPP = (1 - \beta)R/(R - \beta R + \alpha)$  y por lo tanto un hallazgo de una investigación es más probable que sea verdadero que falso si  $(1 - \beta)/R > \alpha$ .

Como la gran mayoría de las investigaciones dependen de un  $\alpha = 0.05$ , esto significa que un hallazgo positivo tiene mayores posibilidades de ser cierto si  $(1 - \beta)/R > 0.05$ .

Como resulta evidente que tanto los sesgos como la cantidad de estudios hechos por equipos independientes alrededor del mundo pueden distorsionar el análisis anterior y conducir a probabilidades aún más pequeñas de que los resultados de una investigación resulten realmente ciertos, Ioannidis modeló estos dos factores y produjo nuevas tablas de  $2 \times 2$  que los incorporan. En aras de la simplicidad voy a abstenerme de publicar acá estas dos tablas, sin embargo, describiré brevemente la forma como el autor entendió el efecto que sesgos y pruebas independientes tienen sobre el VPP.

### Sesgos

Los entendemos como la combinación de varios factores del diseño, el análisis de los datos y su presentación, que tienden a producir hallazgos positivos de investigación cuando en realidad no existen. Un sesgo puede implicar manipulación en el análisis y reporte de los hallazgos mientras que los reportes selectivos o distorsionados son también una forma de sesgo. El autor define entonces como  $u$  al porcentaje de hallazgos que se reportan positivos de manera errónea gracias al efecto de los sesgos y uno puede asumir que  $u$  no depende de si una relación es verdadera o no. Pues bien, cuando hay sesgos (casi siempre los hay), uno puede establecer que el  $VPP = ([1 - \beta]R + u \beta R) / (R + \alpha - \beta R)$ . Siendo así, el VPP baja cuando suben los sesgos ( $u$ ), a no ser que  $1 - \beta < \alpha$  (o sea  $1 - \beta < 0.05$ , en la mayoría de los casos). Dicho de otro modo, mientras más sesgos, la probabilidad de que un hallazgo positivo sea realmente cierto disminuye considerablemente, independiente del nivel de poder del estudio y de la probabilidad pretest de que la hipótesis sea cierta.

### Investigaciones de varios equipos independientes

Varios equipos de investigadores pueden estar estudiando las mismas preguntas de investigación. Si uno de estos grupos afirma haber encontrado una relación positiva recibirá toda la atención. Ahora bien, la probabilidad de que al menos uno de los grupos de investigación (entre muchos) encuentre una relación positiva es fácil de estimar: para  $n$  estudios independientes de igual poder, el  $VPP = R(1 - \beta^n) / (R + 1 - [1 - \alpha]^n - R\beta^n)$  (sin considerar los sesgos). Cuando se aumenta el número de estudios independientes el VPP tiende a disminuir, a menos que  $1 - \beta < \alpha$  (o sea  $1 - \beta < 0.05$ , en la mayoría de los casos). Esto es independiente del nivel de poder del estudio y de la probabilidad pretest de que la hipótesis sea cierta.

Esto se me parece mucho a lo que antes llamábamos corrección de Bonferroni<sup>12</sup>, que se hacía imperativa cuando se emplean muchas pruebas de  $p$  en un estudio pues, si se realizaban muchas de  $p$ , aparecía una probabilidad de que alguna diera significativa por el simple hecho de hacer tantas pruebas de  $p$ . Mejor dicho, si muchos investigadores independientes trabajan en la misma hipótesis, es más posible que alguno reclame un hallazgo positivo que realmente no existe.

Pues bien, a partir de este tipo de modelos Ioannidis pudo obtener varios corolarios interesantes sobre la probabilidad de que los hallazgos positivos de una investigación sean realmente ciertos:

**Corolario 1.** Mientras más pequeño sea un estudio, menor la posibilidad de que sus hallazgos sean ciertos. Mientras más pequeño el tamaño de la muestra menor el poder del estudio y, para todas las funciones descritas arriba, el VPP de un hallazgo positivo disminuye en la medida en que el poder baja hacia  $1 - \beta = 0.05$ .

**Corolario 2.** Mientras menor sea el tamaño del efecto, menor es la probabilidad de que un hallazgo positivo sea realmente cierto. Esto se debe a que el poder está relacionado con el tamaño del efecto. Como la investigación moderna se enfoca (obligada en la medida en que avanza) en efectos cada vez más pequeños, es de esperar que la proporción de hallazgos realmente verdaderos sea cada vez menor. De hecho, si hay un campo de investigación en el cual el tamaño del efecto estudiado es intrínsecamente bajo, es muy probable que dicho campo esté plagado de hallazgos positivos falsos. Un ejemplo de esto es el estudio de los determinantes genéticos o nutricionales de una enfermedad compleja, que le confieren a la población un riesgo relativo menor de 1.05. En estos casos la epidemiología genética o nutricional se convierte en una empresa utópica.

**Corolario 3.** Mientras mayor sea el número y menor la selección de relaciones estudiadas en un campo científico, es menos probable que los resultados de una investigación sean ciertos. Como se indicó atrás, el VPP (probabilidad de que un hallazgo positivo sea realmente cierto) depende mucho de la probabilidad pretest de que la relación estudiada sea cierta ( $R$ ). Siendo así, es mucho más probable que un hallazgo positivo sea cierto en un estudio confirmatorio (como un ensayo clínico aleatorizado grande o un metaanálisis) que en un experimento exploratorio que solo pretende generar hipótesis.

**Corolario 4.** Mientras mayor sea la flexibilidad de los diseños, definiciones, desenlaces y modelos analíticos en un campo científico, menor será la probabilidad de que un hallazgo positivo sea realmente cierto. Esto se debe a que la flexibilidad es sinónimo de sesgos. Siendo así, es fácil entender que los hallazgos realmente verdaderos son más comunes cuando los desenlaces son claros y universalmente reconocidos (como muerte) que cuando se usan desenlaces multifactoriales (como escalas cualitativas de una enfermedad). De forma similar, cuando los estudios usan métodos analíticos conocidos y ampliamente usados (como Kaplan Meier o Logrank) tienen mayor probabilidad de que los hallazgos positivos sean realmente ciertos que cuando se usan métodos analíticos experimentales (como los de inteligencia artificial). En cualquier caso, por más estricto que parezca el diseño de un estudio, los sesgos casi siempre terminan siendo un problema inevitable. Para decirlo de otra forma, hay mucha evidencia de que el informe selectivo de resultados y la manipulación (consciente o no) de los desenlaces y análisis reportados representa un serio problema, incluso en los ensayos clínicos aleatorizados<sup>13</sup>.

**Corolario 5.** Mientras mayores sean los intereses financieros u otros intereses y prejuicios en un campo de investigación, menor será la probabilidad de que un hallazgo positivo sea realmente cierto. Esto se debe a que los conflictos de intereses y los prejuicios son una fuente infinita de sesgos. No parece necesario profundizar en la cantidad de conflictos de intereses financieros que hay en la investigación biomédica, donde médicos e investigadores frecuentemente tienen vínculos económicos con la industria de medicamentos y dispositivos. De otro lado, los prejuicios no necesariamente tienen que ver con dinero, sino que los investigadores pueden estar sesgados simplemente porque creen en una teoría científica o en sus hallazgos de investigación. En muchos casos los investigadores sienten la presión académica (*publish or die*) porque obtener resultados significa promoción y crecimiento. Es bien conocido que muchos investigadores prestigiosos pueden detener, vía evaluación por pares, la aparición de resultados que contradicen los suyos, condenando un campo científico a creer de forma perpetua en un dogma falso.

**Corolario 6.** Mientras más “caliente” (con mayor cantidad de grupos de investigación involucrados) sea un campo de investigación, menor la probabilidad de que un resultado positivo sea realmente cierto. La explicación de este fenómeno fue descrita algunos párrafos atrás y puede explicar por qué es tan frecuente que campos de investigación que atraen mucho la atención se vean llenos de resultados positivos y posteriores decepciones: además de la baja probabilidad de que un resultado positivo resulte realmente cierto, muchos investigadores se sienten atraídos por refutar cosas publicadas en revistas de prestigio. No es raro ver cómo se alternan con gran velocidad publicaciones cuyos resultados son opuestos, lo que se conoce como fenómeno de Proteus<sup>14</sup>, el dios griego que podía cambiar rápidamente de apariencia.

Pues bien, según Ioannidis, en este tipo de modulación o marco de referencia es muy difícil obtener un VPP > 50%. En la publicación original hay una tabla que ilustra los resultados de simulaciones hechas con las fórmulas descritas y de acuerdo con la influencia del poder del estudio, de la proporción del número de relaciones verdaderas sobre el número total de relaciones investigadas en ese campo ( $R$ ) y de los sesgos, para varias situaciones que podrían ser características de diseños típicos:

$1 - \beta$	$R$	$u$	Ejemplo práctico	VPP
0.80	1:1	0.10	Ensayo clínico controlado con buen poder, pocos sesgos y $R$ de 1:1	0.85
0.95	2:1	0.30	Metaanálisis confirmatorio de ensayos clínicos controlados de buena calidad	0.85
0.80	1:3	0.40	Metaanálisis de estudios pequeños y no concluyentes	0.41
0.20	1:5	0.20	Ensayos clínicos controlados de fase I o II, de bajo poder pero bien realizados	0.23
0.20	1:5	0.80	Ensayos clínicos controlados de fase I o II, de bajo poder y pobre realización	0.17
0.80	1:10	0.30	Estudio epidemiológico exploratorio con poder adecuado	0.20
0.20	1:10	0.30	Estudio epidemiológico exploratorio con bajo poder	0.12
0.20	1:1000	0.80	Investigación exploratoria (con muchas pruebas) orientada a hacer descubrimientos	0.0010
0.20	1:1000	0.20	Como la anterior, pero con menores sesgos debido a que es mejor estandarización	0.0015

Los VPP se derivan asumiendo un  $\alpha = 0.05$

DOI: 10.1371/journal.pmed.0020124.t004

Podemos ver, por ejemplo, que un resultado positivo de un ensayo clínico aleatorizado bien hecho, con poder adecuado y con una probabilidad pretest del 50% resulta ser realmente cierto el 85 % de las veces (VPP). Un desempeño similar se podría esperar de un metaanálisis confirmatorio hecho con ensayos clínicos controlados de buena calidad porque, aunque el sesgo potencial probablemente sube, el poder y la probabilidad pretest son mejores que los de un solo ensayo clínico aleatorizado.

De otro lado, un metaanálisis de estudios pequeños y no concluyentes, donde se intenta usar la agrupación para “corregir” el bajo poder de los estudios individuales, es poco confiable si el  $R$  es  $\leq 1:3$ . Los hallazgos positivos que provienen de ensayos clínicos controlados de fase I o II, de bajo poder pero bien realizados, son realmente ciertos en uno de cada 4 casos, o incluso en menos si se presentan sesgos.

Un estudio epidemiológico exploratorio con  $R$  de 1:10 se desempeña de peor manera (VPP = 0.2), en especial cuando el poder es bajo (VPP = 0.12).

Finalmente, la probabilidad de que un resultado positivo sea real en una investigación exploratoria orientada a hacer descubrimientos a partir de muchas pruebas, es decir, cuando las hipótesis evaluadas exceden en 1000 a 1 a las verdaderas es paupérrima (como cuando se estudian 30.000 genes con la idea de que 30 puedan ser los verdaderos culpables del problema). Esto es así aunque se eviten los sesgos mediante una adecuada estandarización del laboratorio, los métodos estadísticos, los resultados y los informes.

Una forma de ver el tema es pensar en que, cuando un investigador anuncia un hallazgo positivo, dicho hallazgo puede no ser más que una medición precisa del sesgo predominante. Como ya se ha visto, la mayoría de la investigación biomédica moderna opera en áreas donde los hallazgos positivos tienen muy bajas probabilidades pre y post test de ser ciertos. Por ejemplo, supongamos que no existen patrones nutricionales o dietéticos que sean determinadores reales del riesgo de desarrollar determinado tumor. Supongamos también que la literatura científica ha examinado 60 nutrientes y reclama que todos ellos se relacionan con el riesgo de desarrollar el tumor, con riesgos relativos que van de 1.2 a 1.4 al comparar los terciles superior en inferior de ingesta. Pues bien, es claro entonces que el tamaño de las relaciones encontradas son una simple medida de nada más ni nada menos que del sesgo neto del estudio.

En aquellos estudios con un VPP muy bajo, las pocas relaciones verdaderas que existan no logran distorsionar mucho el panorama de lo que estamos hablando. Aunque realmente existan algunas relaciones positivas verdaderas, la forma de la distribución de los efectos observados aún arrojaría una medida clara de los sesgos involucrados en ese campo de investigación. Este concepto pone boca abajo la forma como vemos los resultados científicos. De forma tradicional los investigadores han visto los efectos grandes y significativos con mucho entusiasmo, pues se trata de signos de un descubrimiento importante. Hoy en día, un efecto muy grande y muy significativo puede ser signo de un gran sesgo en muchos campos de la investigación moderna y debe hacer que los investigadores revisen con mucho cuidado los datos y el estudio, buscando qué pudo salir mal con los números obtenidos, el análisis y los resultados. La única ventaja que

veo en todo esto es que, el obtener medidas del sesgo neto implícito en un área de investigación, puede servirnos para establecer cuál podría ser la magnitud de los sesgos que operarían en otros campos donde se empleen métodos analíticos, tecnologías y conflictos similares.

Ioannidis termina esta maravillosa publicación<sup>1</sup> preguntándose cómo se podría mejorar la situación actual. ¿Resulta inevitable que la mayoría de los hallazgos positivos sean falsos? ¿Podríamos hacer algo al respecto? Un problema muy importante es que es imposible saber con 100% de certeza cuál es la verdad en ninguna pregunta de investigación. El estándar de oro es sencillamente inalcanzable. Ahora bien, existen cosas que podemos hacer para mejorar la probabilidad post test:

- Hacer estudios con muy buen poder, es decir, estudios grandes o metaanálisis no sesgados. Lo que pasa es que los estudios grandes implican sesgos inevitables y, por lo menos yo, no sé cómo se hace un metaanálisis que no tenga sesgos. Aun así, esta debería ser la ruta correcta.
- Los estudios a gran escala deberían usarse para preguntas de investigación que tengan una probabilidad pretest realmente alta, de modo que un hallazgo positivo resultaría en un VPP que podría considerarse casi que definitivo.
- Los estudios a gran escala resultan particularmente indicados cuando pueden evaluar conceptos grandes en lugar de preguntas estrechas muy específicas. Un hallazgo negativo refutaría no solo una pregunta propuesta, sino todo un campo o una porción significativa de este. Cuando se hacen estudios a gran escala con criterios estrechos de evaluación, como la promoción y mercadeo de una droga específica, generalmente terminamos con unos datos que van a dormir el sueño de los justos: dinero, tiempo y trabajo perdidos.
- Hay que tener cuidado con los estudios muy grandes pues tienen mayor probabilidad de encontrar una relación estadísticamente significativa para un efecto trivial que no se diferencia en nada de los efectos nulos. Es algo que solemos llamar un efecto estadísticamente significativo pero clínicamente irrelevante.
- Tratar de disminuir los sesgos mediante estándares de investigación mejorados y aboliendo los prejuicios sería muy útil. El tema es que esto requiere un cambio en la mentalidad científica que puede ser difícil de lograr.
- Hay mejor probabilidad de lograr una investigación exitosa si se registran con anterioridad los estudios, en especial los ensayos clínicos controlados.
- Finalmente, en lugar de perseguir la significancia estadística, deberíamos mejorar nuestra comprensión del rango de valores de  $R$  y la probabilidad pretest, que es donde los esfuerzos de investigación realmente operan<sup>15</sup>. Antes de montarse en un estudio, el investigador debería pensar en cuál es la posibilidad de que esté evaluando una hipótesis verdadera en lugar de una falsa. En ocasiones es posible determinar valores altos de  $R$  mediante pura especulación y análisis. Como se describió arriba, siempre y cuando sea éticamente aceptable, los estudios grandes con sesgos mínimos deberían hacerse sobre relaciones que se consideran más o menos bien establecidas, para ver con qué frecuencia se confirman. Ioannidis sospecha que muchos “clásicos”

establecidos no pasarían la prueba. Como se dice en el mundo del golf: “las apuestas se ganan en el tee del hoyo 1 (cuando se plantean).

La mayoría de los nuevos descubrimientos seguirán surgiendo de hipótesis que se generan con probabilidades pretest muy bajas. Debemos entonces reconocer que las pruebas de significancia estadística reportadas en un solo estudio nos brindan solo una imagen parcial de la situación real del área de investigación involucrada si no sabemos qué tantas investigaciones se han hecho por fuera del estudio y en el campo de investigación en general. Es usualmente imposible saber que tanto “*dragado*” se le ha hecho a los datos por parte de los autores que reportan un hallazgo, o de otros grupos de investigación que han precedido un hallazgo reportado (se conoce como “*data dredging*”, “*data fishing*” o “*p-hacking*” a un proceso que involucra probar de forma automática miles de hipótesis sobre una sola base de datos, buscando combinaciones de variables que pudieran demostrar alguna correlación<sup>16</sup>). Aún si fuese posible establecer esto, no nos serviría de ayuda a la hora de establecer la probabilidad pretest, por lo que es inevitable que uno deba asumir de forma aproximada qué tantas relaciones se espera que sean verdaderas, entre todas las evaluadas en un área relevante de investigación. Aunque el *R* asumido resulte siendo subjetivo, es de todos modos muy útil al interpretar los hallazgos positivos de un estudio y ponerlos en contexto.

Todos los conceptos arriba mencionados tienen que ver con el primer artículo de Ioannidis<sup>1</sup>, según el cual la mayoría de los hallazgos positivos reportados son falsos. Como ya vimos, en esta publicación del 2005 el autor propuso algunas estrategias que podrían resultar en que una proporción mayor de hallazgos positivos resulten realmente ciertos. De todos modos, en 2014, luego de mucha meta-investigación y cientos de publicaciones adicionales, Ioannidis presentó un nuevo ensayo llamado “Cómo hacer para que más resultados publicados sean verdad”<sup>11</sup>. Como es un ensayo muy relacionado con el primero, me permito resumirlo a continuación:

### **COMO HACER PARA QUE MÁS INVESTIGACIONES PUBLICADAS RESULTEN CIERTAS**

A pesar del panorama que hemos presentado hay que reconocer que la medicina ha progresado un montón. Es innegable que cientos de nuevos tratamientos benefician a muchos pacientes. Lo que pasa es que la historia podría ser mucho mejor si la ruta de la investigación no se hubiese desviado como lo ha hecho. Estamos hablando de millones de dólares desperdiciados, miles de horas de trabajo arrojadas por la borda, cientos de vidas académicas entregadas a callejones sin salida, muchísimos pacientes tratados con intervenciones ineficaces o perjudiciales y muchas preguntas simples que siguen sin respuesta, a pesar de toneladas y toneladas de investigación.

Aun así, los logros de la investigación científica son sorprendentes. Hoy en día es toda una industria con más de 15 millones de investigadores que publicaron más de 25 millones de artículos solo entre 1996 y 2011<sup>17</sup>. Lo que pasa es que los descubrimientos nuevos realmente aplicables son pocos porque muchas de las asociaciones o efectos propuestos son falsos o descaradamente exagerados, lo que hace que la transición de conocimiento a aplicaciones útiles es con frecuencia lenta e ineficiente<sup>18</sup>. Debido a la increíble cantidad de datos

disponibles, investigar las investigaciones (meta investigación<sup>9</sup>) puede darnos estimados empíricos de la prevalencia de los factores de riesgo que existen para la enorme proporción de falsos resultados positivos. Estos factores de riesgo son por ejemplo:

- Estudios con bajo poder.
- Pequeño tamaño de los efectos.
- Baja probabilidad pretest de las hipótesis.
- Diseños flexibles.
- Definiciones flexibles.
- Desenlaces flexibles.
- Análisis flexibles.
- Sesgos y conflictos de interés.
- El “efecto bandwagon<sup>19</sup>” (fenómeno psicológico en el cual la gente asume algo básicamente porque otros lo hicieron, sin tener en cuenta la evidencia. Se le conoce como “efecto de arrastre”, “efecto de la moda” o “subirse al bus del triunfo”).
- Falta de colaboración.

Por todo lo anterior se ha estimado que el 85% de los recursos de investigación se despilfarran<sup>20</sup>.

### **Se requieren intervenciones efectivas**

Es evidente entonces que se requieren intervenciones efectivas para mejorar la credibilidad y eficacia de la investigación científica. Algunos factores de riesgo de resultados falsos son inmutables (como los efectos de pequeño tamaño), pero otros sí son modificables. Se requiere entonces disminuir los sesgos y los conflictos de intereses, así como evitar los esfuerzos fragmentados con la idea de obtener investigación más estandarizada, no sesgada, transparente y colaborativa. Hay que tener cuidado de no exagerar en las correcciones hasta el punto de congelar a los investigadores del mundo. Toca entonces implementar soluciones pragmáticas, factibles y medibles, para que podamos saber cómo se desempeñan.

Hoy en día, muchas de las decisiones sobre la forma de hacer investigación se toman con base en convenciones o inercia y no en evidencias que puedan soportar una u otra política. Por ejemplo, quienes revisan los grants (esta palabra se puede traducir como solicitudes de financiación o subvención, pero me parece que el término en inglés es más claro) casi siempre tienen una hoja de vida bastante modesta, mientras que los científicos más influyentes no revisan grants ni suelen ser financiados con fondos del gobierno. De otro lado, las prácticas no meritocráticas como el nepotismo, el sexismo y el conservatismo injustificado están muy diseminadas<sup>21</sup>. Es también real que la lealtad y los sesgos de confirmación pululan en el proceso investigativo. Sea como sea, a la hora de la práctica clínica, aunque la MBE ha crecido mucho, algunos argumentan que está en crisis<sup>22</sup> y que ha sido desplazada para promover los conceptos de los expertos y las agendas de la industria<sup>23</sup>. En realidad no hay buena evidencia experimental que nos permita saber cómo debería hacerse la revisión por pares (basada en el protocolo, en el manuscrito o después de la publicación) o cómo deberían asignarse los fondos para

investigación. De hecho, muchas de las estructuras científicas dominantes datan de la edad media (como las jerarquías académicas que usamos) o del siglo XVII (como las sociedades científicas o las revistas) y su idoneidad para enfrentar el crecimiento actual de la ciencia no es clara. Podemos hacer muchos cambios y, aunque algunos podrían perjudicar a la ciencia y deberíamos evitarlos, otros podrían mejorarla mucho.

Hay varias prácticas que se han utilizado con éxito en algunos campos de la ciencia y deberíamos trasplantarlas a la mayor cantidad de disciplinas científicas que podamos:

- Hacer estudios colaborativos a gran escala.
- Adoptar una cultura de replicación de hallazgos.
- Registrar los estudios, los protocolos, los códigos de análisis, los datos crudos y los resultados.
- Compartir los datos, los protocolos, los materiales usados, el software y otras herramientas.
- Adoptar prácticas para hacer los estudios reproducibles.
- Contener (o evitar) los patrocinadores y autores que tengan conflictos de intereses.
- Utilizar métodos estadísticos más apropiados.
- Promover la estandarización de las definiciones y los análisis.
- Adoptar umbrales más estrictos a la hora de aceptar hipótesis como verdaderas.
- Mejorar los estándares de diseño de los estudios.
- Mejorar la revisión por pares, los reportes y la forma como se disemina la investigación.
- Entrenar mejor a los científicos en metodología de la investigación y estadística.

## Las partes interesadas

A la hora de diseñar, probar e implementar intervenciones sobre las prácticas investigativas se hace necesario entender quién se verá afectado y cómo se le da forma a la investigación científica. Los científicos son solo uno de los grupos de personas involucradas y hacen parte de una red mucho mayor en la que cada parte interesada tiene expectativas diferentes, como podemos ver en la siguiente tabla:

doi:10.1371/journal.pmed.1001747.t001

Parte interesada	Grado de interés en los resultados de una investigación			
	Publicable	Financiable	Traducible*	Rentable
Científicos	+++	+++	+	
Industria (ventas y mercadeo)				+++
Industria (investigación y desarrollo)			+++	+++
Inversionistas privados (fodos de cobertura incluidos)			++	+++
Inversionistas del sector público (abiertos, como el NIH)	++		+	
Inversionistas del sector público (cerrados, como los militares)			+++	
Inversionistas filantrópicos	++		+++	
Editores de revistas	+++			+
Empresas que publican con ánimo de lucro	+			+++
Sociedades científicas y profesionales	+			
Universidades	+	+++		+
Instituciones de investigación sin ánimo de lucro	+++	+++	+	+
Personal no científico de soporte		+++		
Hospitales y otros entes que ofrecen servicios relacionados			+	+++
Otras entidades afectadas por estos servicios (aseguradoras)				+++
Gobierno				++
Consumidores de productos y servicios			+++	

\*: Traducible significa que se puede llevar a la práctica clínica

Como vemos, las partes interesadas pueden valorar una investigación de acuerdo con diferentes criterios, pues se fijan si será publicable, financiable, traducible o rentable y es evidente que las expectativas no son las mismas para cada uno de los actores. Los científicos pueden continuar haciendo publicaciones y obteniendo grants sin hacer progresos reales, si todo lo que les importa es publicar y obtener más patrocinios. Si la ciencia es financiada principalmente por inversionistas privados que buscan patentes y utilidades, esto llevaría a que los experimentos que funcionan (o que parece que funcionan) hagan su tránsito con mayor facilidad, pero perjudica la transparencia y el interés que tenemos en que la información se comparta. La influencia corporativa puede dejar de lado la ciencia, privilegiando el mercadeo mediante publicaciones en revistas prestigiosas, congresos de sociedades reconocidas y, a través de un fenómeno muy común y frecuentemente subestimado por los colegas: la conversión de los profesionales líderes de opinión en ramas de sus departamentos de mercadeo<sup>19</sup>. Las diferentes partes interesadas pueden tener sus propias preferencias sobre en qué casos la reproducibilidad debe ser promovida o rechazada: los grupos de la industria farmacéutica promueven la reproducibilidad en la investigación preclínica, pues dependen de ella para localizar nuevas drogas potencialmente útiles, sin embargo, la misma industria es muy reticente a la hora de compartir los datos de los ensayos clínicos, porque estos suceden en un momento de la historia del producto donde un reanálisis de los datos<sup>24</sup> podría invalidar la evidencia que soporta un medicamento en el cual han invertido mucho dinero.

Las dinámicas entre las partes interesadas son complejas, de hecho, a veces la misma persona puede representar diferentes partes interesadas durante el desarrollo de un estudio porque, por ejemplo, un investigador académico puede ser editor de una revista, socio de una empresa, ejecutivo de una sociedad científica, consejero gubernamental y/o beneficiario de la industria.

### **Las investigación como moneda**

Las publicaciones y los grants son una especie de moneda en la ciencia, porque compran bienes académicos, como promociones y poder. Mejor dicho, los títulos académicos y el poder son riqueza agregada para quien los posee. La tasa de cambio de estas “monedas” y el valor de los bienes académicos varía en cada ambiente institucional o en cada disciplina científica. A su vez, ambas cosas pueden verse afectadas por características propias de cada microambiente, como el nepotismo, el amiguismo o la corrupción. El poder administrativo, el trabajo en redes cerradas, el cabildeo universitario, las sociedades profesionales emparentadas y las academias distorsionan aún más el panorama. Siendo así las cosas, el estatus quo puede con facilidad seleccionar ganadores y perdedores del juego académico, produciendo gran cantidad de investigaciones mediocres o irreproducibles, controlando la revisión por pares en las revistas y promoviendo a aquellos que piensan de una determinada manera.

Ahora bien, estos hechos pueden representar también oportunidades de mejora. Por ejemplo, la necesidad de registrar los ensayos clínicos ha funcionado porque la mayoría de las revistas grandes adoptaron esta práctica como prerrequisito para la publicación de los resultados. De otro lado, los muy interesantes esfuerzos de revisión post publicación (como el PubMed Commons<sup>25</sup>) no han podido explotar su potencial como herramienta para evaluar la calidad de

una investigación porque actualmente no hay recompensa para este tipo de revisión por pares, posterior a la publicación de un manuscrito.

## Modificar el sistema de recompensas

El sistema actual de estímulos y recompensas podría ser modificado de manera significativa<sup>26</sup>. Sabemos que la escala actual valora las publicaciones, los grants, los títulos académicos y el poder previamente acumulado. Los investigadores de mayor rango tienen más publicaciones y más grants, sin embargo, los académicos que están en la parte superior de la escalera (como los rectores universitarios) suelen tener un registro pobre o mediocre de publicaciones y citas, quizás porque su destreza para el cabildeo compensa la carencia de credenciales y su éxito viene a expensas de otros candidatos más competentes que le traerían mayor rigor intelectual a la toma de decisiones de alto nivel. Es también posible que solo sea que los directivos hacen bien el trabajo burocrático que se requiere para mantener funcionando una alucinante maquinaria académica y sus habilidades permiten a los colegas con más talento científico concentrarse en la investigación. El sistema actual no premia, o incluso a veces castiga, a quienes desean replicar con rigurosidad trabajos previos. De hecho, se presiona a los investigadores para que presenten información nueva y significativa. Hoy en día no se incentiva el compartir datos, códigos de análisis y demás, bajo el precepto de que “los competidores podrían robar mis datos, mis ideas y hasta mi financiación”. Tampoco es adecuadamente valorada una revisión por pares de calidad. De hecho, la revisión por pares funciona muchas veces como una red de seguridad y un mecanismo para argumentar buena calidad. Puede ser superficial y desembocar solo en modestas mejorías del trabajo revisado, permitiendo con frecuencia la aceptación de documentos flagrantemente errados. Que la revisión por pares sea tan poco valorada no ayuda a potenciar sus beneficios ni a minimizar sus problemas.

doi:10.1371/journal.pmed.1001747.t002

	Tipos de sistemas de recompensa		
	Actual	Con un cambio moderado	Con un cambio más profundo
<b>MONEDAS</b>			
<b>Publicaciones (por unidad)</b>	Gana 1	Sin valor	Sin valor
Publicación replicada y reproducible (por unidad)	Gana 1	Gana 2	Gana 2
Publicación llevada con éxito a la práctica clínica (por unidad)	Gana 1	Gana 5	Gana 5
Publicación refutada (por unidad)	Gana 1	Pierde 1	Pierde 1
<b>Compartir datos, protocolos, códigos de análisis (por unidad)</b>	Sin valor	Gana 2	Gana 2
<b>Contribución a una revisión por pares (por unidad)</b>	Sin valor	Gana 2	Gana 2
<b>Contribución a la educación y/o entrenamiento (por unidad)</b>	Sin valor	Gana 1	Gana 1
<b>Obtención de grants (por cada R01*)</b>	Gana 5	Gana 5	Pierde 5
<b>OTROS ARTÍCULOS QUE IMPLICAN RIQUEZA</b>			
<b>Profesor asistente, título en buena universidad</b>	Gana 3	Gana 3	Sin valor
<b>Profesor asociado, título en buena universidad</b>	Gana 10	Gana 10	Sin valor
<b>Profesor titular, título en buena universidad</b>	Gana 20	Gana 20	Sin valor
<b>Director de grupo de investigación (por cada estudiante**)</b>	Gana 2	Gana 2	Pierde 2
<b>Poder administrativo, trabajo en red, cabildeo</b>	Gana hasta 200	Sin valor	Pierde hasta 200

\*: R01 es un tipo de grant del NIH.

\*\* : Estudiante haciendo un doctorado o investigador post doctorado

La tabla anterior, con propósitos ilustrativos y para que podamos entender el tipo de recompensas existentes y la forma como sesgan el proceso investigativo, compara la situación actual con dos potenciales cambios en el sistema de recompensas, uno moderado y otro más profundo. En la columna de la izquierda se presentan las actividades que “generan valor” en la comunidad científica, divididas en dos grandes categorías (“monedas” y “otros artículos que implican riqueza”). Cuando se dice que una determinada actividad gana 1 (punto o moneda), el autor está tratando de medir el valor relativo de dicha actividad en el sistema de recompensas (es mejor ganar 5 que ganar 1 o que perder 2).

Vemos entonces que hoy en día se le da un valor de 1 a una unidad publicada (como un artículo de primer autor o de autor de alto nivel, publicado en una revista respetable del campo de investigación correspondiente), 5 puntos de valor a un grant importante (como un R01), 2 por tener un fellow de postdoctorado, lo que significa que un científico obtiene el mismo valor publicando 5 artículos como primer autor (o autor de alto grado) que consiguiendo un grant R01 del NIH como investigador principal. También se obtiene lo mismo con dos artículos publicados que teniendo un fellow de postdoctorado. Hay que tener en cuenta que el valor relativo de publicar un artículo puede variar, pues hay muchas disciplinas donde se publican cientos de artículos con muchísimos autores, caso en el cual el valor relativo de cada publicación podría variar. Un sistema en el que se valore el aporte relativo de cada autor en una publicación ayudaría a asignar los créditos de un modo más adecuado.

Pues bien, en la primera propuesta de modificación de recompensas (que hemos llamado cambio moderado), se diferencia el valor de los artículos publicados dependiendo de la replicación y el traslado de los resultados a la práctica clínica, dándole entonces mayor valor a ideas sonoras cuyos resultados son replicables y reproducibles. Publicar por publicar no vale nada. Se le asigna mayor valor (5 puntos) a publicaciones que resultan en cosas que funcionan, como tratamientos efectivos, pruebas diagnósticas o herramientas pronósticas que demuestren mejoría de desenlaces clínicos importantes. También se obtiene valor adicional al compartir información, por participar de forma efectiva en una revisión por pares y al realizar actividades docentes de eficacia probada. A veces un editor o un par que revisa un manuscrito puede aportar tanto como el mismo autor.

La segunda propuesta de modificación al estatus quo es mucho más profunda. Además de los cambios adoptados en la primera propuesta, obtener grants, premios u otras formas de poder son consideradas negativas, a no ser que alguna de esas cosas derive en más ciencia de buena calidad. Los recursos y el poder son vistos como oportunidades y los investigadores tendrán que lograr que sus resultados se compadezcan con las oportunidades que tienen. Los rangos académicos no tienen valor en este modelo y podrían eventualmente eliminarse: los investigadores simplemente tienen que mantener un balance positivo entre resultados y oportunidades. En este escenario deliberadamente provocativo, los investigadores se resisten a obtener grants o a volverse poderosos porque eso se consideraría una carga. El problema potencial de este escenario es que podría desestimular el liderazgo y las aplicaciones ambiciosas a patrocinios importantes.

Cuando hablamos de hacer este tipo de modificaciones al sistema de recompensas hay que entender que pasa lo mismo que cuando se modifica una fisiopatología en un intento por luchar contra una enfermedad: las intervenciones pueden tener tanto efectos benéficos, como efectos secundarios no deseados. Cualquier cosa que se haga deberá entonces medirse y evaluarse con rigurosidad.

### **¿Qué sigue entonces?**

La magnitud en la que se podrían mejorar las prácticas de investigación actuales es desconocida. La cosa es que hay tantos problemas que es muy factible que se puedan implementar mejoras sustanciales con relativa facilidad. El ajuste fino de las políticas actuales y algunas intervenciones más radicales y disruptivas deben considerarse seriamente. Existen muchos escenarios diferentes para la evolución de la investigación biomédica y la investigación científica en general, cada uno más o menos compatible con la búsqueda de la verdad y el bienestar humano. Las intervenciones para cambiar el sistema actual no deberían aceptarse sin un escrutinio adecuado, incluso cuando sean razonables y bien intencionadas. Idealmente, deberían ser evaluadas experimentalmente. Los logros de la ciencia son sorprendentes, pero la mayoría del esfuerzo se desperdicia. Sea como sea, la ciencia es todavía la mejor forma de adquirir conocimientos que tenemos. Habrá que mejorarla, sin duda. Las intervenciones para hacer que la ciencia sea menos derrochadora y más efectiva podrían ser sumamente beneficiosas para nuestra salud, nuestro bienestar y nuestra comprensión de la verdad, y podrían ayudar a la investigación científica a perseguir con éxito sus nobles objetivos.

Hasta acá hemos visto como Ioannidis presenta de forma brillante su postulado fundamental más conocido: “la mayoría de los hallazgos publicados son falsos<sup>1</sup>”. Se trata de conceptos que muchos hemos sabido por mucho tiempo, pero Ioannidis ha tenido la lucidez para presentarlos de una manera clara y convincente, lo que lo ha hecho merecedor del respeto y el aprecio de sus pares y colegas. También propone cambios a la forma como está estructurada la investigación en el mundo<sup>11</sup> y estoy seguro que la Medicina Basada en la Evidencia verá cambios sustanciales, para mejor, en el futuro próximo.

Pues bien, no satisfecho con todo lo anterior, Ioannidis publicó en 2016 un nuevo ensayo<sup>2</sup> en el que plantea un nuevo tema: “la mayoría de la investigación clínica es inútil”. Parecería más de lo mismo, sin embargo, se trata de una sabia reflexión sobre la verdadera utilidad que debería tener la investigación. Con el fin de redondear estas líneas que compilan la información de estas tres maravillosas publicaciones<sup>1,2,11</sup>, me permito resumir entonces la que acabo de plantear:

### **POR QUÉ LA MAYORÍA DE LA INVESTIGACIÓN CLÍNICA NO SIRVE PARA NADA**

La investigación en ciencias básicas (la que en inglés se conoce “Blue-sky research”) no se hace pensando en que tenga impacto práctico inmediato. Es un tipo de investigación sin un objetivo muy claro porque se deriva de la curiosidad humana, del simple deseo de “saber”, que es en el fondo la razón de ser de la ciencia. Pero algo muy distinto sucede con la investigación clínica,

que se supone debería ser útil, se supone que debería hacer una diferencia para la salud humana o para los desenlaces clínicos y, por lo tanto, debería concebirse con esa idea en mente.

Cuando se habla de “investigación clínica”, en general nos referimos a investigaciones que abordan preguntas sobre el tratamiento, prevención, diagnóstico, detección o pronóstico de una enfermedad. Aunque los ensayos clínicos son el mejor diseño que tenemos para encontrar este tipo de respuestas, los estudios observacionales también pueden arrojar información valiosa. Pues bien, la investigación clínica útil es aquella que produce cambios favorables en la toma de decisiones, bien sea por sí misma o cuando se integra con otras investigaciones en la forma de revisiones sistemáticas, metaanálisis y guías de práctica clínica.

Varios millones de artículos de investigación clínica se han publicado hasta hoy, sin embargo, la gran mayoría de ellos no son útiles (es decir, no sirven para nada). Se ha estimado que el desperdicio de dinero en investigación médica supera el 85% de los recursos destinados para tal efecto<sup>20</sup>. Como se expuso en detalle arriba, Ioannidis había dicho que la mayoría de los hallazgos positivos de la investigación médica son falsos<sup>1</sup> y había propuesto también algunas potenciales intervenciones para actuar sobre este problema<sup>8,11</sup>. Es perfectamente claro que para que una investigación sirva para algo tiene que producir resultados ciertos (no falsos), pero esto no es suficiente. Se requiere además que los resultados positivos y verdaderos sean también útiles. Y lo primero no es sinónimo de lo segundo. En los siguientes párrafos el autor describe las características más importantes que una investigación debe tener para ser útil, presenta el estado actual de las cosas y propone cambios para mejorar.

La siguiente tabla es un resumen de lo que Ioannidis considera son las características que deben mirarse a la hora de establecer si una investigación clínica resulta de utilidad:

<b>Característica</b>	<b>Pregunta por hacer</b>
<b>Basada en un problema real</b>	¿Refleja un problema de salud lo suficientemente grande e importante como para querer arreglarlo?
<b>Situada en un contexto</b>	¿Se ha evaluado la información previa de manera sistemática para estar seguro de que se necesita un nuevo estudio?
<b>Representa una ganancia de información</b>	¿Es el estudio propuesto del tamaño suficiente como para que sea una fuente real de información?
<b>Pragmática</b>	¿La investigación representa la vida real?
<b>Centrada en el paciente</b>	¿Refleja el estudio las necesidades reales de los enfermos?
<b>Le da valor a la inversión</b>	¿Vale la pena gastarse el dinero que cuesta esta investigación?
<b>Factible</b>	¿Es posible hacer este estudio?
<b>Transparente</b>	¿Son los métodos usados, los datos y los análisis verificables y libres de sesgos?

## **Basada en un problema real**

Es de mucha utilidad resolver un problema que representa una gran carga de enfermedad para la sociedad, sin embargo, un problema de baja prevalencia pero con graves consecuencias para quien lo sufre es también importante. De hecho, al estudiar enfermedades raras es posible que el conocimiento obtenido sea relevante para condiciones clínicas más comunes (por ejemplo, al estudiar la importancia de la PCSK9 en la hipercolesterolemia familiar es posible que se encuentren nuevos tratamientos para muchos pacientes con enfermedad cardiovascular).

De otro lado, la investigación clínica se torna inútil cuando nos inventamos enfermedades inexistentes para acomodárselas a gente sana. El término que se usa en inglés es “disease mongering<sup>27</sup>”, que no es otra cosa que “venderle” enfermedades a los médicos y a la sociedad ampliando los límites de la definición de una patología, con el objetivo (casi siempre oculto) de hacer crecer los mercados de la industria farmacéutica y de dispositivos. En estos casos, tratar personas por el solo hecho de que ingresaron al nuevo rango de definiciones de una enfermedad no tiene sentido porque no hay una enfermedad real para tratar.

Esto me hizo acordar del título que tengo pensado para una conferencia que nunca le pienso dictar a un grupo de médicos en su grado: “si van a intervenir a alguien... procuren que esté enfermo”.

## **Situada en un contexto – Representa una ganancia de información**

La investigación útil proporciona una ganancia de información clínicamente relevante<sup>28</sup>. Esto significa que la investigación le agrega conocimientos importantes a lo que ya sabíamos. Para que esto suceda debemos tener claridad sobre lo que ya se sabe para que podamos poner los nuevos datos en contexto. Más aún, los estudios se deberían diseñar para que puedan proporcionar suficiente cantidad de evidencia que le asegure a los actores del sistema que las decisiones que tomen están suficientemente soportadas como para cambiar la práctica clínica. Siendo así, los estudios ideales deberían ser útiles de entrada, sin importar sus eventuales resultados. Cuando pensamos que un estudio será útil solo si arroja determinado resultado, existirá tanta presión para obtener dicho resultado que la interpretación de los datos podría hacerse difícil.

Muchas de las investigaciones actuales no están precedidas por revisiones sistemáticas de la información existente<sup>29</sup>. En otros casos una intervención se compara contra placebo o contra cuidado estándar a pesar de que hay intervenciones efectivas demostradas y bien estudiadas.

Es frecuente que se usen desenlaces subrogados que carecen de significado clínico y se ha hecho muy común el empleo de desenlaces combinados que los autores usan con la simple idea de demostrar que su estudio tiene un poder adecuado<sup>30</sup>. Además de lo anterior, rara vez se discute el valor de un resultado negativo a la hora de diseñar un estudio.

## **Pragmática**

Las conclusiones de una investigación deberían ser aplicables a situaciones de la vida real. Cuando el contexto de la investigación (como los criterios de inclusión, por ejemplo) se desvía mucho de los pacientes reales, el lector crítico debería preguntarse qué tanto afecta esto la validez de las conclusiones del estudio. Uno tiende a creer que la población de un estudio debe representar muy bien a todos los pacientes (para tratamiento) o a toda la sociedad (para prevención) para que los resultados puedan generalizarse. Resulta que, como los ensayos clínicos necesitan ser consentidos por los participantes, ningún estudio de este tipo resulta tener una muestra perfectamente aleatoria de la población general, sin embargo, los efectos del tratamiento pueden ser similares en los no participantes aunque la muestra no sea tan rigurosamente representativa, si se hacen estudios con diseños pragmáticos que, aunque muy defendidos y reclamados, pocas veces se ven. En resumen, un estudio es considerado pragmático si fue hecho en una población que se parece a los pacientes de la vida real.

## **Centrada en el paciente**

La investigación útil debe centrarse en el paciente, se hace con la idea de beneficiar a los enfermos, con el objetivo de preservar su salud y bienestar y no pensando en las necesidades de los médicos, los investigadores o los patrocinadores. Todo estudio debe alinearse con las prioridades del paciente, con la utilidad que éste le asigne a los problemas y desenlaces y con qué tan aceptable es para él una intervención evaluada en una investigación. Cuando se propone un desenlace subrogado éste debe correlacionarse muy estrechamente con lo que para el paciente real es importante en la clínica.

Hoy en día hay mucho interés en que la investigación se haga pensando en el paciente. De hecho, en el 2012 se lanzó en Estados Unidos una organización llamada PCORI (Patient-Centered Outcomes Research Institute), con la idea de fomentar el tema<sup>31,32</sup>. Aunque hay iniciativas similares en otros lugares es muy poco frecuente que se involucre a los pacientes a la hora de establecer una agenda investigativa. Tanto médicos como pacientes son constantemente bombardeados con información que pretende convencerlos de que los desenlaces subrogados (y otros con poco significado) tienen importancia real, cuando lo único que en el fondo representan son beneficios comerciales y/o el avance de alguien en su carrera académica.

## **Le da valor a la inversión**

A la hora de establecer si un estudio es útil, es importante ver si los resultados de la investigación valen el dinero que se paga por ellos. Esto se puede modelar formalmente con la idea de darle un valor a la información obtenida<sup>33</sup>, que siempre debe superar con creces el costo de conseguirla. La investigación clínica sigue siendo extremadamente costosa, aunque se ha estimado que hasta el 90% del costo de los estudios podría suprimirse sin lesionar la investigación<sup>34,35</sup>. Reducir los costos mediante la racionalización de la investigación podría hacer más que simplemente permitir que se realicen más investigaciones, porque podría

reducir la presión para cortar esquinas, asunto que suele resultar en estudios con bajo poder, poca duración, poca precisión y desenlaces inadecuados, es decir, estudios que no convencen a la hora de generar un cambio de la práctica clínica.

### **Factible**

Aunque muchas de las características descritas se cumplan, algunos estudios resultarán muy difíciles o incluso imposibles de conducir. Hay ocasiones en que las dificultades de un potencial estudio no pueden preverse de antemano, por lo que es usual ver exceso de optimismo entre investigadores y patrocinadores.

Muchos estudios clínicos se abandonan por futilidad<sup>36,37</sup>. Es importante entonces examinar con honestidad y rigurosidad la factibilidad de cada protocolo antes de decidirse a ejecutarlo, para evitar abandonarlo a mitad de camino, con todo lo que eso implica.

### **Transparente (confiable)**

La utilidad de una investigación disminuye en la medida en que no es transparente. Si los datos del estudio, el protocolo y otros procesos no están disponibles para verificación o para posterior uso por analistas independientes, entonces el estudio pierde transparencia. La confianza en el estudio se deteriora cuando aparecen sesgos importantes en el diseño, la conducción o el reporte del estudio.

Aunque se ha promovido (incluso en algunas revistas se ha exigido) que los ensayos clínicos se registren con anterioridad para que la comunidad científica se entere del protocolo y sus pretensiones, la realidad es que, por motivos que no vale la pena discutir acá, muchos estudios nunca se registran, muchos hacen un registro parcial y la mayoría no presentan sus datos crudos para el escrutinio de otros.

### **Otras consideraciones**

**Incertidumbre.** Algo de incertidumbre siempre existe sobre cada una de las características de una investigación clínica, aunque ésta siempre es menor que la que existe en el blue-sky research. La incertidumbre evoluciona en el tiempo, en especial cuando una investigación dura varios años. Las preguntas pueden perder importancia cuando cambian las circunstancias. A muchos investigadores les ha pasado que, cuando su estudio está listo para ser publicado, la pregunta de investigación se torna irrelevante porque una nueva tecnología la resolvió de un solo tajo.

**Otras fuentes de evidencia además de los ensayos clínicos.** Los estudios observacionales suelen aportar más confusión que respuestas al intentar llenar los vacíos de información. Los metaanálisis, los análisis de decisión (disciplina que se aproxima a las decisiones difíciles de manera formal) y las guías de diagnóstico y manejo no suelen salvar la situación cuando se basan en enormes estudios inútiles y pueden, además, agregar sus propios problemas y sesgos.

**Enfocándonos en revistas importantes.** Algunos clínicos prefieren leer solo investigación publicada en las revistas médicas generales más reconocidas (*The New England Journal of Medicine, The Lancet, BMJ, JAMA y PLOS Medicine*). Lo que pasa es que estas revistas cubren solo una minoría de la investigación médica que se publica. De los 730.447 artículos que se publicaron como “ensayos clínicos” en PubMed hasta mayo 26 de 2016, solo 18.231 aparecieron en las revistas más reconocidas. La mayoría de los artículos que presentan guías de manejo y de práctica clínica se publican en otras revistas. Es cierto que los artículos que aparecen en las revistas importantes suelen enfocarse en problemas clínicos mayores, pero dada su visibilidad, corren el riesgo de propagar más “*disease mongering*” que las revistas menos visibles. Es verdad que los estudios publicados en revistas conocidas suelen ser los más grandes, sin embargo, los ensayos clínicos pequeños que se publican en estas mismas revistas suelen tener resultados más exagerados que estudios similares publicados en otras partes<sup>38</sup>. La siguiente tabla describe qué tanto se satisfacen las características de utilidad de un estudio, según el tipo de revista donde aparece publicado:

FRECUENCIA DE SATISFACCIÓN DE CADA CARACTERÍSTICA DE UTILIDAD, ESTUDIOS PUBLICADOS EN REVISTAS RECONOCIDAS Y EN TODO EL AMBIENTE DE INVESTIGACIÓN CLÍNICA*		
Característica	Estudios publicados en revistas reconocidas	Toda la investigación clínica
<b>Basada en un problema</b>	Varía mucho	Minoría
<b>Situada en un contexto</b>	Varía según la revista (de poco común a casi siempre)	Poco común
<b>Ganancia de información</b>	Mayoría	Raro
<b>Pragmatismo</b>	Raro	Raro
<b>Centrada en el paciente</b>	Raro / poco común	Raro
<b>Le da valor a la inversión</b>	Desconocido, rara vez evaluado	Desconocido, rara vez evaluado
<b>Factible</b>	Casi siempre	La mayoría
<b>Transparente</b>	Rara (datos compartidos)**, casi siempre (registro de ensayos clínicos), poco común (registro de otros estudios)	Raro, poco común, excepto por el registro de ensayos clínicos (aún una minoría)

\* Raro: < 1% de los estudios; poco común: satisfecho en 1 a 20% de los estudios; minoría: satisfecho en el 20 al 50% de los estudios; mayoría: satisfecha en 50 a 80% de los estudios; muy común: satisfecha en el 80 al 99% de los estudios; casi siempre: satisfecha en > 99% de los estudios. Para la evidencia que soporte estos estimados se debe consultar el estudio original

\*\* La situación está mejorando en los últimos años, al menos en los ensayos clínicos

doi:10.1371/journal.pmed.1002049.t002

The Lancet requiere una colocación rutinaria y sistemática de la información en contexto para los ensayos clínicos. Cada vez más revistas exigen el protocolo completo de todo ensayo clínico que van a publicar. La verdad es que los valores que hacen a una investigación útil permanecen subóptimos en la investigación clínica publicada, aún en grandes revistas.

Como se observa en la tabla de arriba, las investigaciones publicadas fallan en casi todas las características que las harían útiles, con excepción quizás de la factibilidad. Estudios que satisfagan todas las características pueden ser descritos como de una rareza extrema, incluso en las revistas más conocidas y exigentes y selectivas.

## Mejorando la situación

El problema de la investigación clínica carente de utilidad no debe ser visto como un ataque contra los investigadores, sino como una oportunidad de mejorar. El esfuerzo tendiente a resolver los problemas debe ser hecho por todas las partes involucradas, incluyendo el público y los pacientes, con la idea de adoptar prácticas más sanas de manera amplia y exitosa.

### Fuerza de trabajo investigativa y médicos

El número de personas que trabajan en investigación clínica es enorme: millones de individuos han sido coautores de al menos un artículo de la literatura biomédica y la mayoría lo han hecho solo una vez. Se espera que estudiantes, residentes y fellows hagan algo de investigación. Esta exposición al ambiente investigativo puede ser interesante, sin embargo, los participantes son juzgados por su capacidad para producir publicaciones rápidamente y este criterio le hace un flaco servicio a la producción de estudios grandes, de largo plazo y realizados en equipo, muy necesarios para que obtengamos conocimientos importantes sobre salud, enfermedad y cuidado médico. Algunos de estos investigadores suelen ser explotados como recurso humano de bajo costo (o gratuito, pues muchos trabajan como voluntarios) y, lo que sucede, es que un personal no entrenado y no comprometido, difícilmente puede producir investigación de calidad.

Otra receta perversa en investigación clínica es que las universidades y otras instituciones suelen presionar para que su personal produzca publicaciones, en lugar de incentivar investigación de utilidad e impacto clínico real. El criterio de utilidad no hace parte de la métrica con la que se evalúan los artículos y el desempeño académico. En lugar de tratar de convertir a todos los médicos en investigadores prolíficos, deberíamos capacitarlos en la comprensión de los métodos de investigación y la MBE, con la idea de inculcarles un escepticismo saludable y habilidades de pensamiento crítico.

### El dipolo industria-reguladores y los socios académicos

La industria y los reguladores suelen ser un dipolo estrechamente conectado a la hora de obtener licencias para medicamentos y dispositivos. La industria responde a los requerimientos de los reguladores, que actúan a la vez como guardianes del bien común y facilitadores de la industria. Esto crea cierta tensión pues hay algo de ambigüedad en la misión. La industria debería poder defender la investigación clínica útil ante unos reguladores que combinen las recompensas comerciales con la utilidad clínica de los productos, ayudando así a que las buenas compañías superen a las malas y alineando los intereses de los accionistas con los de la sociedad. En este orden de ideas, las agencias reguladoras deberían asumir un papel más enérgico promoviendo la realización de estudios grandes y clínicamente útiles.

El sistema de patrocinio actual de la investigación incentiva los estudios pequeños y de corta duración que pueden generar datos publicables en el corto plazo, mientras que resolver preguntas importantes suele requerir estudios de largo plazo cuyas necesidades financieras

suelen exceder los recursos de la mayoría de los ciclos de investigación disponibles. Las alianzas entre las iniciativas de investigación centradas en el paciente y la academia pueden ayudar a resolver los desafíos que se presentan al diseñar e implementar estudios más pragmáticos (aplicables en la vida real). Hay que reconocer que, incluso en los estudios que cuidan y racionalizan sus costos, el dinero que hay que gastar es muy importante cuando muchos estudios solicitan fondos públicos. La industria puede participar aportando recursos a un fondo común controlado por el estado, con la idea de hacer estudios con investigadores académicos libres de conflictos de interés. Con solo desviar 1% o 2% de las ventas de un medicamento exitoso hacia este tipo de fondos se podría obtener amplia financiación para muchas iniciativas importantes.

#### Agenda de financiación para investigación en ciencias básicas, investigación preclínica y clínica

La investigación que simplemente pretende descubrir cosas, sin entregables especificados con antelación (blue-sky research), es muy importante y debería recibir recursos públicos. Resulta que mucha investigación básica cuenta con entregables predefinidos, como cuando se investiga con la idea de desarrollar una nueva droga o una nueva prueba diagnóstica. Este tipo de investigación básica podría ser financiado por la industria o por aquellos que quieran obtener un beneficio económico cuando se entregue un producto que resulta efectivo. Una buena parte de los recursos públicos podrían usarse en investigación clínica útil, en especial cuando no hay una patente que proteja un producto y por lo tanto no habría una razón comercial para que la industria gaste su dinero en estudios que, mejorarían el cuidado médico, pero no representarían utilidades para los accionistas.

#### Revistas científicas

Las revistas pueden influir mucho a la hora de establecer estándares que definan la investigación considerada aceptable. Algunos grupos externos también podrían evaluar la utilidad clínica de los artículos que se publican en las revistas, estableciendo algún tipo de puntaje de utilidad clínica de una revista, teniendo en cuenta los factores mencionados arriba.

#### Pacientes y grupos para la defensa de los pacientes

Los pacientes y las asociaciones que los representan son los que más ganan con un aumento de la utilidad clínica de la investigación médica. Pueden influir mucho en la medida en que se hagan conscientes de los métodos de la ciencia y ayuden a protegerla de influencias sesgadas. Los medios de comunicación tienen un papel importante que desempeñar en la cruzada por obtener investigación útil para la gente.

#### **Conclusiones (Juan José Arango)**

Los que de alguna manera presenciemos el desarrollo de la MBE hemos sido testigos del rumbo que ha tomado la investigación médica en el mundo, al punto de llegar a desconfiar de sus métodos, sus estrategias de financiación, sus objetivos reales y sus conclusiones. La llegada de

científicos con la claridad mental, la inteligencia, la preparación y la lucidez de John P. A. Ioannidis<sup>39</sup> y la creación de iniciativas como METRICS<sup>9</sup>, PCORI<sup>32</sup>, PLOS<sup>40</sup>, etc., nos renuevan el entusiasmo y la fe en lo que hacemos y en lo que leemos.

Es mi opinión que estudiantes, médicos, residentes, investigadores, instituciones académicas y el público en general deberían poner mucha atención a las publicaciones de Ioannidis y sus colegas, si realmente queremos que la investigación médica esté dirigida a la búsqueda del bienestar de los enfermos, sin perjuicio de la industria y los gobiernos, actores fundamentales de todo este proceso.

Porque es cierto que la investigación clínica no está funcionando bien: **“la mayoría de los hallazgos publicados son falsos y, además, inútiles<sup>1,2</sup>.”**

## Referencias

1. Ioannidis JP. Why most published research findings are false. *PLoS Med.* 2005;2(8):e124.
2. Ioannidis JP. Why Most Clinical Research Is Not Useful. *PLoS Med.* 2016;13(6):e1002049.
3. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *Jama.* 1993;270(17):2093-2095.
4. Estruch R, Ros E, Salas-Salvadó J, et al. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet Supplemented with Extra-Virgin Olive Oil or Nuts. *N Engl J Med.* 2018;378(25):e34.
5. predimed.es/uploads/8/0/5/1/8051451/\_1estr\_protocol\_olf.  
[http://www.predimed.es/uploads/8/0/5/1/8051451/\\_1estr\\_protocol\\_olf.pdf](http://www.predimed.es/uploads/8/0/5/1/8051451/_1estr_protocol_olf.pdf).
6. Estruch R, Ros E, Salas-Salvadó J, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med.* 2013;368(14):1279-1290.
7. Eric Topol | Medscape Editor in Chief. 2018; <https://www.medscape.com/author/eric-topol>.
8. Ioannidis: Most Research Is Flawed; Let's Fix It. 2018.
9. Meta-Research Innovation Center at Stanford | METRICS. 2017; <https://metrics.stanford.edu/>.
10. John Ioannidis on Statistical Significance, Economics, and Replication from EconTalk. 2018; <http://www.stitcher.com/s?eid=52979187>.
11. Ioannidis JP. How to make more published research true. *PLoS Med.* 2014;11(10):e1001747.
12. Armstrong RA. When to use the Bonferroni correction. *Ophthalmic Physiol Opt.* 2014;34(5):502-508.
13. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004;291(20):2457-2465.
14. de Winter J, Happee R. Proteus phenomenon - Wikipedia. *PLoS ONE* 2013-06-20; [https://en.wikipedia.org/wiki/Proteus\\_phenomenon](https://en.wikipedia.org/wiki/Proteus_phenomenon). Accessed 6, 8.
15. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst.* 2004;96(6):434-442.
16. Data dredging - Wikipedia. 2018; [https://en.wikipedia.org/wiki/Data\\_dredging](https://en.wikipedia.org/wiki/Data_dredging).
17. Boyack KW, Klavans R, Sorensen AA, Ioannidis JP. A list of highly influential biomedical researchers, 1996-2011. *Eur J Clin Invest.* 2013;43(12):1339-1365.

18. Contopoulos-Ioannidis DG, Alexiou GA, Gouvias TC, Ioannidis JP. Medicine. Life cycle of translational research for medical interventions. *Science*. 2008;321(5894):1298-1299.
19. Bandwagon effect - Wikipedia. 2018; [https://en.wikipedia.org/wiki/Bandwagon\\_effect](https://en.wikipedia.org/wiki/Bandwagon_effect).
20. Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383(9912):101-104.
21. Wenneras C, Wold A. Nepotism and sexism in peer-review. *Nature*. 1997;387(6631):341-343.
22. Greenhalgh T, Howick J, Maskrey N, Group EBMR. Evidence based medicine: a movement in crisis? *BMJ*. 2014;348:g3725.
23. Stamatakis E, Weiler R, Ioannidis JP. Undue industry influences that distort healthcare research, strategy, expenditure and practice: a review. *Eur J Clin Invest*. 2013;43(5):469-475.
24. Christakis DA, Zimmerman FJ. Rethinking reanalysis. *JAMA*. 2013;310(23):2499-2500.
25. Witten DM, Tibshirani R. Scientific research in the age of omics: the good, the bad, and the sloppy. *J Am Med Inform Assoc*. 2013;20(1):125-127.
26. Ioannidis JP, Khoury MJ. Assessing value in biomedical research: the PQRST of appraisal and reward. *JAMA*. 2014;312(5):483-484.
27. Moynihan R, Doran E, Henry D. Disease mongering is now part of the global health debate. *PLoS Med*. 2008;5(5):e106.
28. Evangelou E, Siontis KC, Pfeiffer T, Ioannidis JP. Perceived information gain from randomized trials correlates with publication in high-impact factor journals. *J Clin Epidemiol*. 2012;65(12):1274-1281.
29. Clarke M, Chalmers I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents? *JAMA*. 1998;280(3):280-282.
30. Ferreira-González I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ*. 2007;334(7597):786.
31. Selby JV, Lipstein SH. PCORI at 3 years--progress, lessons, and plans. *N Engl J Med*. 2014;370(7):592-595.
32. PCORI. 2018; <https://www.pcori.org/>.
33. Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. *Med Decis Making*. 2011;31(6):E1-E22.
34. Minelli C, Baio G. Value of Information: A Tool to Improve Research Prioritization and Reduce Waste. *PLoS Med*. 2015;12(9):e1001882.
35. Eisenstein EL, Collins R, Cracknell BS, et al. Sensible approaches for reducing clinical trial costs. *Clin Trials*. 2008;5(1):75-84.
36. Kasenda B, von Elm E, You J, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA*. 2014;311(10):1045-1051.
37. Chapman SJ, Shelton B, Mahmood H, Fitzgerald JE, Harrison EM, Bhangu A. Discontinuation and non-publication of surgical randomised controlled trials: observational study. *BMJ*. 2014;349:g6870.
38. Siontis KC, Evangelou E, Ioannidis JP. Magnitude of effects in clinical trials published in high-impact general medical journals. *Int J Epidemiol*. 2011;40(5):1280-1291.
39. John Ioannidis - Wikipedia Republished // WIKI 2. 2018; [https://wiki2.org/en/John\\_Ioannidis](https://wiki2.org/en/John_Ioannidis).
40. PLOS | Public Library of Science. 2018; <https://www.plos.org/>.