



SERIES EN MEDICINA INTENSIVA: ACTUALIZACIÓN EN METODOLOGÍA EN MEDICINA INTENSIVA

Errores metodológicos frecuentes en la investigación clínica



L.C. Silva Aycaguer

Escuela Nacional de Salud Pública de Cuba, Ciudad de La Habana, Cuba

Recibido el 2 de noviembre de 2017; aceptado el 25 de diciembre de 2017

Disponible en Internet el 7 de marzo de 2018

PALABRAS CLAVE

Ensayos clínicos;
Pruebas de significación estadística;
Fraude;
Sesgo de selección;
Tamaño muestral;
Revistas depredadoras

Resumen Se enumeran, discuten e ilustran diversos errores que se presentan con frecuencia en la investigación clínica. Se hace una distinción entre lo que puede considerarse un «error» surgido de la ignorancia o el descuido, de aquello que dimana de una falta de integridad de los investigadores, aunque se reconoce y documenta que no es fácil establecer cuándo estamos en un caso y cuándo en otro. El trabajo no se propone hacer un inventario exhaustivo de tales pifias, sino que se concentra en aquellas que, sin dejar de ser frecuentes, suelen ser menos evidentes o menos señaladas en los diversos listados que se han publicado con este tipo de problemas. Se ha optado por desarrollar en detalle los ejemplos que ilustran los problemas señalados en lugar de hacer una relación de errores acompañados de una descripción epidérmica de sus características.

© 2018 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

KEYWORDS

Clinical trials;
Statistical significance tests;
Fraud;
Selection bias;
Sample size;
Predatory journals

Frequent methodological errors in clinical research

Abstract Several errors that are frequently present in clinical research are listed, discussed and illustrated. A distinction is made between what can be considered an “error” arising from ignorance or neglect, from what stems from a lack of integrity of researchers, although it is recognized and documented that it is not easy to establish when we are in a case and when in another. The work does not intend to make an exhaustive inventory of such problems, but focuses on those that, while frequent, are usually less evident or less marked in the various lists that have been published with this type of problems. It has been a decision to develop in detail the examples that illustrate the problems identified, instead of making a list of errors accompanied by an epidermal description of their characteristics.

© 2018 Elsevier España, S.L.U. y SEMICYUC. All rights reserved.

Correo electrónico: lcsilva@infomed.sld.cu

<https://doi.org/10.1016/j.medin.2017.12.012>

0210-5691/© 2018 Elsevier España, S.L.U. y SEMICYUC. Todos los derechos reservados.

Introducción

A comienzos del siglo XXI, el epidemiólogo y estadístico de origen griego, profesor de la Universidad de Stanford, John Ioannidis, publicó un artículo¹ que da por sentado que los hallazgos de la investigación contemporánea en salud arriban en su mayoría a conclusiones incorrectas. Un artículo destinado a fundamentar una afirmación de ese calibre estaba llamado a sacudir a la comunidad científica y a concitar un enorme interés. En efecto, como el propio investigador testimonia en una breve entrevista difundida en YouTube (<https://www.youtube.com/watch?v=KOZAV9AvIQE>), mucho más de un millón de personas lo han leído. De hecho, se trata del artículo que más veces ha sido consultado en la prestigiosa revista de acceso abierto *PLOS Medicine* desde su fundación en 2004.

Once años después, Ioannidis acaba de publicar en la misma revista un artículo² complementario, tanto o más inquietante. En esta contribución, titulada «¿Por qué la mayor parte de la investigación clínica no es útil?» se precisa que: «En general, no solo la mayoría de los resultados de investigaciones son falsos, sino que la situación es peor: la mayor parte de los resultados verdaderos no son útiles».

Las causas del problema son numerosas, pero es obvio que, en principio, ni autores ni editores desean cometer errores ni hacer esfuerzos inútiles. Esta sombría realidad legitima que la revista *MEDICINAINTENSIVA* promueva un análisis de los errores que con frecuencia aquejan a los estudios clínicos o a su difusión. La presente contribución procura pasar revista a algunos de ellos.

La apropiación de resultados ajenos, las omisiones deliberadas o las interpretaciones manipuladoras no pueden considerarse «errores»; son simples fechorías que merecen repudio directo. La historia reciente de la investigación clínica está, lamentablemente, plagada de triquiñuelas de todo tipo para conseguir que determinados procedimientos tengan éxito mercantil aunque sean inútiles o, incluso, dañinos. Las evidencias son muchas e incontrovertibles; acaso el más notable prontuario lo ofrezca el libro de Peter Gøtzsche³, director del *Nordic Cochrane Center*, titulado «Medicamentos que matan y crimen organizado», recientemente publicado en España.

De hecho, de un total de 2.047 artículos científicos que fueron descalificados luego de su publicación, según se registra en PubMed (los llamados «retractions») en los últimos 4 decenios, solo la quinta parte fueron debidos a errores⁴; el resto se debió a fraudes o sospecha de tales.

En un estudio de 2015 que abarcó 57 ensayos clínicos controlados realizados entre 1998 y 2013 se da cuenta de que la *Food and Drug Administration* había descubierto que 22 (casi el 40%) daban información falsa, que 35 (61%) tenían registros inadecuados o inexactos, uno de cada 4 fallaba en la notificación de reacciones adversas, 3 de cada 4 habían violado el protocolo y más de la mitad fallaban en la protección de los pacientes (problemas de seguridad o del consentimiento informado)⁵. Las publicaciones correspondientes no mencionaban estas expresiones de mala praxis y lo cierto es que resulta difícil dirimir cuándo ello se debió a decisiones deliberadas o no.

Entendemos que cuando los editores de *MEDICINAINTENSIVA* decidieron destinar un espacio a los errores que suelen

cometerse no se interesaban en alertar acerca de la fabricación o falsificación de datos o el plagio, sino más bien en llamar la atención sobre yerros que pueden ocasionalmente dimanar de la desactualización o el descuido. Sin embargo, ha de señalarse que la frontera entre los errores y las conductas fraudulentas es delgada.

Siendo así, en lo que sigue señalaremos como errores algunas prácticas que bien pueden estar siendo implementadas con finalidad espuria, aunque no necesariamente ese haya sido el caso. De todos modos, sería útil para los profesionales de la medicina intensiva que evitaran pifias que pudieran ser consideradas estafas académicas, una duda que no siempre se podrá disipar inequívocamente⁶. Por esa razón, algunos de los ejemplos que el lector hallará en este artículo pueden corresponder tanto a supercherías como a inocentes consecuencias de la ignorancia o la inadvertencia de normas exigidas por las buenas prácticas investigativas⁷.

La idea de este artículo, en cualquier caso, no es hacer un inventario exhaustivo de posibles errores. La literatura recoge diversos inventarios de los más conocidos⁸⁻¹²; la más exhaustiva (que describe una cincuentena de errores comunes) es la debida a Strasak et al.¹³. Puesto que tales fuentes pueden ser consultadas fácilmente, aquí intentaremos pasar revista no a los más reiterados, sino a los menos evidenciados y más controversiales, aunque no por ello infrecuentes. Además, se ha optado por desarrollar con cierto detalle cada problema antes que tratar muchos de manera epidérmica.

A continuación se enuncian y comentan 6 errores que cumplen la condición señalada y que yo he elegido por su trascendencia y por su pertinaz aparición en las publicaciones.

Irrelevancia clínica e irresponsabilidad social

Un estudio¹⁴ realizado a partir de los 704 artículos enviados en el bienio 2000-2001 a 2 revistas del más alto nivel (*British Medical Journal* y *Annals of Internal Medicine*) con vistas a su posible publicación arroja que 514 (el 73%) contaron con la ayuda de un experto en metodología y estadística. La tasa de rechazos fue del 71% para los que no tuvieron dicho apoyo y solo del 57% para quienes contaron con él. Tales datos parecen poner de manifiesto el provecho de tener a un estadístico entre los autores.

Desafortunadamente, en no pocos profesionales de la salud se ha ido cincelandando la absurda convicción de que cualquier investigador puede (y, peor incluso, que cualquiera debe) dominar la estadística, como si se tratara de la gramática básica. Sin embargo, el dominio genuino de las técnicas estadísticas exige de conocimientos altamente especializados. Quizás ello explique una mala praxis extendida en esta materia que merece especial atención: el empleo erróneo de las pruebas de significación estadística (PSE)¹⁵.

Las críticas a los «valores p» son varias y de diferente naturaleza, y los trabajos destinados a fundamentarlas son muy numerosos. Hace ya decenios, el *Comité Internacional de Directores de Revistas Médicas*¹⁶ reniega explícitamente de que los «valores p» sean empleados desdeñando la estimación de los efectos. Para profundizar en los efectos nefastos que supone desoír esta recomendación puede acudir a las actuales declaraciones de 72

destacadísimos investigadores de distintos campos¹⁷, así como a un detallado análisis recientemente publicado sobre el tema¹⁸.

Nos concentraremos en la consecuencia más grave del empleo acrítico y adocenado de las PSE: a su amparo, de una manera generalizada y sistemática, se atribuye trascendencia clínica (o para la salud pública) a procedimientos terapéuticos o preventivos que están muy lejos de tenerla.

Imaginemos que se comparan las tasas de recuperación para cierta enfermedad tras la administración de sendos tratamientos (por ejemplo, uno convencional y otro novedoso). Supongamos que al aplicar una prueba de hipótesis se obtiene un «valor p» inferior al sacralizado 0,05. Lo que se podría afirmar es que tenemos indicios fundados para creer que los 2 tratamientos no son iguales. Ni más ni menos. Sin embargo, se ha generalizado la costumbre de suplir esta declaración por la afirmación de que la «diferencia entre los tratamientos es significativa». Es un abuso de lenguaje cuya envergadura radica en que, implícitamente, se está diciendo que «la diferencia es trascendente o sustantiva».

Para ilustrar que no estamos ante un mero problema semántico, así como la incapacidad de los «valores p» para convertir efectos triviales en recomendaciones genéricas que exaltan falsas trascendencias, detengámonos en un ejemplo especialmente expresivo. Se trata de un ensayo clínico publicado en *The Lancet* sobre el valor preventivo de 2 inhibidores de la agregación plaquetaria: Plavix® (clopidogrel) y Aspirina® (ácido acetil salicílico)¹⁹.

Allí se «prueba» que el Plavix® (comercializado por Sanofi, laboratorio que financió el estudio) es más eficaz que la Aspirina® para reducir la aparición de eventos vasculares graves en población con alto riesgo de padecerlos.

La población objeto de estudio fue la de aquellos sujetos con historia reciente de infarto de miocardio dentro de los 35 días anteriores, o con antecedentes de un accidente cerebrovascular isquémico durante los 6 meses anteriores. Tras una asignación aleatoria, los 2 tratamientos se aplicaron a unas 20.000 personas con los rasgos mencionados. El *end-point* (desenlace a evaluar) fue la ocurrencia de un nuevo accidente cerebrovascular isquémico o un nuevo infarto al miocardio, fatales o no, o cualquier muerte de origen vascular.

El megaestudio permitió estimar que el riesgo ascendió a D1=583 eventos por cada 10.000 personas-año de tratamiento con Aspirina® y a D2=532 por cada 10.000 personas-año cuando se empleó Plavix®. Tras aplicar una PSE, el estudio arriba triunfalmente a un valor $p=0,043$ y, a partir de ello, se proclama que el Plavix® tiene mayor capacidad preventiva que la Aspirina®.

Salta a la vista que la colectividad de pacientes de alto riesgo se ahorraría apenas D1 – D2 = 51 eventos por cada 10.000 personas-año de tratamiento preventivo si este se basa en Plavix® en lugar de en Aspirina®.

Consecuentemente, en principio un decisor racional, suponiendo que se base en este estudio, se abstendría de recomendar el Plavix® (un medicamento mucho más caro). Pero incluso si alguien considerara que tal diferencia es «importante», lo que está fuera de discusión es que el valor de p no participa ni tendrá jamás sentido que participe en la recomendación que se adopte. Cualquier decisión se puede (y se debe) tomar sobre la base de los valores de D1 y D2, acompañados de otros datos tales como el costo y los efectos

adversos. El único papel que desempeñó la p en el artículo (y luego en sus numerosas citas) es dar cobertura a quienes comercializan el Plavix®, aprovechando la inercia acrítica en el uso de la estadística y la sumisión a las PSE, así como en el error (o truco²⁰) de escamotear el significado real en términos prácticos de aquello que recomiendan.

No se trata de un ejemplo aislado; esta «dictadura» de las PSE conserva plena vigencia en la literatura biomédica. Inculpándolas, el profesor finlandés Esa Läärä escribe: «La comunicación de “resultados” triviales y sin sentido que no proveen información cuantitativa adecuada que tenga interés científico es masiva»²¹. Un par de ejemplos bastan para convalidar tal afirmación: un análisis de 71 fármacos anticancerosos consecutivamente aprobados para el tratamiento de pacientes con tumores sólidos entre 2002 y 2012²² arrojó que la supervivencia global y el intervalo libre de enfermedad aportados por dichos fármacos ascendía a 2,1 y 2,3 meses, respectivamente.

Los estándares de la *American Society of Clinical Oncology* para considerar que el fármaco producía un «beneficio clínico significativo» nada tienen que ver con la «significación estadística». Entre abril de 2014 y febrero de 2016 la *Food and Drug Administration* aprobó 47 antineoplásicos; un estudio recientemente publicado en una revista especializada²³ puso al descubierto que solo 9 (19%) cumplan con dichos estándares.

Generalizaciones inaceptables

Los resultados de un ensayo clínico no se pueden aplicar precipitadamente a personas que caen fuera del ámbito de la muestra empleada para realizarlo. Todo estudio ha de tener unos criterios de inclusión (aquellos que indican qué sujetos se pueden seleccionar para integrar la muestra que representa a la población que se estudia) y unos criterios de exclusión (rasgos de quienes no se habrán de incorporar al ensayo). Por ejemplo, se establece que se estudiarán adultos no asmáticos y que las mujeres no deben estar embarazadas. Si el tratamiento fue exitoso para determinada dolencia y cierto fármaco luego de un mes de aplicación, entonces no se debe recomendar su empleo en embarazadas ni asmáticos, ni el fármaco debe aplicarse durante otro lapso que no sea un mes (ni más largo ni más corto).

Obviamente, la manera en que se aplica un resultado no es responsabilidad de quienes lo obtuvieron. Sin embargo, en no pocas ocasiones se comete el error de omitir en los resúmenes y las conclusiones de los estudios toda referencia a esas restricciones. Y como estas secciones suelen ser las que se leen con más detenimiento, lo que suele llegar al lector es una invitación implícita a que se apliquen los resultados más allá de lo legítimo.

Sesgo de selección: la tragedia del Vioxx®

El rofecoxib, poderoso analgésico comercializado por Merck bajo el nombre de Vioxx® y que fue usado por unos 84 millones de personas de todo el mundo, salió a la venta en 1999 en más de 80 países. Se usaba para tratar los síntomas de la osteoartritis y la artritis reumatoide, el dolor agudo y las menstruaciones dolorosas. En el año 2000, la revista *The New England Journal of Medicine* se hizo eco del estudio

VIGOR, en el que habían participado 8.076 pacientes, la mitad de los cuales (4.029) fueron asignados aleatoriamente al grupo que recibiría 500 mg de naproxeno (antiinflamatorio no esteroideo de vieja data) 2 veces al día, en tanto que al resto (4.074) se le administrarían 50 mg de rofecoxib una vez al día²⁴.

El *endpoint* era cualquier efecto adverso gastrointestinal, precisamente porque la supuesta gran ventaja del medicamento era que permitía tratar el dolor de enfermedades crónicas sin propiciar úlceras gástricas sintomáticas ni hemorragias gastrointestinales, efectos indeseables propios de los analgésicos clásicos. El resultado fue altamente favorable al Vioxx[®].

Según se supo mucho después, la vicepresidenta de investigación clínica de Merck, y firmante del estudio citado, había propuesto en un documento interno de la empresa en 1997 que del ensayo clínico que planeaban realizar se excluyese a las personas con problemas cardiovasculares previos «para que las diferencias de complicaciones entre los que tomasen Vioxx[®] y los que recibiesen un analgésico clásico no fuesen evidentes». Es decir, Merck no solo conocía con anterioridad la posible toxicidad miocárdica, sino que planificó un sesgo en la selección para que dicha complicación no se pusiera de manifiesto en el ensayo.

La letal influencia en ataques cardíacos, según estimaba David Gram, investigador de la *Food and Drug Administration*, en una entrevista con un diario británico²⁵, produjo «la mayor catástrofe debida a medicamentos en la historia: entre 89.000 y 139.000 estadounidenses podrían haber fallecido o tenido trastornos cardíacos por concepto del fármaco». Ante esta situación, el 30 de septiembre de 2004 Merck anunciaba la retirada en todo el mundo de su medicamento.

En este doloroso y aleccionador episodio, la empresa planificó y consumió el sesgo de selección (una de las fechorías relacionadas con el Vioxx[®] por las cuales la empresa fue luego millonariamente multada). Pero lo que se quiere subrayar es que cuando dicho sesgo no se evita (aunque sea por mera inadvertencia) se está cometiendo un error que puede llegar a ser muy grave.

Mitos y equívocos sobre los tamaños de muestra necesarios

Es crucial comprender algo que suele ser olvidado por quienes teorizan en torno al tamaño muestral y se adhieren a posiciones ortodoxas y muchas veces esquemáticas y rituales^{26,27}. Se trata de que, en última instancia, virtualmente sea cual sea el tamaño muestral, el estudio es potencialmente útil. La información disponible para determinar con precisión un tamaño muestral correcto (un calificativo polémico de por sí) es a menudo inexistente o, como mínimo, mucho menor que la necesaria; y en cualquier caso, estará inexorablemente atravesada por la especulación inherente a la elección de los datos que han de emplearse para el cálculo²⁸.

Ocasionalmente, se hacen juicios según los cuales «el tamaño de la muestra es insuficiente». Este sigue siendo un criterio valorativo muy consolidado en comités de ética, agencias de evaluación de proyectos de investigación y árbitros o editores de revistas. La legitimidad de este tipo de

pronunciamientos es, sin embargo, muy discutible. Tales juicios son conflictivos y casi siempre improcedentes por al menos 2 razones.

En primer lugar, porque las fórmulas para determinar el tamaño muestral son intrínsecamente especulativas y comportan una inevitable carga de subjetividad²⁹.

En segundo lugar, porque la noción de «tamaño suficientemente grande» solo tendría sentido en el supuesto de que pudieran derivarse reglas operativas o conclusiones definitivas de cada estudio aislado. Sin embargo, esta es una ilusión tan arraigada como incorrecta, por la simple razón de que la ciencia no funciona así^{30,31}.

Nuestras convicciones científicas pueden ser más o menos firmes, pero siempre son provisionales, y nuestras representaciones de la realidad tienen en cada momento un cierto grado de credibilidad, pero están abiertas a cambios y perfeccionamientos en la medida en que nuevos datos lo aconsejen. La consolidación del nuevo conocimiento es gradual y cualquier aporte es bienvenido. Unos serán más trascendentes y otros menos. Pero todos pueden hacer alguna contribución en este proceso, independientemente del tamaño muestral, generalmente por conducto del metaanálisis³².

Los errores en la investigación clínica no solo los cometen quienes realizan las investigaciones, sino también quienes difunden sus resultados. Este es un ejemplo: constituye a mi juicio un error exigir que los autores incluyan las fórmulas usadas (y, en general, que expliquen sus tamaños muestrales). Por lo general, lo que se consigue es que ellos apliquen lo que en un destacado artículo publicado en *The Lancet*³³ llaman *retrofitting* (hacer la elección de los valores que conducen a un tamaño muestral elegido de antemano) y, por ende, forzar a incurrir en cierta falta de integridad por parte de los autores. El problema es conflictivo, ya que la idea de que es menester explicar los tamaños muestrales está muy arraigada, a pesar de ser, simplemente, irracional³⁴.

El iceberg de los «estudios negativos»

Recientemente se ha fundado una curiosa revista llamada *Negative Results*³⁵ con el propósito de publicar trabajos que niegan o desafían conocimientos dados por ciertos, sea para restarles apoyo o para refutarlos. La iniciativa reacciona contra el estigma que padecen los llamados «hallazgos negativos».

El llamado «sesgo de publicación», así denominado por primera vez en un entorno de 1980³⁶, consiste en la tendencia a favorecer la publicación de ciertos trabajos en función de los resultados obtenidos y no necesariamente de sus méritos. La razón fundamental de su existencia es la negativa de muchos editores de revistas médicas a publicar aquello que no constituya un aporte «novedoso», como si el rechazo a una arraigada convicción errónea no lo fuera.

Desde entonces se ha venido tomando creciente conciencia sobre el problema, conocido como el *file drawer problem*³⁷: las revistas se llenan con una pequeña fracción de los estudios que se realizan (útiles o no), mientras que las gavetas de los laboratorios se llenan con la mayoría restante de los trabajos (aquellos con resultados negativos, vale decir, estudios para los que no se obtuvo *significación*). Como se dice en la presentación de la mencionada revista:

«los hallazgos científicos son como un iceberg, este flota alrededor del 10% de los descubrimientos publicados sobre un 90% de resultados negativos».

El más grave problema que trae consigo este sesgo, denunciado una y otra vez³⁸, es que distorsiona la visión que tenemos de la realidad y compromete, por tanto, la capacidad de adoptar decisiones atinadas. Cada trabajo puede ser objetivo, pero la mirada global deja de serlo si se suprime una parte de la evidencia. Naturalmente, una de las áreas que se ve más seriamente limitada por este concepto es el metaanálisis, cuyo fin declarado es el de formalizar la identificación del consenso.

Quizás el ejemplo más espectacular en años recientes lo ofrece el manejo de la pandemia de gripe A del 2009. Según se comunica en un excelente trabajo que cuenta los pormenores de este episodio³⁹, como respuesta a la amenaza, se promovió el uso de dos antivirales Tamiflu® (oseltamivir) y Relenza® (zanamivir), de los que se hicieron compras millonarias en euros por parte de numerosos gobiernos.

Tal decisión estuvo respaldada por una revisión de los antivirales realizada en 2006 y basada fundamentalmente en un trabajo de 2003 donde se agruparon 10 estudios (todos financiados por Roche, empresa comercializadora del Tamiflu®) que *demonstraban* su efectividad y seguridad⁴⁰. Sin embargo, surgieron muchas dudas acerca de tales propiedades, y la presión científica y política terminó obligando a las compañías farmacéuticas a compartir los datos brutos de los ensayos clínicos de los antivirales, que habían sido ocultados hasta entonces con el argumento de que constituían un secreto comercial. Primero Glaxo (sobre Relenza®) y unos meses después Roche (Tamiflu®) hicieron llegar tales datos a los autores de una nueva revisión en 2013. El gigantesco esfuerzo de examinar unas 160.000 páginas permitió concluir que los antivirales tenían una eficacia muy modesta en el alivio de los síntomas y ningún impacto para disminuir complicaciones o muertes, con el añadido de producir importantes efectos adversos y sin capacidad de modificar los mecanismos de contagio⁴¹. Del mismo modo, se pudo corroborar que el ocultamiento de numerosos resultados negativos, la práctica de privilegiar la difusión de los hallazgos deseados y de invisibilizar los de signo opuesto produjo conclusiones sesgadas desde el principio.

En relación con la problemática de la supresión de resultados negativos, por su claridad y amenidad, recomiendo enfáticamente al lector que disfrute de la conferencia a cargo del Dr. Ben Goldacre, investigador del *London School of Hygiene and Tropical Medicine*, que puede hallarse en Youtube (http://www.ted.com/talks/ben_goldacre_battling_bad_science?language=en).

Prácticas editoriales depredadoras

Los investigadores recibimos cotidianamente invitaciones para que enviemos artículos a determinadas revistas, presentando muy interesadas en nuestros aportes científicos. Se trata con extrema frecuencia de las llamadas «editoriales y revistas depredadoras»: un modelo lucrativo fraudulento, caracterizado por rasgos tales como prometer una rápida publicación de los artículos, ofrecer tentadores precios, comprometerse a una exigente revisión por pares que en realidad no se produce o es en extremo débil, atribuirse

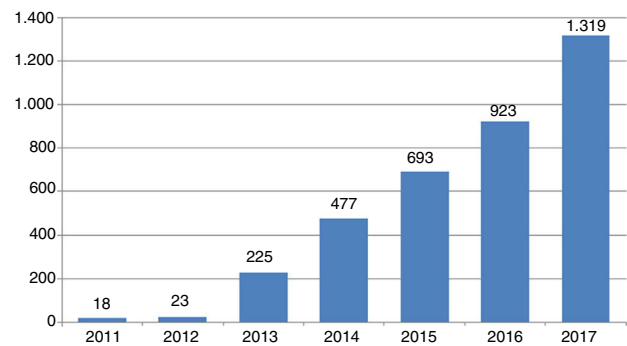


Figura 1 Evolución del número de revistas depredadoras existentes entre 2011 y 2017.

falsos factores de impacto e incluir académicos inexistentes en sus consejos editoriales.

En esta materia se destacó el trabajo de Jeffrey Beall, bibliotecario de la Universidad de Colorado, famoso por haber acuñado el término y por la publicación en el pasado de listas de «potenciales, posibles o probables editoriales y revistas de acceso abierto depredadoras». Curiosamente, en enero de 2017 Beall canceló el sitio de Internet donde actualizaba periódicamente sus listas⁴². Solo informó de que lo hacía por haber sido amenazado. Sin embargo, un equipo que conserva el anonimato, precisamente para no ser coaccionado, ha continuado esa tarea y la lista actualizada (<https://predatoryjournals.com/>) puede consultarla cualquiera.

Para aquilatar la magnitud del fenómeno basta advertir la tendencia que ha experimentado la emergencia de revistas de este tipo desde su aparición en 2011: 18 revistas ese año, 23 al siguiente y a partir de entonces un crecimiento exponencial hasta llegar a 1.319 en octubre de 2017 (fig. 1).

Resulta evidente la nefasta polución informativa que entraña esta plaga académica; pero la razón por la cual se menciona en la presente contribución es porque considero un error (y no pequeño) enviar artículos a tales publicaciones. Hacerlo no solo contribuye a dar alas a una modalidad editorial espuria, sino a comprometer el propio prestigio de quien cae cándidamente en sus redes⁴³.

A modo de resumen

Los errores metodológicos posibles en materia de investigación clínica son muy numerosos y variados. La presente contribución recoge un pequeño grupo de ellos con el propósito de llamar la atención a los especialistas de medicina intensiva sobre desaciertos en los que no siempre se repara. El manejo estadístico impropio, las omisiones de elementos cruciales, las conductas inerciales y el candor a la hora de publicar se han puesto de manifiesto usando ejemplos, no siempre debidos a la candidez, pero igualmente lesivos para la ciencia médica, que vive momentos de crisis desde que John Ioannidis sacudiera a la comunidad científica con las penetrantes contribuciones mencionadas al comienzo del artículo.

Conflicto de intereses

El autor declara no tener ningún conflicto de intereses.

Bibliografía

- Ioannidis JPA. Why most published research findings are false. *PLoS Med.* 2005;2:e124.
- Ioannidis JPA. Why most clinical research is not useful. *PLoS Med.* 2005;13:e1002049.
- Gøtzsche PC. Deadly medicines and organized crime. How big pharma has corrupted health care. London: Radcliffe Publishing Ltd.; 2013.
- Fang F, Steen G, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci U S A.* 2012;109:17028–33.
- Mayor S. Research misconduct found by FDA inspections of clinical trials is rarely reported in journal studies. *BMJ.* 2015;350:h747.
- Nath SB, Marcus SC, Druss BG. Retractions in the research literature: Misconduct or mistakes? *Med J Aust.* 2006;185:152–4.
- Steen RG. Retractions in the scientific literature: Do authors deliberately commit research fraud? *J Med Ethics.* 2011;37:113–7.
- Freemantle N, Rait G. Catalogue of errors in papers reporting clinical trials. *BMJ.* 2015;351:h4843, <http://dx.doi.org/10.1136/bmj.h4843>
- Clark GT, Mulligan R. Fifteen common mistakes encountered in clinical research. *J Prosthodont Res.* 2011;55:1–6.
- Kim MY, Goldberg JD. The effects of outcome misclassification and measurement error on the design and analysis of therapeutic equivalence trials. *Stat Med.* 2000;20:2065–78.
- Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing clinical research. 3rd ed. Philadelphia: Lippincott, Williams and Wilkins; 2007. p. 51–63.
- Henderson WG. Mistakes in clinical trials. En: Itani K, Reda D, editores. Clinical trials design in operative and non operative invasive procedures. Cham: Springer; 2017.
- Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research—A review of common pitfalls. *Swiss Med Wkly.* 2007;137:44–9.
- Altman PG, Goodman NS, Schroter S. How statistical expertise is used in medical research. *JAMA.* 2002;287:2817–20.
- Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *Eur J Epidemiol.* 2016;31:337–50.
- ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals [Internet]. 2016 [consultado Oct 2017]. Disponible en: <http://www.icmje.org/icmje-recommendations.pdf>
- Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nat Hum Behav.* 2018;2:6–10, <http://dx.doi.org/10.1038/s41562-017-0189-z>.
- Silva-Aycaguer LC. Las pruebas de significación estadística: seis décadas de fuegos artificiales. *Rev Fac Nac Salud Publica.* 2016;34:205–12 [consultado Oct 2017]. Disponible en: http://lcsilva.sbhac.net/Articulos/75.Las-PSE_6_decadas_de_fuegos_artificiales_RFSAP.pdf
- CAPRIE Steering Committee. A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events. *Lancet.* 1996;348:1329–39.
- Smith R. The trouble with medical journals. *J R Soc Med.* 2006;99:115–9.
- Läärä E. Statistics: Reasoning on uncertainty, and the insignificance of testing null. *Ann Zool Fennici.* 2009;46:138–57.
- Fojo T, Mailankody S, Lo A. Unintended consequences of expensive cancer therapeutics—The pursuit of marginal indications and a me-too mentality that stifles innovation and creativity: The John Conley Lecture. *JAMA Otolaryngol Head Neck Surg.* 2014;140:1225–36.
- Kumar H, Fojo T, Mailankody S. Guidelines for oncology clinical trials. *JAMA Oncol.* 2016;2:1238–40.
- Bombardier C, Laine L, Reicin A, Shapiro D, Burgos R, Davis B, et al; VIGOR Study Group. Comparison of upper gastrointestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. *N Engl J Med.* 2000;343:1520–8.
- Ortín A. El científico que recomendó retirar Vioxx advierte sobre miles de muertes. *El País*, 4 de enero de 2005.
- Thompson B. In praise of brilliance: Where that praise really belongs. *Am Psychol.* 1998;53:799–800.
- Bacchetti P. Current simple size conventions: Flaws, harms and alternatives. *BMC Med.* 2010;8:17.
- Fernandes-Taylor S, Hyun JK, Reeder RN, Harris AH. Common statistical and research design problems in manuscripts submitted to high-impact medical journals. *BMC Res Notes.* 2011;4:304.
- Silva LC, Benavides A. Apuntes sobre subjetividad y estadística en la investigación. *Rev Cubana Salud Publica.* 2003;29:170–3.
- Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: Caveat lector? *J Gen Intern Med.* 1998;13:753–6.
- Marks HM. Rigorous uncertainty: Why RA Fisher is important. *Int J Epidemiol.* 2003;32:932–7.
- Sackett DL, Cook DJ. Can we learn anything from small trials? *Ann N Y Acad Sci.* 1993;703:25–31.
- Grimes DA, Schulz KF. Descriptive studies: What they can and cannot do. *Lancet.* 2002;359:145–9.
- Silva LC, Alonso P. Explicación del tamaño muestral empleado: una exigencia irracional de las revistas biomédicas. *Gac Sanit.* 2013;27:53–7.
- Negative Results [Internet] [consultado Oct 2017]. Disponible en: <http://www.negative-results.org/index.php/about-our-company/>
- Smith ML. Publication bias and meta-analysis. Evaluation in education. 1980;4:22–4.
- Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychol Bull.* 1979;86:638–41.
- Ioannidis JP. Journals should publish all "null" results and should sparingly publish "positive" results. *Cancer Epidemiol Biomarkers Prev.* 2006;15:186.
- Gérvás J. El asunto Tamiflú/Relenza, la salud pública y algunas lecciones para la decisión y la ética. *Rev Cubana Salud Publica.* 2014;40:334–48.
- Cohen D, Carter P. WHO and the pandemic flu "conspiracies". *BMJ.* 2010;340:e2912.
- Heneghan CJ, Onakpoya I, Thompson M, Spencer EA, Jones M, Jefferson T. Zanamivir for treatment and prevention of influenza in adults and children: Systematic review of clinical study reports. *BMJ.* 2014;348:g2547.
- Straumsheim C. No more 'Beall's List'. Inside Higher Ed. 2017 [consultado Oct 2017]. Disponible en: <https://www.insidehighered.com/news/2017/01/18/librarians-list-predatory-journals-reportedly-removed-due-threats-and-politics>
- Silva LC. Las editoriales depredadoras: una plaga académica. *Rev Cub Inform Cien Sal.* 2016;27 [consultado Oct 2017]. Disponible en: <http://www.acimed.sld.cu/index.php/acimed/article/view/996/608>