



Repositorio Institucional de la Universidad Autónoma de Madrid

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Medicina Clínica 146.9 (2016): 408-412

DOI: 10.1016/j.medcli.2015.11.029

Copyright: © 2015, Elsevier Inc.

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

La reproducibilidad de las investigaciones biomédicas: *Quo vadis?*

Reproducibility of biomedical research: *Quo vadis?*

Rafael Dal-Ré* (a), Antonio Bernad (b), Rafael Garesse (c)

- a) Investigación Clínica, Programa BUC (Biociencias UAM+CSIC), Centro de Excelencia Internacional, Universidad Autónoma de Madrid, Madrid
- b) Departamento de Inmunología y Oncología, Centro Nacional de Biotecnología, CSIC, Madrid
- c) Departamento de Bioquímica, Instituto de Investigaciones Biomédicas “Alberto Sols” UAM-CSIC, Madrid. Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER). Instituto de Investigación Sanitaria 12 de Octubre (i+12). Facultad de Medicina, Universidad Autónoma de Madrid, Madrid.

* Autor para correspondencia. Correo electrónico: rafael.dalre@fuam.uam.es

Nº de palabras: 2827

Nº de tablas: 3

Nº caracteres con espacios: 25105

"Доверять, но проверять" ("Confía, pero verifica"; Proverbio ruso)

La primera causa de fracaso en el desarrollo clínico de nuevos medicamentos es la falta de eficacia. Así, en 2011-2012 el 56% de 105 nuevos productos y nuevas indicaciones para productos ya comercializados que fracasaron entre la fase 2 (ensayos exploratorios de eficacia) y la entrega del dossier de registro a las agencias de regulación, lo fueron por falta de eficacia¹. Este hecho ocurre a pesar de que, lógicamente, los productos que alcanzan su administración a humanos han obtenido resultados positivos en la investigación preclínica. ¿Hay que deducir, por tanto, que la investigación básica y preclínica –vg, la realizada en líneas celulares y en animales de experimentación- es responsable, siquiera en parte, de esta alta tasa de fracasos en el desarrollo clínico de nuevos medicamentos? Así parece, cuando una regla no escrita entre las compañías de capital riesgo señala que al menos el 50% de los estudios preclínicos publicados por investigadores académicos, incluso en las revistas de máximo prestigio, no pueden ser replicados en laboratorios de la industria biofarmacéutica². De hecho, una estrategia habitual entre las compañías de capital riesgo es la de exigir la validación de los resultados publicados, bien por otros laboratorios académicos, bien por compañías privadas ajenas al grupo de la investigación original². Esto, sin embargo, no siempre ocurre y, en ciertos casos, una serie de estudios preclínicos mal diseñados y comunicados pueden haber conducido a la realización de ensayos clínicos, de los que no se obtuvieron resultados positivos pero en los que sí se expuso a los participantes a riesgos y efectos secundarios de forma innecesaria³.

Qué sabemos hoy de la reproducibilidad de las investigaciones en biomedicina

Desde hace siglos, la ciencia se ha corregido así misma. El método científico provee un marco sistemático para la formulación y, mediante la máxima del ensayo y error, el refinamiento de hipótesis. El conjunto se sustenta en la capacidad de los investigadores de reproducir los

resultados de sus experimentos, clave para tener confianza en la veracidad de los resultados obtenidos, y así poder plantear nuevas investigaciones. Un hallazgo que no puede ser replicado pierde credibilidad.

La reproducibilidad se puede referir tanto a la capacidad de reproducir unos resultados – usando los mismos datos originales- por investigadores independientes al grupo de investigadores inicial, como a la generación de nuevos datos que sustenten las conclusiones del experimento inicial. Aquí se entenderá siempre la segunda acepción del término, excepto cuando se indique lo contrario. En todo caso, la variabilidad biológica implica que no se debe esperar que se repliquen todos y cada uno de los resultados de una investigación previa, aunque sí que se confirmen las mismas conclusiones clave de aquella⁴.

Durante los últimos años muchos artículos, en revistas científicas y generalistas⁵, han señalado la falta de reproducibilidad de la investigación biomédica que, sin duda, está relacionada con que se estime que un 85% de aquella se desperdicie⁶. Esto, junto a estudios que demuestran que el número de artículos retractados ha aumentado de forma exponencial en los últimos años⁷, ha contribuido a que la opinión pública haya reducido su nivel de confianza en la ciencia: solo una minoría (el 36%) de los americanos cree que la información procedente de los científicos es exacta y fiable⁸.

La irreproducibilidad de las investigaciones que afecta a los estudios preclínicos tiene especial relevancia ya que, como se ha comentado, los resultados positivos pueden conducir a que se inicie la investigación con seres humanos. Las compañías biofarmacéuticas interesadas en los resultados de una investigación preclínica académica suelen intentar reproducir “el experimento” para validar los resultados publicados. A este respecto dos estudios alarmaron a la comunidad científica. Uno realizado por investigadores de Bayer mostró que obtuvieron unos resultados completamente en consonancia con los publicados en solo el 20-25% de 67 proyectos (47 de oncología)⁹. Posteriormente, investigadores de Amgen intentaron reproducir

los resultados de 53 artículos de alto impacto en oncología y solo lo lograron en 6 (el 11%)¹⁰. El que algunos de estos productos oncológicos alcancen la investigación clínica y se ensayen en pacientes con un tipo de cáncer para el que los productos no han mostrado realmente eficacia preclínica, explicaría parte de los fracasos que se observan en los ensayos clínicos. De hecho, oncología es el área terapéutica con una más alta tasa de fracasos en el desarrollo clínico: los productos oncológicos suponen el 30% de todos los nuevos medicamentos que iniciaron la fase 2 que acaban por fracasar¹. La tasa media de traslación con éxito de los resultados en modelos animales a ensayos clínicos en oncología no alcanza el 8%¹¹.

La investigación con seres humanos, mucho más compleja y costosa que la preclínica, muchas veces, simplemente, no se puede replicar. Piénsese, por ejemplo, en los ensayos clínicos de efectividad comparativa y en los grandes estudios prospectivos de cohorte, que requieren de cientos o miles de participantes estudiados durante tiempos prolongados. A pesar de esto, hay ciertos datos que nos indican que la situación no es mucho mejor a la descrita con la investigación preclínica. Basten dos ejemplos. En ensayos clínicos controlados no se pudo confirmar ninguna de las conclusiones fundamentales de 52 estudios epidemiológicos observacionales¹². Como replicar ensayos clínicos es muy costoso –y no es ético exponer a nuevos pacientes a placebo o a intervenciones que se sabe tienen una relación beneficio/riesgo inferior a otras-, se evalúa la calidad de las publicaciones re-analizando los datos originales. Pues bien, el 35% de re-análisis publicados de ensayos clínicos dieron lugar a diferencias en los resultados respecto de los originalmente publicados, que afectaban a los tipos y número de pacientes que deberían ser tratados¹³. Hay que destacar que, habitualmente, las agencias de regulación exigen –excepto para oncología- la realización de dos (o más) ensayos confirmatorios de eficacia (fase 3) para la autorización de nuevos medicamentos¹⁴; estos 2 ensayos, en la mayoría de los casos son prácticamente idénticos (en criterios de selección, posología y variables de evaluación), lo que viene a ser una replicación simultánea.

En psicología (y psiquiatría) la situación es todavía más preocupante si cabe. Se sabe que la tasa de publicación de estudios con resultados ‘positivos’ (el 92%) se sitúa en el primer puesto de todas las ciencias naturales y sociales¹⁵. En 2011 se inició un singular proceso de replicación por 270 psicólogos de todo el mundo de 100 estudios de psicología publicados en 2008. Primero se publicó el diseño experimental de la replicación de cada estudio –con la participación de los autores del estudio original si así lo estimaron oportuno–, luego se recibieron en abierto las revisiones de expertos de cada estudio, y solo después se llevaron a cabo y se informó públicamente de los resultados¹⁶. Aunque el 97% de los 100 estudios originales presentaban resultados estadísticamente significativos ($p < 0,05$), entre los repetidos solo en el 47% se pudieron replicar los resultados; además, la magnitud del efecto medio entre los estudios replicados fue la mitad de la de los originales¹⁷. Hay que resaltar sin embargo que, como dice Nosek, el impulsor de este macro-ejercicio de replicación, con los datos obtenidos no se puede saber qué trabajo –el original o el replicado– es el que obtuvo los resultados verdaderos¹⁸. Un ejercicio similar para replicar 50 estudios preclínicos de oncología publicados y de alto impacto está ya en marcha¹⁹.

Cuáles son las causas de la falta de reproducibilidad de las investigaciones biomédicas

En la actualidad, se calcula que la tasa de irreproducibilidad acumulada en la investigación preclínica alcanza el 53%²⁰. Entre las principales razones que provocan la falta de reproducibilidad de las investigaciones hay que mencionar a: a) los errores en el diseño del estudio; b) los problemas en los reactivos biológicos y materiales de referencia; y c) los sesgos y errores en el análisis y comunicación de los resultados.

Así, por ejemplo, en 2671 estudios publicados entre 1991 y 2011, pertenecientes a revisiones sistemáticas de la eficacia de fármacos en 8 modelos animales, se observó mejoría en los porcentajes de aleatorización –que pasó del 14% al 42%–, y del enmascaramiento –que

aumentó del 16% al 39%- entre los años señalados, respectivamente; sin embargo, el cálculo de la muestra no varió, pasando del 2% al 1%²¹ (Tabla 1). En 49 meta-análisis (que incluyeron 730 estudios) en neurociencias, que estimaron el tamaño del efecto de interés, se observó que el poder estadístico mediano era de tan solo el 21%: alcanzaba un paupérrimo 8% en estudios de resonancia magnética nuclear, mientras que los estudios de experimentación animal en laberinto acuático (*water maze*) o radial (*radial maze*) se situaba en el 18% y 31%, respectivamente²³. El campo para la comunicación de resultados falsos positivos está más que abonado (Tabla 1). Aunque los modelos animales presentan evidentes diferencias biológicas con el ser humano, hay una clara inflación de publicaciones con conclusiones exageradas o “superoptimistas” en relación a la eficacia de los productos ensayados: el 80% de los tratamientos eficaces y seguros en modelos animales fracasan en los ensayos clínicos²⁵.

Se estima que entre un 16% y un 36% de las líneas celulares utilizadas en investigación oncológica –a veces pobremente caracterizadas- están contaminadas con células de diferentes tipos de cáncer, constituyendo uno de los factores que más pueden influir en la irreproducibilidad de los estudios preclínicos en oncología²⁶. Esto conduce a que los investigadores publiquen inadvertidamente resultados prometedores de agentes oncológicos frente a ciertos tipos de cáncer en líneas celulares que, de hecho, pueden corresponder a otro tipo de neoplasia. Esto se evitaría adquiriendo las líneas celulares de un proveedor fiable y, además, validándolas en el laboratorio del investigador.

Se sabe que los anticuerpos –uno de los reactivos más comúnmente usados en la investigación biológica- pueden ser también una importante fuente de irreproducibilidad en ensayos que, por otra parte, han sido correctamente diseñados y realizados. El uso de anticuerpos pobremente caracterizados, que no están adecuadamente validados o que no son validados por los investigadores antes de realizar la investigación de interés, puede dar lugar a resultados prometedores pero, en última instancia, inútiles al no poderse reproducir. En un

análisis de 49 anticuerpos comerciales, dirigidos a 19 receptores específicos, se constató que la mayoría reconocía más de una proteína, lo que no les hacía fiables para distinguir entre varios receptores; en la misma línea, un estudio mostró que el 25% de 246 anticuerpos utilizados en epigenética no superaron las pruebas de especificidad²⁷.

En la investigación preclínica es frecuente que se introduzcan procedimientos en el análisis no incluidos en el protocolo inicial de forma que se altera el manejo de los datos. Así, no es infrecuente que se realicen análisis intermedios, se excluyan datos *ad hoc*, que la variable principal de evaluación se seleccione *post hoc* y que se lleven a cabo pseudo-replicaciones –en las que se confunden la replicación técnica con la replicación biológica³–.

Es muy conocido el sesgo de publicación y el sesgo de comunicación selectiva de los resultados en todo tipo de investigaciones, incluidos los ensayos clínicos²⁸ –para los que se han propuesto diversas medidas correctoras²⁹. El principal medidor de una carrera académica exitosa es el número de publicaciones en revistas de alto impacto. Esto deviene en que la publicación pase de ser un medio a convertirse en el fin, y que la rapidez necesaria en publicar pueda entorpecer una interpretación equilibrada de los resultados. Los hechos demuestran que esta situación provoca que muchas veces se publiquen resultados ‘positivos’ antes de que se puedan confirmar, reproducir^{22,30}.

Por último, se estima que el 49% de todos los fondos (públicos y privados) destinados a la investigación biomédica en los EEUU se invierte en investigación preclínica (incluyendo la básica), alcanzando los 56.000 millones de dólares/año. Con una tasa acumulada de irreproducibilidad del 50%, se calcula que anualmente se invierten unos 28.000 millones de dólares en estudios que no se pueden replicar²⁰.

Cómo prevenir la publicación de estudios irreproducibles

Para que la comunidad científica recobre todo su prestigio hay que lograr que las investigaciones científicas se realicen de forma íntegra, poniendo en marcha iniciativas y medidas a lo largo de toda la vida de un proyecto de investigación, desde su concepción hasta la comunicación de los resultados^{24,31}. La manera más racional para evitar que se publiquen estudios que luego no serán reproducibles es que se lleven a cabo con el máximo rigor y que las revistas exijan que se describa detalladamente el material y métodos utilizados: ello permitirá demostrar que no se han cometido, entre otros, errores de diseño (tabla 1) y que, si fue posible, el análisis de los datos se realizó de forma enmascarada que permitirá controlar los sesgos (conscientes e inconscientes) que conducen a muchos investigadores a describir lo que quieren encontrar y desatender lo que realmente muestran los datos^{32,33}. Como se verá a continuación, hay que realizar y analizar los estudios de investigación preclínica con los mismos estándares de calidad que se exige a la investigación clínica desde hace muchos años. Así, en la evaluación de un trabajo, Begley³⁰ enfatiza seis aspectos que deben ser descritos en el texto y que se recogen en la tabla 2.

Recientemente, para limitar los sesgos de publicación en la experimentación animal, se ha propuesto el registro de estos estudios antes de su inicio³⁴. Antes, sin embargo, y con el objetivo de mejorar los estándares de publicación de los estudios en animales, se acordó la guía ARRIVE³⁵, que fue respaldada por financiadores y revistas de alto prestigio pero que, desgraciadamente, no surtió el efecto esperado³⁶.

Así las cosas, los Institutos Nacionales de la Salud de los EEUU (NIH) –la agencia financiadora más importante de ciencia preclínica (incluyendo la básica) en el mundo- reconociendo que el problema de irreproducibilidad afecta sobre todo a la investigación preclínica³⁷, han publicado una guía de cómo se deben comunicar este tipo investigaciones, que ha sido respaldada por Sociedades Científicas, *Nature Publishing Group*, *Science Journals*, *PLOS Journals*, *The BMJ* y más de 30 revistas³⁸ (Tabla 3). Como los NIH son conscientes de que en el trasfondo del

problema se encuentra una deficiente formación de los investigadores, se está elaborando un módulo de entrenamiento específico. Además, están evaluando si añadir como requisito la validación independiente de los trabajos preclínicos previos para las solicitudes de ayudas para ensayos clínicos³⁷.

El futuro próximo

Todos los agentes que intervienen en la investigación científica deben contribuir a un cambio de rumbo que, en un tiempo razonable, nos permita afirmar que la gran mayoría de los estudios científicos en biomedicina son correctamente realizados y comunicados, y que son reproducibles. Esta es la esencia del método científico.

Como se ha mencionado, la carrera profesional y académica de los investigadores se fundamenta en publicar con rapidez estudios con resultados positivos, en revistas de alto impacto. Esto permite al investigador tener más fácil acceso a nuevas ayudas para sus proyectos y, por ende, nuevas publicaciones, que acabarán por otorgarle un reconocimiento académico y profesional. Hay quienes, para romper este círculo vicioso, reclaman cambiar el foco del interés de la carrera académica de los investigadores y premiar aspectos de la actividad científica que son prácticamente ignorados en la actualidad como, por ejemplo, actuar como revisor de manuscritos y proyectos, facilitar a terceros el acceso a los protocolos y datos de los estudios, formar jóvenes investigadores y replicar estudios previos^{24,37,39}. No se trata de no considerar la publicación de los resultados científicos, que es y seguirá siendo el principal instrumento de evaluación, pero sí de reducir su extraordinaria importancia actual, que impregna todos los estratos de la actividad científica. En todo caso, las instituciones de investigación deben fomentar la “cultura de la replicación”²⁴. Esto, sin embargo, no es fácil de conseguir cuando actualmente prima la ayuda a investigaciones novedosas sobre las de reproducibilidad⁵.

Ya se ha comentado que las revistas científicas y los NIH han tomado importantes decisiones que es de esperar surtan su efecto. Pero también otras muchas agencias financiadoras deben plantearse como prioridad asegurar la reproducibilidad de las investigaciones que financian. Del mismo modo, las instituciones de investigación y las universidades tienen una responsabilidad capital a este respecto^{24,37,39}. Se ha visto que las 5 instituciones más prestigiosas de Reino Unido en experimentación animal fracasan estrepitosamente en poner en práctica métodos para limitar el riesgo de sesgos sistemáticos²¹. Parecería lógico esperar que los centros no solo formasen adecuadamente a los investigadores sino que supervisasen la integridad de la investigación que se realiza en sus instalaciones y establecieran unos estándares mínimos de actuación³⁹. Es probable que todo esto tarde en llegar a nuestro país. Mientras esto acontece, se debería empezar consiguiendo que todos los centros de investigación y universidades tuvieran publicado un código de buenas prácticas de investigación^{40,41} y que contasen con un comité que se ocupe de fomentar la integridad científica y de evaluar los casos de sospecha de falta de esta. Toda institución debería asegurarse de que cualquier investigador en prácticas recibiese la formación y el entrenamiento necesarios para interiorizar los valores de la integridad científica⁴² –además de adquirir una correcta formación en metodología y estadística-. Parece claro que a la investigación biomédica le ha llegado el tiempo de “confiar, pero verificar”.

Conflictos de intereses

Los autores declaran carecer de conflicto de intereses alguno

Referencias

- 1.-Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov.* 2013;12:569.
- 2.-Osherovich, L. Hedging against academic risk. *SciBX* 4(15); doi:10.1038/scibx.2011.416.
Published online April 14 2011. Consultado el 2 de noviembre de 2015. Accesible en <http://www.nature.com/scibx/journal/v4/n15/full/scibx.2011.416.html>
- 3.-Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature.* 2012; 490: 187-91.
- 4.-Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circ Res.* 2015; 116:116-26.
- 5.-Anon. Scientists like to think of science as self-correcting. To an alarming degree, it is not. *The Economist.* 19 October 2013. Consultado el 2 de noviembre de 2015. Accesible en <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble>
- 6.-Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet.* 2009; 374: 86-9
- 7.-Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proc Natl Acad Sci USA* 2012, 109: 17028–33.
- 8.-Swanson E. Americans have little faith in scientists, science journalists: poll. *Huffington Post.* 23 December 2013. Consultado el 2 de noviembre de 2015. Accesible en http://www.huffingtonpost.com/2013/12/21/faith-in-scientists_n_4481487.html

- 9.-Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov.* 2011; 10:712-3.
- 10.-Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature.* 2012; 483: 531-3
- 11.-Mak IWY, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. *Am J Transl Res.* 2014; 6(2): 114–118.
- 12.-Young SS, Karr A., Deming, data, and observational studies: A process out of control and needing fixing. *Significance* 2011, 8, 116–120.
- 13.-Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, Mills EJ ,et al. Reanalyses of randomized clinical trial data. *JAMA.* 2014, 312: 1024-32
- 14.-Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA.* 2014; 311:368-77
- 15.-Fanelli D. “Positive” Results Increase Down the Hierarchy of the Sciences. *PLoS ONE* 2010, 5: e10068.
- 16.-Bohannon J. Many psychology papers fail replication test. *Science* 2015, 349: 910
- 17.-Open Science Collaboration. Estimating the reproducibility of psychological sciences. *Science* 2015, 349: 943, aac4716.
- 18.-Baker M. Over half of psychology studies fail reproducibility test. 28 August 2015. Consultado el 2 de noviembre de 2015. Accesible en <http://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248>
- 19.-Errington TM, Iorns E, Gunn W, Tan FE, Lomax J, Nosek BA. An open investigation of the reproducibility of cancer biology research. *Elife.* 2014; 3: e04333

- 20.-Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. PLoS Biol. 2015; 13:e1002165.
- 21.- Macleod MR, Lawson McLean A, Kyriakopoulou A, Serghiou S, de Wilde A, Sherratt N, et al. Risk of Bias in Reports of In Vivo Research: A Focus for Improvement. PLoS Biol 2015, 13: e1002273.
- 22.-Ioannidis JP. Why most published research findings are false. PLoS Med. 2005; 2: e124.
- 23.-Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14:365-76
- 24.-Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet. 2014; 383: 166-75.
- 25.- Perrin S. Make mouse studies work. Nature 2014, 507: 423-5
- 26.-Lorsch JR, Collins FS, Lippincott-Schwartz J. Fixing problems with cell lines. Science 2014, 346: 1452–1453.
- 27.-Baker M. Reproducibility crisis: Blame it on the antibodies. Nature. 2015; 521: 274-6
- 28.-Fanelli D. Negative results are disappearing from most disciplines and countries. Scientometrics 2012, 90: 891–904
- 29.-Dal-Ré R, Caplan AL. Journal editors impasse with outcome reporting bias. Eur J Clin Invest. 2015; 45: 895-8
- 30.-Begley CG. Six red flags for suspect work. Nature. 2013; 497: 433-4.
- 31.-Ioannidis JP, Fanelli D, Dunne DD, Goodman SN. Meta-research: Evaluation and Improvement of Research Methods and Practices. PLoS Biol. 2015;13:e1002264.
- 32.-Nuzzo R. Fooling ourselves. Nature 2015, 526: 182-5
- 33.-MacCoun R, Perlmutter S. Hide results to seek the truth. Nature 2015, 526: 187-9

- 34.-Jansen of Lorkeers SJ, Doevendans PA, Chamuleau SA. All preclinical trials should be registered in advance in an online registry. *Eur J Clin Invest*. 2014; 44:891-2.
- 35.-Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010, 8: e1000412.
- 36.-Baker D, Lidster K, Sottomayor A, Amor S. Two Years Later: Journals Are Not Yet Enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal Studies. *PLoS Biol* 2014, 12: e1001756.
- 37.-Collins FS, Tabak LA. NIH plans to enhance reproducibility. *Nature*. 2014; 505: 612-3.
- 38.-National Institutes of Health. Principles and Guidelines for Reporting Preclinical Research. Consultado el 2 de noviembre de 2015. Accesible en <http://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>
- 39.-Begley CG, Buchan AM, Dirnagl U. Robust research: Institutions must do their part for reproducibility. *Nature*. 2015; 525: 25-7.
- 40.-Universidad Autónoma de Madrid. Código de Buenas Prácticas de Investigación. Febrero 2013. Consultado el 2 de noviembre de 2015. Accesible en https://www.uam.es/ss/Satellite/es/1234886377819/contenidoFinal/Comite_de_etica_de_la_investigacion.htm
- 41.-CSIC. Consejo Superior de Investigaciones Científicas. Código de Buenas Prácticas Científicas del CSIC. Marzo 2010. Consultado el 2 de noviembre de 2015. Accesible en <http://www.csic.es/etica-en-la-investigacion#BPC>
- 42.-European Science Foundation. ALLEA All European Academies. The European Code of Conduct for Research Integrity. March 2011. Consultado el 2 de noviembre de 2015. Accesible en

http://www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf

Tabla 1.- Algunos de los errores en el diseño de estudios científicos²²⁻²⁴.

- Tamaño de la muestra: a menor tamaño, más dificultad de que el resultado obtenido sea verdadero. Un tamaño de muestra pequeño implica un poder estadístico bajo. Un poder estadístico bajo implica una baja probabilidad de encontrar efectos reales, verdaderos, y facilita el hallazgo de resultados falsos positivos.
- Tamaño del efecto: a menor tamaño del efecto encontrado, es menos probable que el efecto sea verdadero. El poder estadístico también está relacionado con el tamaño del efecto. Con un poder estadístico bajo, lo más probable es que el tamaño del efecto encontrado esté exagerado; además, varios estudios con poder estadístico bajo darán lugar al ‘efecto vibración’, es decir, a un amplio intervalo en la magnitud de los efectos encontrados de la variable en estudio. Cuando, como ocurre actualmente, la mayoría de las investigaciones biomédicas describen tamaños de efecto pequeños, es más probable que se describan muchos hallazgos falsos positivos. De hecho, obtener en la actualidad un tamaño de efecto muy grande o altamente significativo, debe ser observado con cautela pues puede ser indicativo de un sesgo importante.
- Relación entre el hallazgo y los datos previos al estudio: la probabilidad de que un hallazgo sea real, verdadero (es decir, el valor predictivo positivo, VPP), está fuertemente influenciado por la probabilidad de que así sea antes de realizar el estudio. Un ensayo controlado aleatorizado de fase 3 (confirmatorio de eficacia) correctamente realizado, con una probabilidad inicial del 50% de que se alcancen resultados verdaderos y un poder estadístico del 80%, tiene un VPP del 85%: es decir, en el 85% de las ocasiones el resultado obtenido en el ensayo será verdadero. Por el contrario, un ensayo controlado aleatorizado de fase 2 (exploratorio de eficacia) correctamente realizado, con una probabilidad inicial del 20% de que se alcancen resultados verdaderos y un poder estadístico del 20%, tiene un VPP del 23%: es decir, en el 77% de las ocasiones el resultado obtenido en el ensayo será falso.
- Definición de las variables de evaluación: a mayor flexibilidad en la definición de las variables, menor probabilidad de que los resultados sean verdaderos. Así, los resultados de un ensayo clínico tendrán más probabilidades de ser verdaderos cuando se usan variables bien definidas (vg, muerte) que si se usan variables con cierto grado de interpretación como, por ejemplo, las escalas de evaluación, de uso frecuente en psiquiatría. Las variables con alto grado de subjetividad favorecen la ‘vibración’ del efecto dependiendo de los análisis estadísticos realizados.

- Validez de las variables de evaluación: variables válidas son las clínicamente relevantes para la patología en estudio, aspecto de especial relevancia cuando se usan escalas de evaluación clínica o variables compuestas.

Poder estadístico: es la probabilidad de que se rechazará correctamente la hipótesis nula cuando ésta es falsa; es decir, la probabilidad de no cometer un error tipo II o falso negativo.

Tamaño del efecto: es una medida estandarizada que cuantifica el tamaño de la diferencia entre dos grupos o la fuerza de una asociación entre dos variables.

Valor predictivo positivo (VPP): es la probabilidad de que un resultado positivo refleje un efecto real, verdadero, es decir, que sea un hallazgo positivo verdadero.

Tabla 2.- Aspectos a comprobar en la evaluación de artículos de investigación preclínica para asegurar la consistencia de los resultados descritos según Begley³⁰

<i>Factor</i>	<i>Comentario</i>
Enmascaramiento de los experimentos	Los estudios <i>in vitro</i> , en animales y lectura de geles, deben realizarse de forma enmascarada o, al menos, el evaluador debe estarlo
Repetición de los experimentos	Antes de enviar a publicar los resultados hay que repetir los experimentos, algo que solo se hace de forma excepcional
Comunicación de todos los resultados	Hay que evitar la publicación selectiva de resultados, algo que es difícilmente detectable por los revisores. Se deben comprobar las imágenes que, se ha visto, pueden ser manipuladas de forma digital.
Grupos control	Siempre se debe informar de los resultados obtenidos en el grupo control, algo que no siempre se hace
Validación de reactivos (*)	Es vital tener constancia de que, por ejemplo, los anticuerpos detectan solo el antígeno de interés
Análisis estadísticos adecuados	Es frecuente el uso de pruebas estadísticas inadecuadas en los estudios de experimentación animal

(*) Aquí se puede añadir cualquier otro tipo de material como, por ejemplo, líneas celulares, de vital importancia en la investigación oncológica (ver el texto)

Tabla 3.- (Resumen de los) Principios y guía para la comunicación de la investigación preclínica³⁸

1. Análisis estadístico riguroso. La revista debe disponer de un mecanismo que compruebe la exactitud estadística de los manuscritos.
2. Transparencia. Las revistas no deben imponer límites -o tener límites generosos- en la extensión de la sección de 'métodos'. Las revistas dispondrán de un procedimiento editorial que asegure la comunicación a revisores y lectores de la información clave relativa a la metodología y análisis.

Normas para una comunicación rigurosa:

- *Estándares.* Usar estándares aceptados tales como ARRIVE⁴⁶
 - *Replicaciones.* Comunicar cuántas veces se replicó cada experimento y si los resultados se confirmaron por repetición y bajo qué condiciones
 - *Estadística.* Exigir que se informe toda la estadística incluyendo el tipo de análisis, el valor exacto de la N, la definición de centro, y las medidas de dispersión y precisión
 - *Aleatorización.* Para todos los estudios con animales, informar como mínimo sobre si las muestras fueron aleatorizadas y el método utilizado.
 - *Enmascaramiento.* Para todos los estudios con animales, informar como mínimo sobre si los investigadores estaban enmascarados para la asignación a los grupos y en la evaluación.
 - *Estimación de la muestra.* Qué método estadístico se utilizó en el cálculo de la muestra
 - *Criterios de inclusión y exclusión.* Se debe informar de cualquier criterio para excluir cualquier dato o sujeto, e incluir todo resultado experimental que haya sido omitido del manuscrito por la razón que sea.
3. Acceso a los datos y material. Como poco, todos los conjuntos de datos (*datasets*) en los que se fundamentan las conclusiones del manuscrito deben estar disponibles para el peticionario (cuando éticamente sea adecuado) durante la evaluación de aquél (por el director de la revista y revisores) y tras petición razonable inmediatamente después de su publicación. Se recomendará depositar el conjunto de datos en repositorios de acceso público y compartir el software (se debe informar si éste está disponible y cómo obtenerlo).

4. Consideración de refutaciones. Al publicar un trabajo, la revista asume la responsabilidad de considerar para su publicación refutaciones de ese trabajo, que deben cumplir los estándares de calidad usuales

 5. Considerar el establecimiento de guías de mejores prácticas para:
 - Datos fundamentados en imágenes (vg, *Western blots*).
 - Descripción del material biológico que identifique de forma inconfundible los reactivos, en especial, anticuerpos, líneas celulares y animales
-