



# Más allá de la reproducibilidad:

## Replicabilidad y otras cuestiones relativas a la validez de la investigación

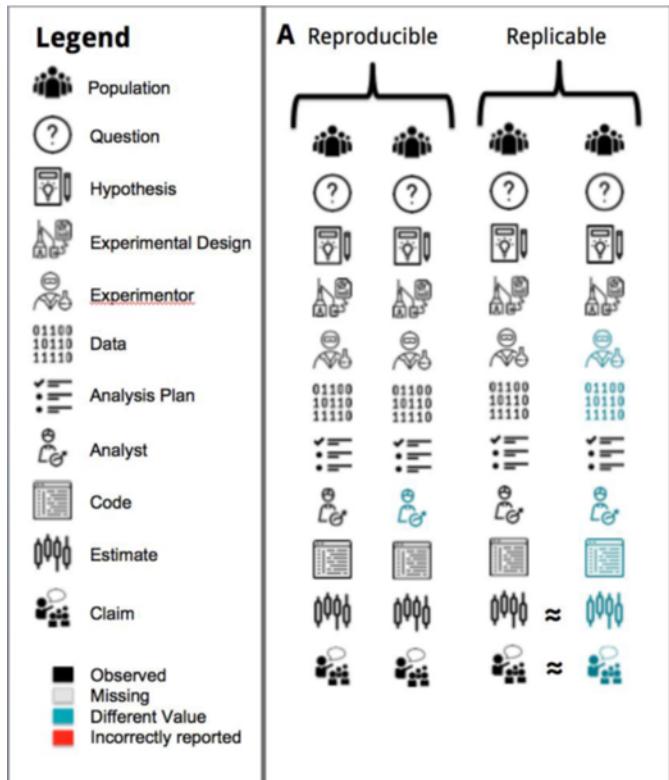
Miguel A. Martínez-Beneito - Grupo de investigación Bayensians. Fundación FISABIO (Valencia).  
25 de noviembre de 2019

# Estructura de la sesión:

- **Definición e ilustración** del concepto de replicabilidad.
- **Causas** de la crisis de replicabilidad.
- **Abordaje** de la replicabilidad.

# **Replicabilidad: Definición e ilustración del concepto de replicabilidad.**

# Reproducibilidad y replicabilidad



**Reproducibilidad: Mismos datos, distintos experimentadores, resultados iguales (presumiblemente) -> Transparencia de procesos.**

**Replicabilidad: Distintos datos, similares condiciones experimentales, resultados similares (constatables) -> Generabilidad de los resultados.**

**Reproducibilidad y replicabilidad** son dos conceptos a menudo confundidos en la literatura.

En ningún caso **reproducibilidad** garantiza la **generabilidad** de los resultados.

“Results that are not reproducible are **hard to verify** and results that do not replicate in new studies are **harder to trust.**” (<http://simplystatistics.org/2017/03/02/rr-glossy/>)

# ¿También crisis de replicabilidad?

## Replication crisis

From Wikipedia, the free encyclopedia

The **replication crisis** (or **replicability crisis**) refers to a **methodological crisis** in **science** in which scientists have found that the results of many **scientific experiments** are difficult or impossible to **replicate** on subsequent investigation, either by independent researchers or by the original researchers themselves.<sup>[1]</sup> While the crisis has long-standing roots, the phrase was coined in the early 2010s as part of a growing awareness of the problem.

[Wikipedia \(https://en.wikipedia.org/wiki/Replication\\_crisis\)](https://en.wikipedia.org/wiki/Replication_crisis)

SCIENCE NEWS | Wed Mar 28, 2012 | 7:09pm BST

In cancer science, many "discoveries" don't hold up



[http://uk.reuters.com/... \(http://uk.reuters.com/article/us-science-cancer-idUSBRE82R12P20120328\)](http://uk.reuters.com/article/us-science-cancer-idUSBRE82R12P20120328)

## Cancer Research Is Broken

There's a replication crisis in biomedicine—and no one even knows how deep it runs.

By Daniel Engber



<http://www.slate.com/...>

([http://www.slate.com/articles/health\\_and\\_science/future\\_tense/2016/04/biomedicine\\_facing\\_a\\_worse\\_replication\\_crisis\\_than\\_1](http://www.slate.com/articles/health_and_science/future_tense/2016/04/biomedicine_facing_a_worse_replication_crisis_than_1)

# Replicabilidad en Medicina

## Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD



**Context** Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most

Ioannidis (JAMA, 2005) ([jamanetwork.com/journals/jama/fullarticle/201218](http://jamanetwork.com/journals/jama/fullarticle/201218)) lleva a cabo un **estudio de replicación** de trabajos publicados entre 1990 y 2003 en New England, JAMA, Lancet o revistas médicas especializadas con  $IF > 7$ .

**Criterio de inclusión:** número de citas  $> 1000$ .

**Busqueda bibliográfica** de estudios similares, con criterios de calidad (tamaño muestral, diseño, ...) **similares o superiores**, que pudieran **corroborar o refutar** dichos estudios.

Los **49** artículos encontrados como **elegibles**, se clasifican como:

- **Negative**, no presentaba resultados positivos (**4**, todos refutaban estudios anteriores).
- **Unchallenged**, no se ha valorado su replicabilidad en la literatura (**11**).
- **Contradicted**, el efecto original no se ha podido replicar (**7**).
- **Initially stronger effects**, el efecto original o su duración se reduce a menos de la mitad (**7**).
- **Replicated**, el efecto original ha sido corroborado (**20**).

**41.2%**(=14/34) de los estudios testados, con efectos positivos, son **refutados** por estudios de similares características.

# El “Decline Effect”

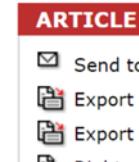
Pero ni siquiera la publicación de réplicas positivas de trabajos es **garantía** de nada (<http://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>).

**Facciones asimétricas** se consideran signo de **mutaciones genéticas**.

*Nature* 357, 238-240 (21 May 1992) | doi:10.1038/357238a0; Accepted 17 March 1992

Female swallow preference for symmetrical  
male sexual ornaments

Anders Pape Møller\*



Møller (Nature, 1991)

(<https://www.nature.com/nature/journal/v357/n6375/pdf/357238a0.pdf>) descubre que **hembras de golondrinas prefieren** machos con **plumaje más simétrico**.

**Simetría** de plumaje sería indicador indirecto de **calidad genética**.

Hembras **aplican** este criterio (mecanismo de selección natural) de forma **inconsciente**.

En los 3 años siguientes **9 (de 10)** artículos corroboran la teoría anterior.

Dicha teoría se estudia en **humanos**:

- Mujeres prefieren el **olor** de hombres con facciones **simétricas**, aunque sólo durante su periodo **fértil**.
- Mujeres tenían más **orgasmos** con hombres **simétricos**.
- **Bailarines** simétricos son consistentemente **evaluados** como mejores.

Pero:

- En 1994 sólo **8 de 14** artículos corroboran el efecto original.
- En 1995, **4 de 8** artículos lo corroboran.
- En 1998, **4 de 12** artículos lo corroboran.
- De 1992 a 1997, el **efecto** originalmente encontrado **decae en un 80%**.

Decline effect, estructura general:

Tras proponerse un paradigma, los **procesos de publicación científica tienden a corroborarlo** (sesgo de publicación).

**Años después** los incentivos editoriales cambian, dando preferencia a aquellos **resultados que desaprueban** el paradigma establecido.

Decline effect ha sido documentado en muchas **más situaciones**:

- **Efectividad** de fármacos **antipsicóticos** de segunda generación.
- **Percepción extra-sensorial**: Capacidad de predecir hechos futuros.
- ...

Decline effect es una **expresión más de la crisis** de replicabilidad.

# Replicabilidad en Psicología

PSYCHOLOGY

## Estimating the reproducibility of psychological science

Open Science Collaboration\*†

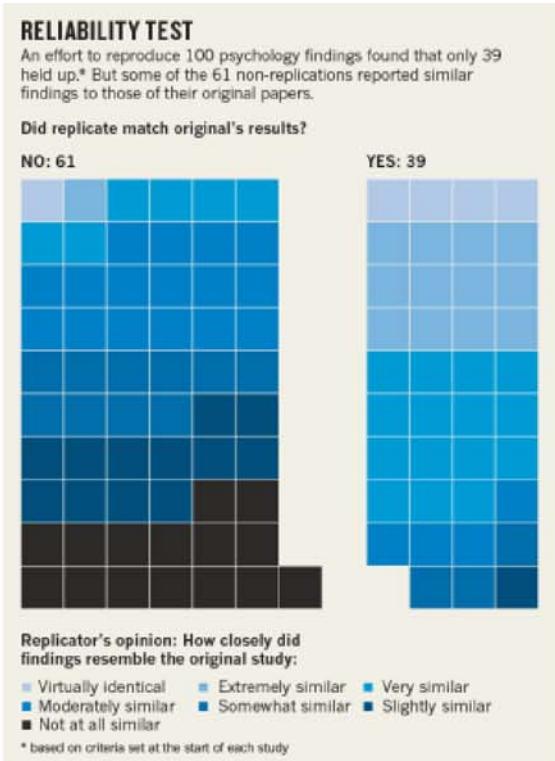
Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original

En 2015 (Science) (<http://science.sciencemag.org/content/349/6251/aac4716>), el Center for Open Science publica los resultados de un **estudio colaborativo de replicación** en psicología.

**Mayor esfuerzo** de replicabilidad llevado a cabo hasta la fecha. “One of the top 10 scientific breakthroughs of the year (Science)” (<https://projects.iq.harvard.edu/psychology-replications>)

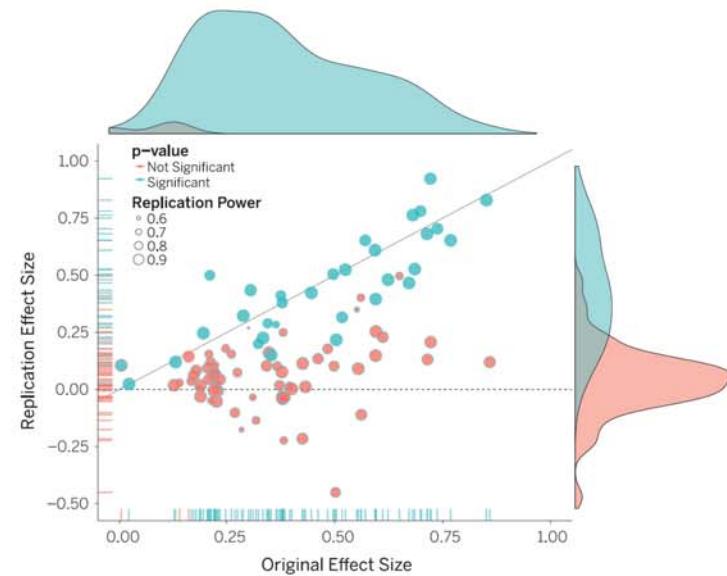
Distintos **grupos** de forma independiente **replican 100 estudios** publicados en la literatura (artículos de 2008 en revistas influyentes).

A diferencia del estudio de Ioannidis, este estudio **replica específicamente** los estudios de la literatura (**evidencia empírica**).



**Sólo 39 de los 100 estudios corroboran los resultados originales.**

El problema podría ser **todavía peor** ya que se trata de estudios en revistas “**de prestigio**”.



**Original study effect size versus replication effect size (correlation coefficients).** Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

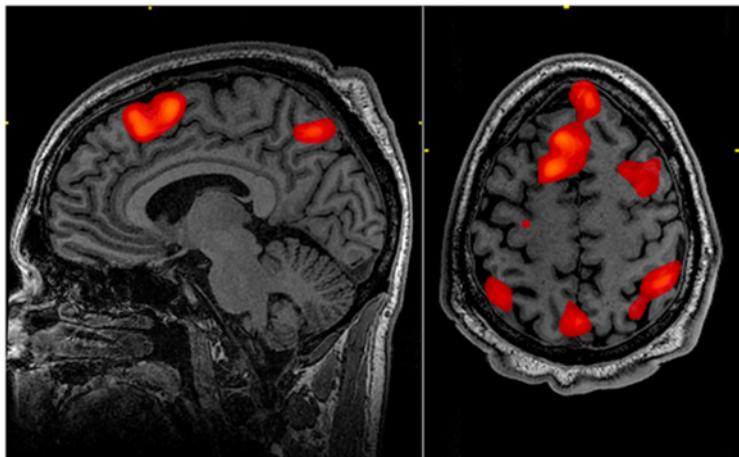
**Sólo parte** de los estudios replicados parecen reproducir los efectos originales.

## **Replicabilidad: causas de la crisis**

# 1.- Deficiencias en los procedimientos

## fMRI

**fMRI** ha sido la herramienta principal para estudiar la **funcionalidad** de cada región cerebral.



An fMRI scan during working memory tasks.

Habitualmente, **individuos** se someten a **estímulos** y, mediante resonancia magnética, se determinan las **regiones del cerebro** con mayor consumo de hemoglobina tras dicho estímulo.

Dichas **áreas**, teóricamente, serían las **encargadas de procesar** y asimilar el estímulo.

# Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund<sup>a,b,c,1</sup>, Thomas E. Nichols<sup>d,e</sup>, and Hans Knutsson<sup>a,c</sup>

PNAS, 12/07/2016 (<http://www.pnas.org/content/113/28/7900>)

“We used **resting-state fMRI data** from 499 healthy controls ... **Using this null data** with different experimental designs, **we estimate the incidence of significant results**. In theory, **we should find 5% false positives** (for a significance threshold of 5%), but instead **we found** that the most common software packages for fMRI analysis (SPM, FSL, AFNI) can result in **false-positive rates of up to 70%**. These results question the validity of a number of fMRI studies and may have a large impact on the interpretation of weakly significant neuroimaging results.”

El tratamiento estadístico de la **dependencia entre observaciones** empleado en los 3 paquetes de software estudiado es deficiente, disparando el número de falsos positivos.

According to many of the headlines that greeted “Cluster failure”, the paper is a devastating bombshell that could demolish the whole field of functional magnetic resonance imaging (**fMRI**):

[Discover](#)

Bug in fMRI software calls 15 years of research into question  
(Wired)

A bug in fMRI software could invalidate 15 years of brain research. This is huge. (ScienceAlert)

New Research Suggests That Tens Of Thousands Of fMRI Brain Studies May Be Flawed (Motherboard)

(<http://blogs.discovermagazine.com/neuroskeptic/2016/07/07/false-positive-fmri-mainstream/#.WQZCLNryiM8>)

Este trabajo ha supuesto un auténtico **terremoto** dentro del campo de fMRI.

Se estima que unos **3500 artículos** podrían estar **afectados** (alrededor del 9% de la literatura del campo).

([http://blogs.warwick.ac.uk/nichols/entry/bibliometrics\\_of\\_cluster/](http://blogs.warwick.ac.uk/nichols/entry/bibliometrics_of_cluster/))

## 2.- Conflictos de intereses

Conflictos de intereses **distorsionan la literatura** científica, sesgandola en **direcciones interesadas**.

No se publica en función de la **evidencia** sino de la **conveniencia**.

**Conflictos** de interés están, en parte, **detrás de la crisis** de replicabilidad.

**Tipos** de conflictos de interés en publicaciones científicas:

- Intereses **comerciales, políticos y sociales**.
- Presión por publicar de los **científicos**.
- Interés de las **revistas** por publicar resultados novedosos, rompedores.

# **Intereses comerciales, políticos y sociales.**

**Investigación científica** ofrece un estupendo **argumento a decisiones o intereses** arbitrarios que de otra forma serían difíciles de justificar.

La **aureola de veracidad** de la ciencia **legitima decisiones** políticas y sociales (“evidence based policy”) o intereses comerciales.

Grupos de presión, asociaciones, lobbies ... buscan **sustento en literatura científica**.

El **problema** surge cuando la ciencia **no** se usa para **guiar** las decisiones sino **para justificarlas**.

“**Scientists** working in the policy arena are often **naïve** about the impact of their findings” ([https://en.wikipedia.org/wiki/Science\\_on\\_the\\_Verge](https://en.wikipedia.org/wiki/Science_on_the_Verge))

# Danone

"Instituto Danone para la Nutrición y la Salud"

([http://www.institutodanone.es/ayudas\\_proyectos\\_investigacion\\_cientifica.html](http://www.institutodanone.es/ayudas_proyectos_investigacion_cientifica.html))

ofrece **ayudas a la investigación:**



## AYUDAS A PROYECTOS DE INVESTIGACIÓN CIENTÍFICA

Con el fin de fomentar la actividad científica, y apoyar particularmente a los grupos de investigación más jóvenes, el Instituto Danone convoca desde 2014 dos Ayudas de 40.000€ cada una para promover la labor de investigación científica en las áreas de la Alimentación, Nutrición y Salud.

Se trata de ayudas de dos años de duración para la realización de Proyectos de Investigación Científica sobre Alimentación, Nutrición y Salud en sus aspectos básico, clínico y aplicado.

[BASES DE LA CONVOCATORIA 2017](#)

[FORMULARIO DE SOLICITUD](#)

Leyendo la **convocatoria** con más detalle uno se de cuenta:

1. Se aceptarán aquellas propuestas sobre Alimentación, Nutrición y Salud en sus aspectos básico, clínico y aplicado a realizar en universidades, instituciones o institutos de investigación públicos o privados en España, y en relación con cualquiera de los siguientes campos de actividad: nutrición clínica, metabolismo, epidemiología, genética y biología molecular, y ciencia y tecnología de los alimentos. En la concesión de una de las dos ayudas se dará preferencia a las propuestas que, teniendo en cuenta lo anterior, tengan como objetivo el estudio de las propiedades y efectos positivos para la salud del consumo de yogur.

Financiación se destina a estudiar **propiedades beneficiosas** de sus productos (para que posiblemente engrosen la literatura científica).

La **intencionalidad** de estos estudios es evidente. Se busca **sustento** no conocimiento.

## + Danone

**Actimel** es uno de los productos estrellas de Danone con una facturación anual de alrededor de 5300 millones de dólares (**25% de la facturación del grupo**) (<https://scientiablog.com/2012/06/14/la-verdadera-historia-del-actimel-i/>).

Se trata de un **alimento funcional** con unas supuestas propiedades beneficiosas para la salud (“**mejora tus defensas**”).

Su principal propiedad, reducción de diarreas, se atribuye a la presencia de una cepa patentada de **Lactobacillus Casei Imunitass**.

Sus **efectos positivos** se sustentan (supuestamente) en **literatura** científica generada al efecto. De ahí la **importancia comercial** de disponer de literatura científica que permita aducir estas propiedades.

La European Food Safety Administration (**EFSA**) emite en 2010 un **informe** sobre los **efectos** de este producto sobre la salud.

El informe “**rechazó** los más de **20 trabajos** de investigación que pretendían avalar los beneficios del Actimel” (<https://scientiablog.com/2012/06/14/la-verdadera-historia-del-actimel-i/>).

“The Panel concludes that the **evidence** provided is **insufficient to establish a cause and effect relationship** between the consumption of Actimel and a reduction of the risk of C. difficile diarrhoea” (<http://www.efsa.europa.eu/en/efsajournal/pub/1903>)

La **EFSA destaca errores** repetidos de procedimiento en los estudios evaluados.

Las **propiedades beneficiosas** publicitadas de Actimel, y otros prebióticos, se sostienen (y por tanto permiten) sólo por la presencia de vitamina B6 (<https://scientiablog.com/2012/06/18/la-verdadera-historia-del-actimel-ii-me-siento-enganado-por-danone/>) en su composición (aunque su dosis es, por ejemplo, un tercio inferior a la de un plátano).



En cualquier caso el halo ilusorio de **credibilidad de la ciencia** sigue presente en el producto.

# Presión por publicar de los científicos

La ciencia vive una **proliferación “industrial” de datos y de publicaciones** que los explotan. Producción científica tratada como **producto de consumo** disminuye su calidad.

El sistema académico considera el **número de publicaciones** como uno de sus principales parámetros para valorar a sus integrantes. Científicos son en gran parte **evaluados por su volumen** de publicaciones.

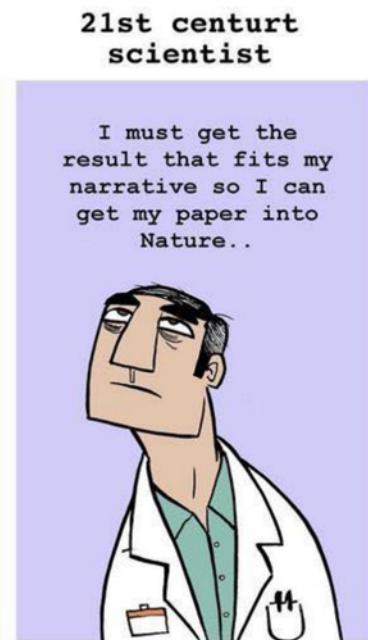
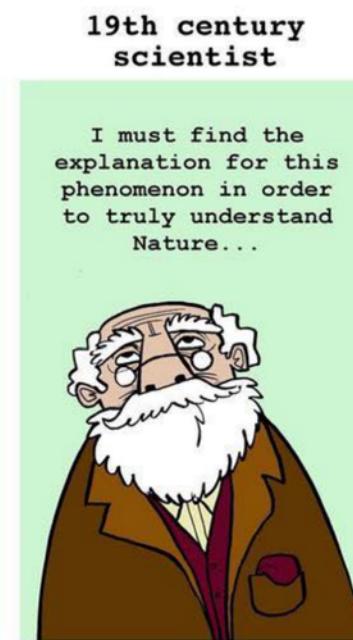
Sería conveniente poner el **foco en la calidad** de las publicaciones más que en su cantidad.

La **presión**, o simplemente el interés, por **publicar** supone un claro **conflicto de interés** (más allá de lo económico) de todas las publicaciones.

# Presión de las revistas por publicar resultados novedosos

Además, las revistas exigen resultados novedosos: “Results need to be exciting, eye-catching, even implausible ... When we review papers we’re often making authors prove that their findings are novel or interesting. We’re not making them prove that their finding are true” (<http://www.nature.com/news/replication-studies-bad-copy-1.10634>)

Esta presión se vive fundamentalmente en aquellas **revistas de mayor impacto**, aquellas que influyen en mayor medida en la literatura.



facebook.com/pedromics

### 3.- Sesgo de publicación

**No todos** los estudios ven la luz de la **publicación científica** con la misma probabilidad.

Resultados **no significativos, o en contra** de la hipótesis inicial de los investigadores, tienden a **no ser publicados**. Este efecto se conoce como “**file drawer problem**” ([https://en.wikipedia.org/wiki/Publication\\_bias](https://en.wikipedia.org/wiki/Publication_bias)).

Este efecto puede **alterar el corpus bibliográfico** ocultando parte de la evidencia científica disponible y por tanto **sesgando el conocimiento acumulado** en ciertas áreas de investigación.

Sesgo de publicación y el **decline effect**: “[Another reason may be the publication bias: scientists and scientific journals prefer to publish positive results of experiments and tests over null results, especially with new ideas. As a result, the journals may refuse to publish papers that do not prove that the idea works. Later, when an idea is accepted, journals may refuse to publish papers that support it.](https://en.wikipedia.org/wiki/Decline_effect)” ([https://en.wikipedia.org/wiki/Decline\\_effect](https://en.wikipedia.org/wiki/Decline_effect))

# Sesgo de publicación y acupuntura

**Revisión sistemática de la literatura sobre la efectividad de acupuntura, de 1966 a 1995, para un grupo de países (Vickers et al., 1998)**

**Table 1** Results of Controlled Clinical Trials of Acupuncture by Country of Research

Country	Total Trials Analyzed	Favoring Test Treatment Number	Favoring Test Treatment Percentage
USA	47	25	53
China	36	36	100
Sweden	27	16	59
UK	20	12	60
Denmark	16	8	50
Germany	16	10	63
Canada	11	3	27
Russia/USSR	11	10	91
Austria	9	8	89
Italy	9	8	89
Australia	6	1	17
France	6	5	83
Taiwan	6	6	100
Japan	5	5	100
Finland	4	2	50
Hong Kong	3	3	100
Netherlands	3	1	33
New Zealand	3	2	67
Poland	3	2	67
Switzerland	3	1	33
Bulgaria	2	2	100
Brazil	1	1	100
Croatia	1	1	100
Israel	1	1	100
Nigeria	1	1	100
Sri Lanka	1	0	0
Vietnam	1	1	100
Total	252	171	68

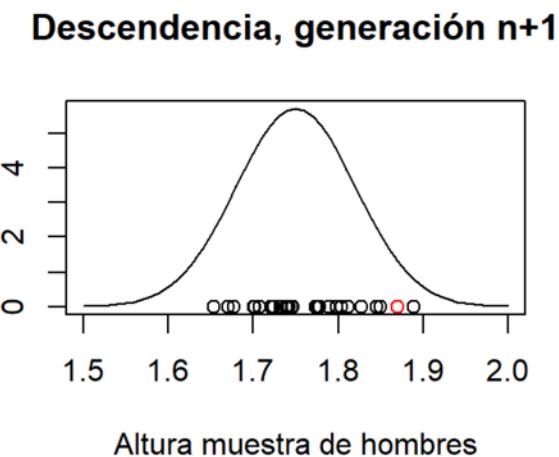
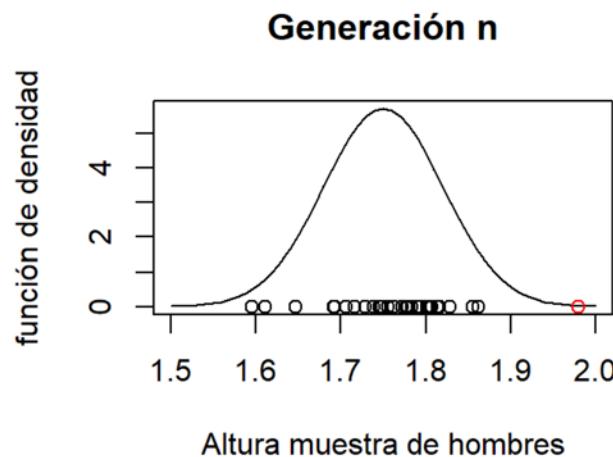
(<https://pdfs.semanticscholar.org/2b01/5f2b2e09f6b06ab236a30adf6eda72c5f6c2.pdf>).

Los países **asiáticos** reportan resultados significativamente **favorables** a la acupuntura en **61 de 63** estudios (96.8%).

Paises **no asiáticos** reportan resultados significativamente **favorables** a la acupuntura en **110 de 191** estudios (57.6%).

## 4.- Regresión a la media

Es el fenómeno estadístico por el que los **individuos con observaciones extremas** tenderán a estar **más cerca a la media** de la población en posteriores observaciones.



Este fenómeno de “regresar” hacia la media **dio nombre** originalmente a los modelos de regresión.

Estudios son **muestras de tamaño 1** del universo de posibles estudios. Sacamos **conclusiones a partir de una única muestra**.

Tomando una **muestra lo suficientemente extrema** podremos demostrar virtualmente **cualquier resultado**. Un 5% de las muestras posibles, simplemente por azar, nos deberían dar un resultado significativo (aunque no existiera efecto).

Así, muchos **hallazgos “significativos”** pueden ser simplemente **outliers estadísticos** (muestras suficientemente anómalas) que al intentar ser replicados **pierden su excepcionalidad**.

Regresión a la media podría estar también **detrás del decline effect**:

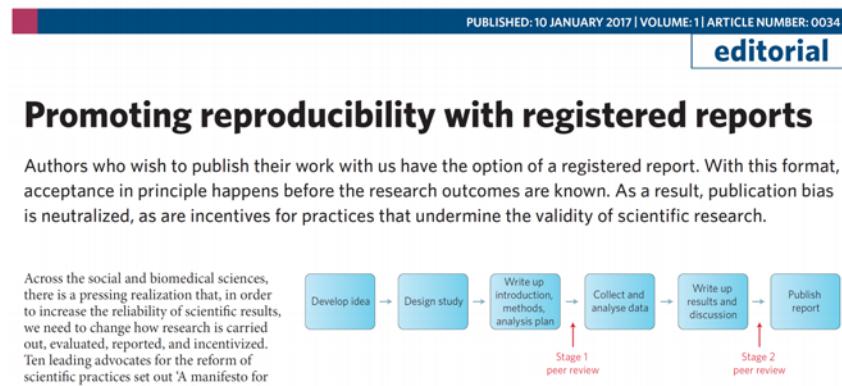
-**Conflictos de intereses** hacen que outliers estadísticos no pasen desapercibidos y acaben publicados.

-**Sesgo de publicación** evita que dichos resultados anómalos sean **contrarestados** inicialmente, aunque sí posteriormente cuando ya no son tan novedosos.

## **Abordaje de la crisis de replicabilidad**

# Pre-registro de estudios

Procedimiento adoptado de **ensayos clínicos**: todos los procedimientos de análisis han de ser **declarados de antemano** a la recolección de datos de análisis. Center for Open Science (<https://cos.io/rr/>), por ejemplo, ofrece esta posibilidad.



<https://www.nature.com/...>

(<https://www.nature.com/articles/s41562-016-0034>)

Una vez **aprobado** el registro previo del estudio por un grupo de evaluadores **se garantiza la publicación** de los resultados del estudio.

Triple objetivo:

- Limita **intervención interesada** de experimentadores en el análisis.
- Fomento publicaciones con **métodos rigurosos y bien documentados**.
- **Disminución de sesgo** de publicación (fomenta publicación de resultados negativos).

# Atención específica a la replicabilidad en revistas

Estudios de replicación son **menos atractivos** que estudios originales.

Replicación poco agradecida. Todos preferimos un esfuerzo creativo más que confirmatorio de los resultados de otros investigadores. Este hecho debería ser combatido en bien de la propia ciencia.

**Revistas** deberían promover estudios de replicación para vigilar la **calidad de sus contenidos**.

“Perspectives in Psychological Science” ha creado una **sección exclusiva** para estudios de replicación. (<http://www.newyorker.com/tech/elements/the-crisis-in-social-psychology-that-isnt>)

Social Psychology, **número monográfico** sobre replicación de estudios de importancia particular en este ámbito.

(<http://www.newyorker.com/tech/elements/the-crisis-in-social-psychology-that-isnt>)

Algunas revistas (Psychological science) **anima** específicamente a la publicación de **estudios con protocolos pre-registrados**.  
([https://en.wikipedia.org/wiki/Replication\\_crisis](https://en.wikipedia.org/wiki/Replication_crisis))

# Atención específica a resultados negativos

Published online 23 February 2011 | *Nature* **470**, 437 (2011) |  
doi:10.1038/470437a

<https://www.nature.com/...>

Column: World View

## Unpublished results hide the decline effect



Some effects diminish when tests are repeated. Jonathan Schooler says being open about findings that don't make the scientific record could reveal why.

Jonathan Schooler

(<https://www.nature.com/news/2011/110223/full/470437a.html>)

Propuesta de **base de datos** open-access de **resultados negativos**.

PsychFileDrawer.org: **Web** donde psicólogos pueden mandar **resultados no publicados**, sean significativos o no. (<http://psychfiledrawer.org/>)

**Revistas** específicas dedicadas a la publicación de **resultados negativos**:

- Journal of Negative Results.

# Financiación específica a estudios de replicación

La introducción de **replicabilidad** en la agenda de las **agencias de financiación** puede hacer significar una gran diferencia. “A mere 3% of scientific funding devoted to replication could make a big difference. The current amount is, he says, near zero” (Nature, 2015). (<http://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248>)

The Netherlands Organisation for Scientific Research en 2016 ha lanzado una **línea de financiación específica para replicar** estudios de gran impacto en ciencias sociales y salud.

# Para concluir

En materia de replicabilidad, como en tantos otros temas, “**lo difícil no es llegar si no mantenerse**”.

Se pretende la generación de **resultados sólidos** que “pasen el **test del tiempo**” (“**Many scientific results will not stand the test of time**”, Editorial de JASA).

El **método científico** debería dar las **bases** para que así sea aunque, por desgracia, en el entente actual es **menos efectivo de lo deseable**.

