



Una rata generada por IA se cuela en una revista de alto impacto o, de nuevo, las limitaciones del Factor de Impacto de una revista para evaluar la calidad de la investigación científica

Alejandro Sandoval-Lentisco^a, Miriam Tortajada^a, Víctor Martínez-Pérez^b, Rubén López-Nicolás^c

^aDept. de Psicología Básica y Metodología, Universidad de Murcia, España

^bDept. de Psicología, Universidad de Castilla-La Mancha, España

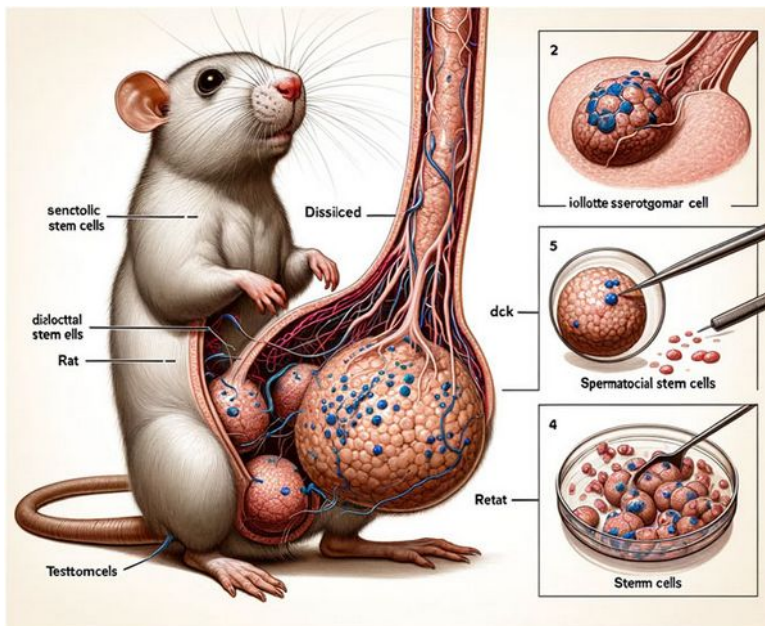
^cDept. de Psicología Básica, Universidad Autónoma de Madrid, España

Tipo de artículo: Actualidad.

Disciplinas: Antropología, Biología, Economía, Filosofía, Inteligencia Artificial, Lingüística, Medicina, Neurociencia, Psicología, Psiquiatría, Sociología.

Etiquetas: Factor de Impacto, evaluación de la ciencia, ANECA, publicación científica.

¿Cómo distinguir a un buen científico de uno que no es tan bueno? Hoy en día, los científicos compiten fieramente entre sí por el acceso a financiación, puestos de trabajo, y cómo no, reputación. El Factor de Impacto de las revistas en las que un científico publica sus artículos (en inglés Journal Impact Factor, JIF) ha sido, hasta el día de hoy, la forma más habitual de valorar la calidad de la investigación. Parece lógico suponer que, si un estudio se publica en una revista que es altamente citada, debe ser bueno. No obstante, instituciones de diversos países, entre las cuales está la española Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA), han decidido modificar la forma en la que se evalúa la investigación restándole peso al JIF. En este artículo, repasamos varias razones por las cuales el JIF no es la forma más adecuada de valorar un estudio y discutimos las nuevas iniciativas que han surgido en respuesta a estas limitaciones.



(cc) Guo et al. (2024) <https://doi.org/10.3389/fcell.2023.1339390>

Sandoval-Lentisco, A., Tortajada, M., Martínez-Pérez, V., y López-Nicolás, R. (2024). Una rata generada por IA se cuela en una revista de alto impacto o, de nuevo, las limitaciones del Factor de Impacto de una revista para evaluar la calidad de la investigación científica. *Ciencia Cognitiva*, 18:4, 14-16.



Evaluar la productividad y el desempeño de los investigadores no es una tarea sencilla y, sin embargo, se debe hacer constantemente en el ámbito científico, por ejemplo, cuando nos enfrentamos a decisiones como elegir quién es el mejor candidato para un puesto en una universidad o a qué proyectos se les concede financiación. Dado que el paso final de cualquier estudio es ser comunicado al resto de la comunidad científica, parece lógico que se usen las publicaciones como un elemento clave para evaluar a la persona. Sin embargo, un conteo de la cantidad de artículos que cada investigador ha publicado es simplista e ignora posibles diferencias en aspectos cruciales, como la calidad de lo publicado. Pero medir la calidad de las publicaciones es complicado.

Tradicionalmente, la principal forma de evaluar un artículo ha sido a través del Factor de Impacto (en inglés Journal Impact Factor, JIF) de la revista en la que estaba publicado. El JIF es, básicamente, el número de citas promedio que reciben los artículos publicados en una revista. Es de suponer que una investigación de gran importancia afecte a un mayor número de otras investigaciones y, por tanto, sea citada en un mayor número de artículos. Y si una revista tiene muchos artículos con muchas citas, debe ser porque publica sólo muy buenos estudios. De aquí se ha pasado a la lógica contraria: si una persona ha conseguido publicar su estudio en una revista con un alto factor de impacto, ese estudio debe ser de gran calidad científica.

Sin embargo, aunque esta métrica ha sido omnipresente en todos los baremos de cualquier disciplina científica, distintas instituciones como la española Agencia Nacional de Evaluación de la Calidad y Acreditación (ANECA) han decidido reconsiderar la forma en que se evalúa la investigación, restándole peso al JIF en favor de otros criterios. Este movimiento está siendo objeto de una intensa discusión. Muchos investigadores defienden que el JIF de la revista es una buena medida para valorar un artículo, principalmente por ser una medida objetiva, eficiente y bien conocida por la comunidad científica. Sin embargo, y aunque este debate no es, ni mucho menos, nuevo (Seglen, 1997; Simons, 2008), volvemos a repasar algunas de las razones por las que el JIF no es la mejor forma de evaluar artículos científicos.

En primer lugar, el JIF, como su propio nombre indica, se trata de una métrica que da información sobre las revistas científicas. Por ello, utilizarlo para evaluar el impacto que puede tener un artículo da lugar a una imagen distorsionada. Esto ocurre porque la distribución de citas que tienen muchas revistas es altamente asimétrica, con muchos artículos muy poco citados y unos pocos artículos muy citados, por lo que su media de citas no es representativa siquiera de su tendencia central y, aún menos, de un artículo en particular. Este hecho hace que el número de citas que recibe un artículo correlacione pobremente con el JIF de la revista (Seglen, 1997).

No obstante, uno podría argumentar que el JIF sigue siendo una medida indirecta de la calidad de los artículos. Dado que las revistas con alto JIF suelen ser muy selectivas, parece razonable que solamente seleccionen artículos de muy buena calidad. De hecho, somos los propios investigadores quienes mandamos a esas revistas los que consideramos que son nuestros mejores trabajos, creando así una especie de profecía autocumplida. Sin embargo, por contraintuitivo que nos parezca, la investigación empírica, así como algunos casos llamativos, muestra que los artículos publicados en revistas con mayor JIF no son necesariamente de mejor calidad.

Por ejemplo, un indicador de calidad de un estudio es que tenga un tamaño muestral grande, ya que ese estudio será más preciso (Button et al., 2013). Teniendo en cuenta diversas revistas del área de la psicología social y de la personalidad, un estudio reciente reveló que la correlación entre los tamaños muestrales de los estudios publicados en una revista y su JIF era, básicamente, cero (Fraley et al., 2022). Otra investigación similar también mostró que el JIF no está relacionado de una manera sustancial con otros indicadores de calidad de los artículos, tales como el número de errores al reportar los resultados, la fuerza de la evidencia de los resultados o su replicabilidad (Dougherty & Horne, 2022).

Como evidencia más anecdótica, tenemos el reciente caso convertido en viral de un artículo que contenía una imagen totalmente inverosímil generada por IA de los genitales de una rata, que es la que abre el

presente artículo. El artículo en cuestión fue publicado en la revista *Frontiers in Cell and Developmental Biology*, que cuenta con un JIF de 5.5, relativamente alto en su disciplina. Otro notable ejemplo podemos encontrarlo cuando, en 2011, el investigador Daryl Bem presentó evidencia de un fenómeno parapsicológico: los eventos futuros podían afectar las respuestas previas de los participantes (Bem, 2011). El artículo, publicado en la prestigiosa revista *Journal of Personality and Social Psychology*, fue duramente criticado por la metodología empleada (Wagenmakers et al., 2011) y fue, entre otras, una de las causas que precipitaron lo que se conoce como “crisis de replicación en psicología” (Pashler & Wagenmakers, 2012). De nuevo, la revista que publicó este artículo tenía, y tiene, uno de los JIFs más altos de las revistas de su disciplina.

Dicho esto, la pregunta más frecuente cuando se cuestiona el uso del JIF es más que razonable: si no utilizamos el JIF para evaluar la investigación, entonces, ¿qué podemos utilizar en su lugar? La respuesta no es para nada obvia. Algunos abogan por una valoración más cualitativa, aunque apoyada por el uso de indicadores cuantitativos más adecuados (véase DORA o CoARA). Otros, sin embargo, critican también ese enfoque por necesitar más recursos y añadir más elementos de subjetividad (Ioannidis & Maniadis, 2023). Y, aunque ya están surgiendo numerosas iniciativas de distintas instituciones a nivel mundial (<https://sfdora.org/reformscape>), lo cierto es que el debate acerca de cuál es la mejor forma de evaluar la investigación sigue siendo una cuestión abierta. No obstante, lo que sí es seguro es que el sistema está pidiendo a gritos un cambio y, con suerte, se encontrará una manera más adecuada de evaluar la investigación.

Referencias

- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407-425.
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, Article 5.
- Dougherty, M. R., & Horne, Z. (2022). Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. *Royal Society Open Science*, 9, 220334.
- Fraley, R. C., Chong, J. Y., Baacke, K. A., Greco, A.J., Guan, H., & Vazire, S. (2022). Journal n-pact factors from 2011 to 2019: Evaluating the quality of social/personality journals with respect to sample size and statistical power. *Advances in Methods and Practices in Psychological Science*, 5, 25152459221120217.
- Ioannidis, J. P. A., & Maniadis, Z. (2023). In defense of quantitative metrics in researcher assessments. *PLOS Biology*, 21, e3002408.
- Pashler, H., & Wagenmakers, E. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530.
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, 314, 497.
- Simons, K. (2008). The misused impact factor. *Science*, 322(5899), 165-165.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100, 426-432.

Manuscrito recibido el 6 de marzo de 2024. Aceptado el 25 de marzo de 2024.