

## ¿Mide el factor de impacto la calidad de las publicaciones? ¿Se deben usar parámetros bibliométricos para evaluar investigadores, grupos o centros?



### Francisco Pérez Vizcaino

Profesor Titular de Farmacología y Vicedecano de Investigación en la Facultad de Medicina, Universidad Complutense de Madrid.



### Giampaolo Co'

Vicedirector del Dipartimento di Matematica e Fisica de la Università del Salento (Lecce, Italia).

El tema en el que se centra esta controversia toca de lleno a la *Revista de Física Médica*. La revista de nuestra sociedad no está “indexada”, es decir, no forma parte del grupo de revistas, numeroso por otro lado, que tienen ese reconocimiento. El hecho de estar indexada significa tener asignado un *índice de impacto* y esto propicia que la publicación tenga más posibilidades de ser seleccionada por los investigadores para intentar publicar sus trabajos. Nos encontramos en un círculo cerrado del que es difícil escapar ya que si el número de artículos que se publican es insuficiente menos periodicidad puede tener la revista y menos artículos pueden ser citados. Esta situación repercute negativamente en el prestigio de la publicación y de los trabajos publicados en ella. Los participantes en esta controversia son los profesores Francisco Pérez Vizcaino y Giampaolo Co', ambos con amplia experiencia investigadora y buenos conocedores de los mecanismos que rigen el mundo editorial de las revistas científicas.

Francisco Pérez Vizcaino (1966), es profesor titular de Farmacología y Vicedecano de Investigación en la Facultad de Medicina de la Universidad Complutense de Madrid. Ha publicado más de 120 artículos indexados en Farmacología Vasculat con más de 2200 citaciones. Su principal línea de investigación se centra en los mecanismos de acción de factores y fármacos vasoactivos.

Giampaolo Co' (1955) es licenciado en Ciencias Físicas por la Universidad de Bolonia (Italia) y Doctor por la Universidad de Bonn (Alemania). Actualmente es Vicedirector del Dipartimento di Matematica e Fisica de la Università del Salento (Lecce, Italia). Su línea de investigación se enmarca en el campo de la Física Nuclear y es autor de más de 100 publicaciones.

\* Pueden enviarse sugerencias sobre temas a debatir a la coordinadora de la sección: Margarita Chevalier  
Email: chevalier@med.ucm.es

## A favor: Francisco Pérez Vizcaino

El factor de impacto (FI) se ha metido en nuestras vidas. Su éxito se debe a que permite representar un hecho complejo mediante un simple número. Es universal y objetivo, se actualiza anualmente, se puede obtener de manera casi inmediata y, aunque no es comparable entre distintas áreas de conocimiento científico, sí lo es en cierta medida la posición relativa del FI dentro del área. Pero no le faltan detractores.

Conviene recordar que el FI de una revista científica en un determinado año, según lo calcula el Institute for Scientific Information (ISI) y se publica anualmente en el *Journal of Citation Reports* (JCR), es el número promedio de veces que se han citado durante ese año los artículos que se han publicado en esa revista durante los dos años anteriores. Por tanto, el FI hace una estimación objetiva del impacto o la repercusión de la revista sobre el trabajo de otros investigadores a corto plazo. Por extensión y de manera razonable se puede decir que es un indicador de la influencia o el prestigio de la revista.

Las limitaciones y los errores del FI son bien conocidos. El periodo de cálculo base para citas es muy corto y por tanto el FI mide más la actualidad o la repercusión a corto plazo. No tiene en cuenta si la citación es en sentido negativo. Tampoco mide el impacto en la sociedad. Las maniobras que realizan las revistas como incluir artículos de revisión, editoriales, forzar las autocitas, etc., para aumentar su propio FI también son un factor de confusión. Por otro lado, el uso del factor de impacto debe ser tratado de manera muy cuidadosa si se aplica a países periféricos o en general a publicaciones que no son en inglés ya que no hay apenas cobertura en ISI de estas revistas<sup>1</sup>. Los índices alternativos como el *Article Influence Score* (basado en el eigenfactor score), que tiene en cuenta el impacto de la revista que cita, o el FI a 5 años, ambos del propio JCR, presentan algunas variaciones cuando se comparan con el FI<sup>2,3</sup> y pueden corregir algunos de sus defectos, pero con algunos matices tienen el mismo significado.

**Por tanto, el FI no es una medida de la calidad de una revista.** Sin embargo, sí es razonable afirmar que existe una asociación entre el FI y la calidad de los artículos que se publican en la revista. Es decir, una publicación de FI 7 no es necesariamente mejor que una de FI 3, pero la calidad promedio de las publicaciones de FI 7 es superior al promedio de las de FI 3, al menos dentro de la misma área. Esta es una opinión generalizada en la comunidad biomédica<sup>4</sup>. Los motivos teóricos que sustentan esta asociación son varios. Por un lado está el hecho de que un investigador cita antes un trabajo que considera de mayor calidad que uno de menos. Es también muy probable que un investigador que quiere sustentar una afirmación con una referencia

cite antes un trabajo de una revista que considera mejor que un trabajo de otra que considera inferior. Por otro lado, el FI ejerce un mecanismo de retroalimentación positiva. Cuanto más alto es el FI de una revista, mayor es su prestigio, mayor es el deseo de los investigadores de publicar en ella, mayor es el número, y posiblemente la calidad, de los manuscritos que se envían, y con ello, mayor la capacidad que tiene el editor para elegir aquellos que considera de mejor calidad. Los evaluadores externos por su parte tienden a ser más estrictos si están juzgando un trabajo que va a ser publicado en una revista de mayor prestigio/impacto. Por todo ello, FI no es sinónimo de calidad **pero sí es un indicio de calidad de una revista.**

Si de evaluar el impacto o la repercusión de un trabajo de investigación se trata, el factor que debemos manejar es el número de citaciones de ese trabajo en concreto. Este parámetro que es muy dependiente del área y del año de publicación se puede normalizar en función de las citaciones medias que han recibido los trabajos publicados ese año en ese área científica. Por otro lado, la calidad de un trabajo es muy difícil de establecer de manera cuantitativa y objetiva. Sin embargo, si un trabajo se publica en una revista es sin duda porque ha pasado sus estándares de calidad habiendo superado la evaluación por pares y por el editor o el equipo editorial. Por tanto, si admitimos que el FI de la revista es un indicio de su calidad, podemos asumir, obviamente con reservas, que el FI es también un indicio de la calidad del trabajo.

Finalmente, los indicadores bibliométricos también permiten estimar la cantidad y la calidad del trabajo publicado de un investigador, grupo, departamento, centro o institución. La cantidad es fácil de estimar contando sencillamente el número de documentos. La valoración de la calidad, de nuevo, es más compleja. El número de citaciones de los trabajos, y mejor si están normalizados por área y año de publicación, nos daría un índice de la repercusión de esas publicaciones. El índice *h* de Hirsch, que se está haciendo también muy popular, nos indica el número de trabajos *h* que han sido citados *h* veces. Se puede calcular para un autor, grupo, departamento, etc., y le han salido también múltiples variantes que normalizan por número de autores en la publicación, por la edad o que dan más peso a los trabajos muy citados<sup>5</sup>. Se pretende que sea un indicador simultáneo de la calidad y la cantidad. Sin embargo, lo que de verdad refleja es la cantidad y la repercusión. En mi opinión, si asumimos que el único indicador verdadero de calidad de una publicación es la evaluación por pares de la revista y que la exigencia de los evaluadores es función del FI de la revista, el uso del FI acumulado (suma de los FI de todas las publicaciones) reflejaría mejor la cantidad y calidad de la producción científica de un investigador o grupo.

**La evaluación de la investigación es necesaria** y el resultado de la evaluación puede, y debe, influir en la distribución de los recursos y en la promoción de los investigadores. La evaluación debe ser objetiva, cuantitativa, fácil de realizar, rápida, barata y transparente. Los parámetros bibliométricos cumplen todos estos requisitos. Entre los detractores del uso de estos parámetros abundan aquéllos que abierta o encubiertamente prefieren que no se lleven a cabo las evaluaciones o prefieren las evaluaciones subjetivas realizadas por comisiones locales fácilmente influenciables. Hay también detractores *bona fide* que consideran que la calidad de la investigación sólo puede valorarse mediante el análisis pormenorizado por expertos del trabajo realizado. Aun estando de acuerdo con el principio, la puesta en práctica es difícil si no imposible, subjetiva y costosa, especialmente cuando se trata de evaluaciones a gran escala. Y la evaluación cuando es subjetiva está sujeta a conflicto de intereses y debe ser realizada por evaluadores externos.

En conclusión, el FI da una excelente medida de la visibilidad y del prestigio de una revista. No refleja de manera estricta la calidad pero es el indicador universal y objetivo que más se le puede aproximar. Las citaciones, el FI acumulado, el índice *h* o cualquier variante deben servir como parámetro fundamental en la necesaria evaluación de la actividad investigadora.

## En contra: Giampaolo Co'

La valoración de la actividad investigadora desarrollada por individuos, grupos y entes es una tarea extremadamente difícil y por ello, como en muchos otros ámbitos de la actividad humana, también en este campo se han buscado criterios cuantitativos que proporcionen resultados objetivos al respecto.

El paso final de la investigación científica pública es la presentación de los resultados obtenidos a la sociedad que, por otro lado, es la que suele proporcionar la financiación necesaria. Los métodos para hacer públicos estos resultados son variados, pero la publicación de un artículo en una revista científica, que lo somete al análisis de una comisión de expertos (revisión por pares), se considera como el mejor método para acreditar la validez científica del trabajo realizado. Por ello, la mayoría de los parámetros objetivos usados para valorar la investigación pública se centran en los artículos publicados en este tipo de revistas.

El parámetro más sencillo es el número de artículos: quien escribe más artículos es mejor investigador. Producir millares de artículos irrelevantes es mejor que desarrollar una investigación de gran impacto cultural limitada a unas pocas publicaciones. Si este fuera el criterio utilizado, las revistas tendrían interés en publi-

car la gran mayoría de los artículos que reciben, reduciendo lo más posible el nivel de exigencia de los pares. Pero en este parámetro, la calidad de lo publicado no juega ningún papel.

La búsqueda de un parámetro que tuviera en cuenta la relevancia científica de los artículos llevó a responsabilizar a la propia comunidad científica de la valoración de esta relevancia mediante la cita de la publicación. El número de citas se convirtió en el parámetro objetivo que definía la calidad de una investigación: a mí me citan más que a ti y, por tanto, soy mejor. Sin embargo, es fácil caer en problemas de *popularidad* como los que ocurren con, por ejemplo, los índices de audiencia de los programas de radio y televisión: un artículo puede ser muy citado, independientemente de su calidad. Un ejemplo reciente es el documento de la colaboración OPERA sobre los neutrinos que viajan a una velocidad mayor que la de la luz (<http://arxiv.org/abs/1109.4897>).

Una mejora respecto a este banal conteo de las citas consiste en valorar la calidad de las revistas donde el artículo se ha publicado. Para determinar esa calidad se utiliza el factor de impacto, FI (del inglés, *impact factor*). El cálculo del FI se lleva a cabo considerando el número de las citas anuales de la revista y dividiéndolo por el número total de artículos publicados. Una revista es tanto más prestigiosa cuanto más alto es este número. Publicar pocos artículos muy citados es mejor que publicar muchos pero menos citados.

Se pueden hacer muchas críticas al FI. Quiero mencionar aquí las dos que en mi opinión son más relevantes. En primer lugar, se trata de un índice dependiente de la disciplina científica. Es evidente que el número de científicos que componen una comunidad dada, que se autocita por definición, influye sobre el valor del FI. Por ejemplo, una comunidad de tres personas puede producir como máximo seis citas, mientras que una de cuatro investigadores, veinticuatro. Además, los tiempos de publicación y de desarrollo de la investigación son típicos de cada disciplina y, por tanto, también lo son las respuestas de confirmación o falsación de un nuevo artículo. Por ejemplo, la física teórica tiene tiempos de respuesta mucho más rápidos que los de la física experimental de altas energías; ésta se organiza en grupos de centenares de personas, mientras que los físicos teóricos trabajan solos o en grupos reducidos.

En segundo lugar, el FI es un valor medio de una distribución que no es normal: un único artículo muy citado desplaza significativamente el valor medio. Esto permite *trampear* sin muchas dificultades con el FI: en 2007 una revista científica que polemizaba sobre el (ab)uso del mismo escribió un editorial en el que autocitaba todos los artículos publicados por la propia revista del 2005 al 2006 provocando un aumento de su propio FI de 0.66 a 1.44.

Por lo que conozco, la última invención en términos de parámetros objetivos es el índice  $h$ , cuyo nombre se debe al apellido de su inventor, Jorge E. Hirsch, un físico de la universidad de California. Más allá de la complejidad de la formulación original, el método consiste en ordenar las publicaciones de una persona en sentido decreciente del número de citas. El índice  $h$  viene dado por la posición del artículo que coincide con su número de citas.

Este índice resuelve el problema de sobrevaloración de un currículum con un gran número de publicaciones irrelevantes, pero adolece de algunos aspectos negativos. Por ejemplo,  $h$  no tiene en cuenta el número de autores de un artículo. El comentario de más arriba acerca de físicos teóricos y experimentales pone en evidencia la dificultad de utilizar este parámetro en contextos amplios. Además, y en cualquier caso, este índice está ligado al número total de publicaciones. Un científico tan relevante por el impacto de sus obras como Galois tendría un  $h=2$  y en 1906, después de haber publicado sus tres famosos artículos sobre el movimiento browniano, el efecto fotoeléctrico y la relatividad restringida, Einstein habría tenido un  $h$  de 4 ó 5, insuficiente para ser admitido a un concurso de investigador del nivel más bajo en Italia.

Entre las distintas tareas que mi estado académico me impone, la que más detesto es formar parte de comisiones de concursos para seleccionar nuevos profesores o investigadores para la universidad. Junto a otros colegas debo confrontar personas que trabajan en contextos diferentes y juzgar cuál es la más adecuada para las tareas objeto del concurso. Utilizar parámetros objetivos sería un alivio y evitaría que la comisión asumiera la responsabilidad de la elección. Pero si tales criterios objetivos existieran, no sería necesaria una comisión: un simple ordenador bastaría para seleccionar a la persona más digna.

En Italia, el Ministerio de la Universidad y de la Investigación indica a las comisiones de concurso criterios objetivos a los que atenerse para expresar sus valoraciones. Un colega ha sintetizado este proceso diciendo que seremos forzados a juzgar currícula y no personas. Pero como suele ocurrir, las acciones preparadas por la burocracia centralizada no resuelven los problemas y, en compensación, crean otros.

El análisis cuantitativo es muy útil en algunos ámbitos del conocimiento, pero no puede ser generalizado. Un análisis cualitativo es esencial para poder estudiar estructuras complejas y no homogéneas. El análisis cualitativo es subjetivo y los sujetos en los que se delega el ejercicio de este análisis deben asumir las correspondientes responsabilidades. ¿Qué pensáis acerca de la hipótesis de seleccionar a los jugadores del equipo nacional de fútbol sólo mediante datos estadísticos (goles marcados, pases realizados, balones recibidos, paradas realizadas respecto al número de goles recibidos, etc.)? ¿Objetivo! ¿no?

## Bibliografía

1. Bordons M, Fernández MT, Gomez I. Advantages and limitations in the use of impact factor measures for the assessment of research performance. *Scientometrics* 2004;53:195-206.
2. Davies PM. Eigenfactor: Does the principle of repeated improvement result in better estimates than raw citation counts? *J Am Soc Inform Sci Technol* 2008;59:2186-8.
3. West J, Bergstrom T, Bergstrom CT. Big Macs and Eigenfactor Scores: Don't Let Correlation Coefficients Fool You. *J Am Soc Inform Sci Technol*. DOI: 10.1002/asi.21374.
4. Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? *J Med Libr Assoc* 2003;91:42-6.
5. Harzing, A.W. (2007) Publish or Perish, available from <http://www.harzing.com/pop.htm>