

Limitaciones en el uso de los indicadores bibliométricos para la evaluación científica

Durante los últimos años se ha evidenciado una creciente necesidad de evaluar los resultados de la actividad investigadora, tanto de países o regiones como de centros o grupos de investigación. Obtener métodos objetivos para evaluar la calidad de la investigación es un deseo compartido y perseguido por los gestores de política científica de todo el mundo. Sin embargo, y a pesar de los esfuerzos realizados, este deseo no está todavía satisfecho, al ser "la calidad de la investigación" una noción subjetiva, de difícil valoración.

ISABEL GÓMEZ CARIDAD Y MARÍA BORDONS GANGAS*

There has been a growing need in recent years for evaluating the results of research activity, both of countries and regions and of research centres or groups. Developing objective methods for evaluating the quality of research is a wish shared and pursued by the administrators of scientific policy the world over. Nevertheless, despite the efforts made, this wish has not yet been fulfilled, given that "quality of research" is a subjective notion, and difficult to assess.

Actualmente se admite que los propios investigadores de un área son los más capacitados (o quizá los únicos capacitados) para evaluar los resultados científicos que se producen en dicha área. Así, la evaluación por expertos es la que prima en distintos procesos del sistema científico, tales como el proceso de selección de los artículos que han de ser publicados por una revista, o de los proyectos que van a ser financiados por entidades u organismos. No obstante, este sistema de evaluación también presenta una serie de inconvenientes como son su alto coste (prohibitivo si se trata de evaluar la producción científica de todo un país) y la subjetividad inherente al juicio humano. En la búsqueda de otros indicadores que aporten objetividad al proceso de evaluación surgen los indicadores bibliométricos, basados en el análisis de las publicaciones científicas.

El uso de este tipo de indicadores con fines de evaluación es hoy una práctica común en los países más avanzados.

¿Qué son los indicadores bibliométricos?

Los indicadores bibliométricos son datos estadísticos deducidos de las publicaciones científicas. Su uso se apoya en el importante papel que desempeñan las publicaciones en la difusión de los nuevos conocimientos, papel asumido a todos los niveles del proceso científico. Estos indicadores son válidos en aquellos contextos en que los resultados de investigación dan lugar a publicaciones científicas, lo cual es habitual en las áreas científicas más básicas.

Los indicadores bibliométricos permiten manejar, clasificar y analizar grandes volúmenes de publicaciones científicas

también a las patentes, ya que el legítimo deseo de las industrias de extraer un rendimiento financiero de sus inversiones e investigación inhibe la publicación de artículos. La aplicación de los resultados de la investigación a la innovación industrial y su posible comercialización requiere el análisis de patentes y la utilización de otros indicadores relacionados con el mercado.

Los indicadores bibliométricos permiten manejar, clasificar y analizar grandes volúmenes de publicaciones científicas. A nivel internacional existen distintos grupos que elaboran estudios bibliométricos, bien para comparar la actividad científica de los distintos países, su evolución temporal y sus áreas de mayor y menor actividad —por ej. la *Information Science and Scientometrics Research Unit* (ISSRU) en Hungría, *Computer Horizons* en EE.UU.—, o bien para conocer la actividad del propio país a distintos niveles de profundidad y como instrumento para establecer la política científica nacional o europea —por ej. el *Observatoire des Sciences et des Techniques* (OST) de Francia, el *Center for Science and Technology Studies* (CWTS) holandés, la *Science Policy Research Unit* (SPRU) del Reino Unido—. Los análisis bibliométricos se han utilizado junto a otros indicadores con fines tan variados como establecer prioridades o identificar áreas de excelencia. De hecho, la bibliometría se ha convertido en una herramienta muy útil para los evaluadores y gestores de política científica, suplementa la opinión de expertos particularmente cuando se trata de muchos documentos cuya lectura individualizada no es factible. Así, son especialmente útiles para la evaluación de grandes unidades (por ej. un país, una universidad), y su fiabilidad disminuye al descender el tamaño de la unidad a analizar. El caso extremo sería el de la evaluación de autores individuales, situación en la que los indicadores deben de utilizarse (si se hace) con especial cautela (1), y siempre como complemento del juicio de expertos, que

Entre los indicadores más utilizados se puede señalar el número de publicaciones, el número de citas recibidas por las mismas, el factor de impacto de la revista de publicación, la vida media de las publicaciones, etc. Para valorar la investigación de las áreas aplicadas es importante recurrir

son los que han de valorar la influencia o importancia de la aportación científica tras un análisis cualitativo de las publicaciones.

Aunque los estudios bibliométricos han recibido numerosas críticas en los últimos años, la mayor parte de ellas no cuestionan el valor de la bibliometría, sino el uso inadecuado que se hace de ella (ej. 2, 3, 4). Dentro de la comunidad de bibliómetros —como en cualquier otra comunidad científica— tiene lugar un importante y constante debate sobre el uso y las limitaciones de los indicadores basados en publicaciones (por ej. se rechaza la equiparación entre calidad y factor de impacto, y se cuestiona el valor de las citas). Sin embargo, lo que constituyen premisas básicas y ampliamente aceptadas por los bibliómetros sobre cómo deben interpretarse determinados indicadores, en ocasiones no es conocido por los profesionales de otras áreas, que o bien utilizan los indicadores como instrumento de evaluación o bien son objeto de la misma.

A continuación se comentan algunos de los indicadores de más amplia difusión, haciendo especial énfasis en sus limitaciones y en las interpretaciones erróneas, que por exceso de simplificación, pueden cometerse al utilizarlos. Pretendemos con ello contribuir a una mayor comprensión por parte de todos del significado y validez de los indicadores bibliométricos, así como evitar incorrectas interpretaciones, a veces con importantes repercusiones en la carrera profesional de los investigadores.

¿Revistas nacionales o internacionales?

Los resultados de la investigación pueden ser publicados, según su tema y alcance, en revistas nacionales o internacionales. Las publicaciones internacionales están escritas preferentemente en lengua inglesa y su destino es la comunidad científica internacional, mientras que las revistas nacionales, destinadas a una audiencia local, suelen estar escritas en el idioma del país de origen.

Podría decirse que la ciencia básica es la ciencia "internacional" por excelencia y que, por tanto, tiende a transmitirse en revistas internacionales. Sin embargo, la ciencia aplicada, extendida como una explotación local de la ciencia básica, puede encontrar un buen vehículo de transmisión en las revistas de carácter nacional. Esto es aún más cierto para muchas áreas de ciencias sociales y humanidades, en las que se realiza investigación de interés claramente local, que afecta a un menor colectivo científico.

Por todo lo dicho, el argumento generalizador por el que tienden a considerarse de baja calidad las publicaciones nacionales de los países más periféricos peca de simplista. La causa de que muchas de estas revistas no estén incluidas en los índices internacionales es en ocasiones su interés local, y no su baja calidad. Las revistas científicas locales son importantes para el desarrollo de la actividad científica de un país, y facilitan la comunicación entre sus investigadores. El reto es fomentar e identificar las revistas locales de calidad, dentro de la gran masa de publicaciones que han proliferado en los últimos años.

En lo que se refiere a las revistas nacionales que abordan temas básicos de investigación, cada vez es más claro que su supervivencia depende de su calidad científica, y de su capacidad de atraer autores prestigiosos de la comunidad científica internacional.

¿Qué significa estar en el *Science Citation Index*?

La cobertura de una revista por bases de datos bibliográficas internacionales se considera un criterio de evaluación indirecta de la calidad de dicha revista, ya que su inclusión suele ir precedida de un examen de calidad. Aunque actualmente existen muchas bases de datos especializadas, hay que destacar por su amplio uso en bibliometría las producidas por el Institute for Scientific Information (ISI) de Filadelfia, y en especial el *Social Science Citation Index* (SSCI), en ciencias sociales, y el *Science Citation Index* (SCI), en ciencias experimentales, recogiendo este último cerca de 3.200 revistas de reconocido prestigio. Las principales razones de su amplio uso son su carácter multidisciplinar (pueden estudiarse todas las áreas consultando una sola base de datos), la presencia de nombre y dirección de todos los autores de los trabajos, y la elaboración anual por parte del ISI de datos estadísticos —con posible utilidad bibliométrica— sobre las revistas recogidas en sus bases de datos (*Journal Citation Reports*).

El hecho de que una revista aparezca recogida en la base de datos SCI se utiliza con frecuencia como un criterio de calidad en sí mismo, pero es importante saber cómo se realiza la selección de revistas, para ser conscientes de posibles desequilibrios en su cobertura. Aparte de aspectos formales (por ej. la puntualidad en la edición) y de la calidad intrínseca de la revista valorada por juicio de expertos, el número de citas recibidas por una revista es un elemento decisivo en la selección, considerado como un indicador del impacto o uso que la comunidad científica hace de una revista. Esto favorece a la comunidad científica angloparlante,

la más numerosa a nivel mundial, y da lugar a los siguientes sesgos:

- todos los países no están igualmente representados, se favorecen las publicaciones anglosajonas (en 1994 el SCI sólo recogía 6 revistas españolas, 3 de ellas escritas totalmente en lengua inglesa);
- la ciencia básica está mejor representada que la aplicada (esta última se cita menos);
- los temas de interés o alcance internacional están mejor representados que los temas de interés local (a éstos últimos suele corresponder un menor colectivo científico que actúa como posible citante).

Aunque la cobertura del SCI experimenta pequeñas variaciones a lo largo del tiempo, estos sesgos se perpetúan por la propia naturaleza del sistema de selección basado en citas.

Por todo lo dicho, utilizar la presencia en el SCI/SSCI como indicador de calidad en sí mismo es más adecuado para las áreas básicas de investigación y para países de lengua inglesa, y resulta especialmente inadecuado para las áreas aplicadas o para las ciencias sociales de países no angloparlantes, en las que los factores lingüísticos y culturales tienen mayor importancia. Hay que señalar que la priorización otorgada en algunos procesos de evaluación científica a las publicaciones en revistas ISI ha influido sobre los hábitos de publicación de los científicos, inclinándolos hacia esas revistas, con el consiguiente aumento de su visibilidad internacional, pero en perjuicio directo de las revistas nacionales.

¿Qué miden las citas?

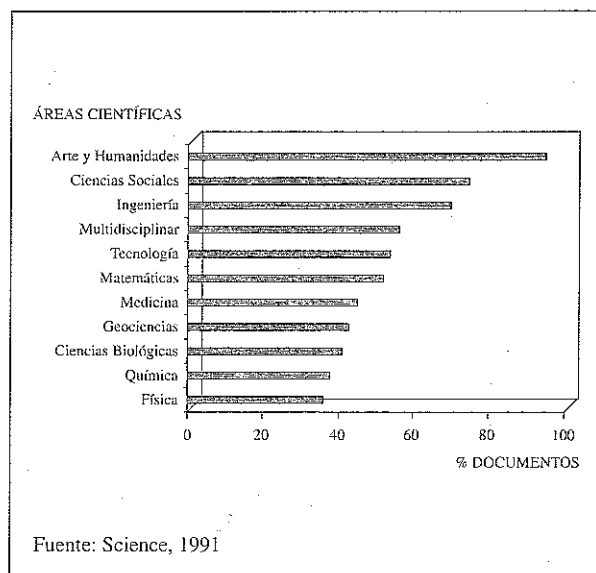
Aunque en un principio se consideró que el número de citas que recibía un documento podía ser indicativo de su calidad, hoy se acepta que el número de citas que mide impacto, uso o influencia de una investigación, es sólo un indicador parcial de la calidad del trabajo, y viene también determinado por otros muchos factores distintos a la calidad

El argumento generalizador por el que tienden a considerarse de baja calidad las publicaciones nacionales de los países más periféricos peca de simplista

(1,6). Si bien es cierto que los documentos "estrella", que introducen importantes descubrimientos o hallazgos, suelen recibir un alto número de citas, también lo es que un documento apenas citado no es necesariamente de mala calidad. Factores como el prestigio del autor, de su institución o país, o la actualidad de su tema de investigación intervienen distorsionando el valor de las citas como indicador de calidad.

Por otro lado, es importante señalar las grandes diferencias que existen en el número de citas según las áreas científicas. En este hecho influyen distintos factores, entre los cuales pueden mencionarse las diferencias en los hábitos de citación, según las áreas temáticas (distinto número medio de referencias por artículo), en el tamaño de la comunidad científica, o en el carácter más analítico o descriptivo de la literatura. Atendiendo a datos del ISI en 1991, es interesante comprobar que más de la mitad de los documentos de sus bases de datos no recibieron ninguna citación en los 5 años posteriores a su publicación (50 por 100 de los documentos del SCI, 75 por 100 de los recogidos en el SSCI y 95 por 100 de los del A&HCI —*Arts & Humanities Citation Index*—) (7). La figura 1 muestra el porcentaje de no citación en distintas áreas científicas.

Figura 1



Porcentaje de documentos del SCI no citados en los primeros cinco años de su publicación.

Una interpretación errónea de los datos anteriores nos podría llevar a plantearnos preguntas como: ¿Significa esto que el 45 por 100 no citado es ciencia de mala calidad, que no contribuye al progreso

científico? En caso afirmativo, ¿podría prescindirse de la misma, y financiar sólo a aquellos grupos que realizan investigación "de calidad", con la consiguiente reducción de los gastos en I+D? Sin embargo, ya hemos comentado las limitaciones de las citas como indicador de calidad. Los documentos no citados no son necesariamente documentos inútiles, que no contribuyen al avance científico. Entre otros factores, hay que tener en cuenta que la utilidad de algunos resultados científicos no puede apreciarse de forma inmediata, bien porque sobrepasa las fronteras de la propia área en la que se describió o porque todavía no existe la tecnología que permita aplicarlos.

¿Qué es el factor de impacto?

El factor de impacto (FI) es un indicador del número de citas que reciben en promedio los trabajos publicados en una determinada revista. Así, por ejemplo, el factor de impacto de la revista X en el año 1994 se calcula dividiendo el número de citas recibidas por los trabajos publicados en la revista X durante 1992 y 1993 (estas citas han de proceder de las revistas fuente del ISI en 1994), entre el número de trabajos publicados por la revista X en esos dos años. Este indicador fue introducido a finales de los años 70 por E. Garfield, y el ISI lo calcula anualmente para todas las revistas que recoge, junto a otros datos estadísticos.

El hecho de publicar en revistas de alto factor de impacto se considera en muchos contextos un criterio de calidad en sí mismo, pues se asume que un artículo publicado en determinada revista es probable que reciba el mismo número de citas que el promedio de la revista. Esta premisa puede ser válida en el análisis de un conjunto amplio de documentos (ej. producción de una universidad), pero lo es menos al descender a pequeñas unidades como la producción de un autor o aún menos si nos referimos a un artículo determinado. La razón es que la distribución de citas entre artículos de una revista es muy asimétrica y es fácil que un punto concreto difiera mucho del promedio calculado (factor de impacto).

¿Es comparable el FI entre disciplinas?

Los indicadores bibliométricos han permitido constatar la existencia de importantes diferencias en el desarrollo de la actividad investigadora según las áreas científicas. Variaciones en niveles de

producción, tasas de colaboración o tipo de documento predominante, indican en ocasiones las diferencias en el quehacer científico de distintas áreas o disciplinas.

En relación concreta al Factor de Impacto, se han observado importantes variaciones entre disciplinas, dependientes del distinto tamaño de las comunidades científicas, distintos hábitos de citación o variaciones en el ritmo de envejecimiento de la literatura según las áreas. Esto hace que no se puedan comparar factores de impacto de distintas disciplinas. A modo de ejemplo, la Tabla I muestra una selección de disciplinas científicas cubiertas por el SCI, con indicación del máximo factor de impacto obtenido por una revista en cada disciplina en 1994. También se indica en dicha tabla el número de revistas incluidas por el SCI en cada tema en dicho año. Se observa, por ejemplo, que la revista número 1 de Ingeniería Aeroespacial tiene un FI = 0,592, frente al FI = 4,166 de la primera de Cirugía, y el FI =

42,169 de la correspondiente a Bioquímica y Biología Molecular. Ni que decir tiene que esto no significa que la investigación realizada en Biología Molecular sea de mayor calidad que la de Cirugía, y ésta mejor que la de Ingeniería Aeroespacial.

Utilizar la presencia en el SCI como indicador de calidad en sí mismo es más adecuado para las áreas básicas de investigación y para países de lengua inglesa

Entre los factores que influyen en las variaciones de factor de impacto según las disciplinas, se puede mencionar el distinto ritmo de envejecimiento de la literatura, que hace que el factor de impacto calculado en base a las citas recibidas en un período de 2 años, no sea igualmente "justo" para todas las áreas o disciplinas científicas. Así, en algunas disciplinas como Biología Celular, Biología Molecular o Virología la literatura envejece muy rápidamente, y se citan fundamentalmente documentos muy recientes, mientras que en otras como Botánica, Anatomía o Ingeniería es frecuente citar documentos de 10 años de antigüedad. En la Fig. 2 se muestra la distribución anual de citas recibidas por una revista de Virología y otra de Telecomunicaciones, ambas son las de FI más alto de su especialidad. Se observa cómo el cálculo de impacto en 2 años tiene en cuenta el 28 por 100 de las citas recibidas por J-Virol y sólo el 8 por 100 de las recibidas por IEEE. Así, el FI de la primera es 6,4 veces superior al de la segunda. Si se considerase una "ventana de citación" de 5 años, los FI sólo diferirían un factor de 3,7. El hecho de considerar sólo citas a corto plazo sitúa a las revistas de Telecomunicaciones en clara "desventaja" respecto a las de Virología a la hora de calcular su factor de impacto.

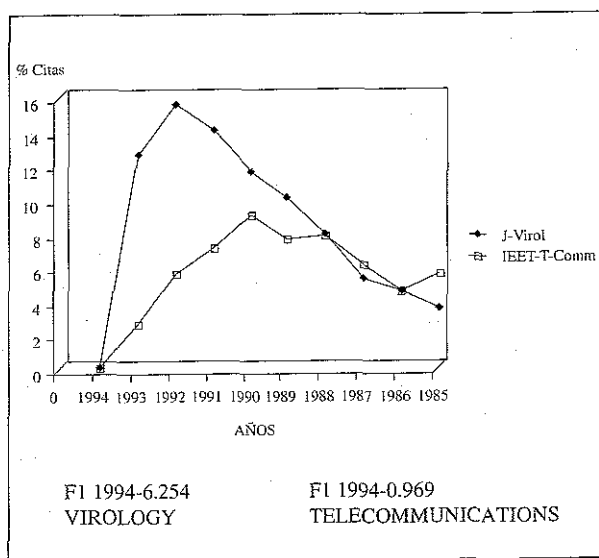
En general, las disciplinas con un FI más bajo en el SCI corresponden a especialidades aplicadas de las áreas de ingeniería, agricultura o informática; o a disciplinas muy pequeñas en las que la comunidad científica, posible citante, es muy reducida. Las disciplinas con mayor FI corresponden en general a ciencia básica y a colectivos muy numerosos. A medio camino entre ambas se situarían las disciplinas de medicina clínica.

Tabla I

Factor de impacto máximo y número de revistas de algunas disciplinas del Science Citation Index en 1994

Disciplina	FI máximo	N.º revistas
Bioquímica y Biol. Molecular	42,169	193
Inmunología	39,426	94
Genética y Herencia	22,568	73
Química Física	9,088	69
Física Partículas	5,471	13
Química Orgánica	4,906	33
Ecología	4,838	72
Ciencia de Materiales	4,792	81
Física Aplicada	4,792	52
Cirugía	4,166	93
Dermatología	3,826	27
Ingeniería Eléctrica y Electrónica	2,618	138
Anatomía y Morfología	2,483	13
Geología	2,394	59
Química Aplicada	2,279	28
Matemáticas	1,704	124
Agricultura	1,342	101
Ornitología	1,132	14
Horticultura	0,776	9
Ingeniería Aeroespacial	0,592	19

Figura 2



Diferencias en la antigüedad de la literatura citada en dos revistas de distintas disciplinas científicas.

Las comparaciones de FI dentro de una misma disciplina presentan mayor validez, aunque también se observan variaciones en función del carácter más básico o aplicado de las revistas. La investigación básica suele recibir mayor número de citas y sus revisas suelen tener mayor factor de impacto que las correspondientes a áreas aplicadas. Esto hace que incluso dentro de una disciplina SCI, las revistas centradas en los aspectos básicos obtengan mayor factor de impacto que las orientadas a los temas aplicados.

De esto deriva una posible repercusión negativa del uso del factor de impacto de forma indiscriminada para la evaluación del personal investigador, como es la baja consideración/valoración de aquellos investigadores dedicados a las áreas clínicas o aplicadas o de interés más local. Esto no sólo resulta claramente injusto, sino que además podría llevarles a abandonar su actividad y reorientarse hacia áreas básicas; al tiempo que desalienta a los jóvenes investigadores a elegir dichas áreas, todo lo cual puede ser altamente perjudicial para el interés general del país.

En resumen

Los indicadores bibliométricos utilizados adecuadamente, y en combinación con otros indicadores, constituyen una valiosa herramienta en la evaluación de la actividad científica de unidades de tamaño grande o medio. Dado su carácter estadístico,

su validez es mucho menor aplicados a pequeñas unidades como investigadores individuales o artículos concretos.

Para valorar la aportación de un trabajo al progreso científico no basta con acudir a los indicadores bibliométricos, sino que la información que éstos aportan tiene que ser complementada con el juicio de expertos. Es

importante evitar el empleo aislado de un indicador como baremo único, sin matizar con otra información y sin atender a las características propias de la disciplina en la que se aplica. La ciencia es una actividad muy compleja que, en ningún caso, puede quedar representada por un solo indicador. ■

* Centro de Información y Documentación Científica, CINDOC, CSIC.

Bibliografía

- ⁽¹⁾GARFIELD, E.; WELLJAMS-DOROF, A. Citation data: their use as quantitative indicators for science and technology evaluation and policy-making. *Science and Public Policy*, Vol. 19:5; 321-327, 1992.
- ⁽²⁾SEGLER, P.O. From bad to worse: evaluation by Journal Impact. *Trends in Biochemical Sciences*, 14 (8): 326-327, 1989.
- ⁽³⁾HANSSON, S. Impact factor as a misleading tool in evaluation of medical journals. *The Lancet*, Vol. 346, Sept. 30, p. 906, 1995.
- ⁽⁴⁾SAIZ-SALINAS, J. I. Failed professor. *Nature*, vol. 381, p. 186, 1996.
- ⁽⁵⁾GARFIELD, E. Is citation analysis a legitimate evaluation tool? *Scientometrics* 1(4): 359-375, 1979.
- ⁽⁶⁾MACROBERTS, M. H.; MACROBERTS, B. R. Problems of citation analysis. *Scientometrics* 36:3:435-444; 1996.
- ⁽⁷⁾HAMILTON, D. P. Research papers: who's uncited now? *Science* 251(25): 25. 1991.

Es importante evitar el empleo aislado de un indicador como baremo único, sin matizar con otra información y sin atender a las características propias de la disciplina en la que se aplica