

SOBRE LOS INDICADORES NUMÉRICOS

EN LA

EVALUACION CIENTIFICA

Grupo de Evaluación científica de la Real Academia de Ciencias Exactas, Físicas y Naturales (GRECA)¹

23 de Enero 2015

Objetivos del GRECA

Los países de nuestro entorno gastan cantidades muy importantes, en muchos casos crecientes, en investigación científica (IC). En un momento en el que la sociedad reclama, con mucha razón, una justificación clara del gasto público se hace necesaria una evaluación seria de dicha actividad (1, 2). En nuestro país, con mejor o peor fortuna, se han llevado a cabo, y se siguen efectuando, evaluaciones de la actividad científica. Así se han hecho y se hacen evaluaciones de personas, por ejemplo para la atribución de complementos de productividad, sexenios, por parte de la CNEAI- (3, 4). También evalúan proyectos – ej. para la financiación y seguimiento de los mismos por parte de la ANEP y/o del Ministerio financiador-, y de programas -en el seguimiento de los programas del Plan Nacional. Incluso se han efectuado, o se efectúan, evaluaciones de centros de investigación -como las llevadas a cabo por el CSIC en muchos de los suyos o también por determinados institutos de acuerdo con sus propias normas- (5). Se echa de menos sin embargo la evaluación de productividad científica de las universidades (6), como también la mayor transparencia en el resultado de las evaluaciones, en particular de las institucionales.

La Real Academia de Ciencias Exactas, Físicas y Naturales (RAC) <http://www.rac.es> es un organismo independiente y multidisciplinar, comparable a otras instituciones similares de los países más avanzados (<http://www.easac.eu>), cuyo objetivo fundamental está comprometido con el progreso científico, y en el caso de la RAC, con el avance de la IC en España. Entre otras cosas se espera de la RAC solvencia y neutralidad a la hora de proporcionar sus opiniones sobre la evolución de la propia ciencia o su discusión sobre enfoques y paradigmas, sobre las investigaciones más trascendentes que se publican o contribuciones críticas en relación con la IC. Pero además o tal vez por ello, la institución está capacitada para emitir juicios en materia de evaluación científica (*scientific assessment*) a cualquier nivel, bien contando con sus propias fuerzas, bien sirviéndose del concurso de otros colegas españoles o extranjeros a los que tiene fácil acceso. Para estimular esta tarea en el seno de la RAC se ha constituido un grupo de trabajo (GRECA) especialmente interesado en poner estas potencialidades de la Institución al servicio de demandas sociales.

La presente nota tiene como objetivo la presentación de este grupo de trabajo a la luz pública aprovechando para hacer algunos comentarios que estimulen el debate interno sobre las tareas del grupo, siempre presentes en el trabajo de la RAC. La publicación, al ser difundida a través

¹* José Bonet, Ana Crespo, Amable Liñán, Ramón Llamas, Carmen Nájera & Juan M. Rojo.

de la página web, podrá también abrir una conexión con colegas científicos y otros interesados del mundo externo a la Academia.

Los indicadores numéricos en la evaluación científica

No se quiere que éste sea un tema programático del trabajo del GRECA sino un ejemplo de cómo GRECA puede ofrecer opinión en asuntos de interés para la comunidad científica y que resultan muy oportunos en tiempos de escasez de fondos públicos y privados para la IC (2). De hecho se vive una situación de estrés que anima a proponer medidas que faciliten la mayor eficiencia en el gasto público y privado.

En particular los comentarios del presente documento se refieren al uso de indicadores tales como el índice de impacto de las publicaciones o el denominado índice-h de personas o instituciones. Estos indicadores intervienen hoy en la mayoría de las evaluaciones que se efectúan en materia de IC.

Hay que recalcar desde el principio que el análisis crítico de estos elementos no constituye una problemática de interés restringido a nuestro país, sino que es objeto de atención en todas las comunidades científicas desarrolladas (7, 8). De hecho aquí se quiere comentar específicamente el contenido de la denominada Declaración de San Francisco (San Francisco declaration on research assessment DORA, <http://am.ascb.org/dora/>) que ha suscitado cierto debate en diversos foros según puede apreciarse en Internet y redes sociales.

Elementos de contorno sobre los índices numéricos

Señalemos dos elementos de contorno en relación con estos indicadores. En primer lugar que ningún indicador numérico puede sustituir a la evaluación llevada a cabo de forma individualizada por colegas de nivel científico adecuado. Todo el sistema de evaluación descansa en la calidad del evaluador y en la confianza que se deposita en él. Por más que se discuta de técnicas de evaluación y de una formación específica para evaluar, esto suele conducir a establecer requisitos burocráticos y no a proporcionar mejores evaluaciones. Desde aquí se quiere insistir en que lo importante es la calidad del evaluador (sus condiciones personales incluyendo las éticas, que se dan por supuestas porque son un requisito *a priori*). La experiencia general indica que, quizás inconscientemente, es muy frecuente que el evaluador establezca *su propio perfil como umbral*. Por eso un evaluador cuyo perfil es medio no suele apreciar la excelencia. Completando este aspecto, que es crítico, digamos que el evaluador deberá estar a una distancia suficientemente lejana, geográfica e institucionalmente, del sujeto evaluado (persona o institución etc) mientras que su distancia en la temática científica es graduable dependiendo de los objetivos que persiga la evaluación.

En otras palabras, el sistema de revisión por pares y sus reglas de aplicación, sigue siendo la base de cualquier evaluación o valoración del trabajo científico (8).

Lo anterior no debe entenderse como una descalificación de los indicadores numéricos. De hecho, el segundo elemento de contorno en este discurso tiene que ver precisamente con la utilidad de dichos indicadores en el ámbito adecuado. El uso de los indicadores numéricos es una gran ayuda en los trabajos de evaluación científica, particularmente cuando se trata de ciencia básica o sus desarrollos o aplicaciones inmediatas. Es fundamental definir con precisión los objetivos de la evaluación porque son éstos los que definen o restringen cuales son los datos

a tener en cuenta o los indicadores que deben manejarse. Así la aplicación de un indicador numérico puede ser el único camino viable, por ejemplo en la evaluación de grandes colectivos, y naturalmente requiere, lo mismo que en el aspecto antes señalado, que sea aplicado juiciosamente también por evaluadores competentes (¡al final también deben ser pares!). Este último punto es importante: ninguna decisión relevante debe tomarse sobre la base de indicadores numéricos “enjuiciados por un ordenador”. La actuación final del evaluador humano cualificado es imprescindible.

Dicho esto, conviene resaltar que, en una sociedad donde la comunicación social y los medios de difusión son abrumadoramente abiertos y que requieren con frecuencia de gran inmediatez, disponer de índices numéricos, de fácil manejo, resulta naturalmente muy importante y más recomendable que el tanteo inexperto o que las opiniones basadas en estadísticas insuficientes o sesgadas. Creemos por ello que las simplificaciones del proceso evaluador que derivan del uso de los indicadores numéricos son extremadamente útiles y están colaborando en una mejor conformación de la opinión pública nacional e internacional. En este sentido es contraproducente elaborar indicadores cada vez más complejos, que traten de captar los innumerables “recovecos” de la actividad científica, en aras de una pretendida “justicia”. El tiempo y el esfuerzo empleados en un indicador extremadamente complejo pueden emplearse mucho más eficazmente en una evaluación directa como la del punto 1.

Sobre el índice de impacto (IF) de las revistas

La aplicación y generalización de este índice ha sido un enorme avance en la política científica, aunque pueda ser blanco de discusiones, algunas interesantes, en el marco de la evaluación de políticas y actuaciones más complejas (8, 10). Particularmente útil es el Journal of Impact Factor de Thomson Reuters (<http://ip-science.thomsonreuters.com/citationimpactcenter>) porque parte de un filtro previo en las citas recogidas, y esta premisa proporciona una mayor confianza aunque ello pueda conducir a que este índice sea más restrictivo en ciertos campos del saber. Cuando un investigador cree haber obtenido un resultado interesante es natural que pretenda que éste sea conocido por el mayor número posible de colegas y por eso le interesa que se haga público en una revista que sea leída por muchos. Aunque a priori no puede saber si su resultado interesará lo bastante, el IF de una revista avanza una estimación del número de sus lectores posibles.

De manera indirecta el IF es también ilustrativo sobre el tamaño de las comunidades que trabajan en los distintos campos lo cual resulta también de utilidad en los análisis de prospectiva científica y de la ciencia misma. Dadas las inabordables dimensiones de la producción investigadora, la revista en la que se publica ya es un indicador de los grados de filtrado que tienen los resultados científicos. Esto simplifica la información, lo que es evidentemente positivo. Anécdotas aparte: es obvio que hay errores en el proceso de selección que hacen los editores, pero también está bastante claro que si éstos errores dejan de ser esporádicos, disminuirá inexorablemente el prestigio de la revista.

Al paso de las consideraciones generales que hace la DORA sobre el origen de este indicador no parece relevante el que haya sido ingeniado con fines diferentes a la evaluación científica. Tantas y tantas cosas en la historia, en la ciencia y en la vida cotidiana han resultado de suma utilidad con fines diferentes a los previstos inicialmente. Es importante, eso sí, que el IF se aplique con cordura y sin perder de vista los objetivos que se quieren alcanzar en cada acto de

evaluación (11). Los campos científicos que son practicados por comunidades científicas muy grandes alcanzarán IF mayores. Y también es cierto que el proceso investigador produce resultados de distinto grado de interés intrínseco que el investigador orienta a colectivos más especializados y que por lo general se vehiculizan en revistas de menor IF.

Dicho esto, es cierto como dice la DORA, que se producen sesgos en las citaciones que favorecen a algunas revistas; por lo general a las más citadas. Pero parece lógico que ello ocurra, al menos en cierta medida, porque el lector les pone confianza. ¿Es o no comprensible que un investigador quiera respaldar sus discusiones o conclusiones en referencias muy fidedignas?; más raro sería que se fundamentase en las aparecidas en medios recónditos o menos filtrados.

Es cierto también, como se dice en la declaración, que los valores del IF son significativos solo cuando se aplican en contextos o campos suficientemente homogéneos. Pero, sin exagerar... El extremo indeseable sería que, mediante el IF de las revistas en que publican, solo se pudieran comparar investigadores que hagan exactamente lo mismo.

También es posible que, como dice la DORA, determinadas políticas editoriales lleguen a conseguir inflar oportunamente los índices que obtengan sus revistas, pero esto nos parece una situación que no es generalizable.

Es decir que, a grandes rasgos seríamos nosotros menos críticos con los IF que el preámbulo de la declaración DORA.

Sobre el índice h

Así como el IF proporciona una información a priori sobre el impacto que un artículo científico podrá tener, el h informa del grado de interés con que se han recibido los artículos de un científico o una institución. Hay mucha literatura reciente resaltando las luces y sombras de este indicador (10, 11) pero es un hecho que ha tenido un éxito espectacular. Es evidentemente un instrumento ingenioso y elegante por lo sencillo. En el extremo califica la obra de una vida con un número. Es una información muy inmediata y transparente que permite ser traducida al modo de juicio absoluto muy del gusto de la sobresimplificación que seduce a los medios de comunicación y a los planificadores inexpertos. No es exagerado afirmar que es un índice no exento de “morbo”.

Parece obvio que está resultando un instrumento de gran utilidad en múltiples procesos de evaluación. También es evidente que solo proporciona indicios o datos fehacientes cuando mide períodos de actividad suficientemente largos, que son variables sobre todo en relación con el tipo de artículo y el campo de trabajo. Y por supuesto proporciona un número que solo puede servir para comparar en el seno de comunidades científicas suficientemente homogéneas.

Entre las críticas o reservas que se señalan en la DORA, probablemente la falta de análisis en la tipificación de los artículos (múltiples autorías, contenidos como metodologías o revisiones) es la más importante y efectivamente puede proporcionar una imagen deformada de un investigador. Por otro lado, dado que el número se ve fuertemente influido por el tamaño de las comunidades científicas, es complejo deducir una calificación si no se tiene en cuenta este

elemento. Como consecuencia el uso del índice h puede ser engañoso o poco informativo cuando no se pueden precisar valores medios en los distintos campos.

Conclusiones

1. Ningún indicador numérico puede sustituir a la evaluación llevada a cabo de forma individualizada por colegas de nivel científico adecuado.
2. El uso de los indicadores numéricos es una ayuda fundamental en los trabajos de evaluación científica, particularmente cuando se trata de ciencia básica o sus desarrollos o aplicaciones inmediatas. Muy especialmente estos indicadores pueden resultar idóneos para comparaciones masivas, permitiendo establecer umbrales entre comunidades afines.
3. Los indicadores deben ser simples, tanto en su obtención como en su comprensión por parte del entorno.
4. Los indicadores numéricos siempre encierran una información científica relevante, que rara vez es ociosa para el análisis emprendido. Sin embargo, en la mayoría de los casos debe complementarse con información de otro tipo. Muchas de las críticas habituales a los indicadores numéricos provienen de ignorar este último punto.
5. La fiabilidad de los indicadores numéricos disminuye con el tamaño de la muestra: son mucho más fiables para valorar un centro que para valorar individualmente a sus investigadores.

Referencias bibliográficas

1. House of Commons, Science and Technology Committee. 2011. Peer review in scientific publications. Eighth Report of Session 2010–12. Volume I: Report, together with formal minutes, oral and written evidence. London: The Stationery Office Limited. 254pp.
2. Research Assessment Exercise - 2014.
http://en.wikipedia.org/wiki/Research_Assessment_Exercise
3. Crespo A. 2005. Luz y taquígrafos en la evaluación. El País: Sociedad, Futuro.Circuito Científico. Pag 34. 02.03.2005.
4. Agraït N & Poves A 2009. Sobre los resultados de las evaluaciones de la CNEAI, la situación en 2009. CNEAI www.mecd.gob.es/dctm/ministerio/.../cneai/2009-info-v5.pdf
5. Rojo JM. 1999. Peer review—Experiences at national and European level. *Scientometrics*, 45 (3): 497-500. DOI-10.1007/BF02457616.
<http://www.akademai.com/content/HRH18667301X5050>
6. Miras-Portugal MT. 2013 Propuestas para la reforma y mejora de la calidad y eficiencia del sistema universitario español.
www.mecd.gob.es/prensa.../2013/02/20130215-comision-expertos.html
7. Fenner M. 2013. What Can Article-Level Metrics Do for You? *PLoS Biol* 11(10): e1001687. doi:10.1371/journal.pbio.1001687
8. Garwood J. 2011. A conversation with Peter Lawrence, Cambridge: “The Heart of Research is Sick”. *LAB Times. Analysis*. 2-2011: 24-31.

9. Rodríguez Navarro A 2009. Sound Research, Unimportant Discoveries: Research, Universities, and Formal Evaluation of Research in Spain. *Journal of the American Society for Information Science and Technology*, 60(9):1845–1858.
10. Imperial J, Rodríguez-Navarro A. 2007. Usefulness of Hirsch's h-index to evaluate scientific research in Spain. *Scientometrics*, 71 (2): 271–282. DOI: 10.1007/s11192-007-1665-4
11. Yong A. 2014. Critique of Hirsch's citation index: a combinatorial Fermi problem. *Notices of the AMS*, 61 (9): 1040-1050