

El manifest de Leiden sobre indicadors d'investigació^a

Diana Hicks,^b Paul Wouters,^c Ludo Waltman,^c Sarah de Rijcke^c i Ismael Rafols^{d,e,f}

Les dades sobre les activitats científiques estan essent cada vegada més utilitzades per a governar la ciència. Avaluacions sobre investigació que van ser en el seu dia dissenyades individualment pel seu context específic i realitzades per parells, ara són rutinàries i estan basades en mètriques.¹ El problema és que l'avaluació va passar d'estar basada en valoracions d'experts a dependre d'aquestes mètriques. Els indicadors han proliferat: normalment benintencionats, no sempre ben informats, i sovint mal aplicats. Quan organitzacions sense coneixement sobre bones pràctiques i interpretació apropiada d'indicadors duen a terme les avaluacions, correm el risc de danyar el sistema científic amb els mateixos instruments dissenyats per a millorar-les.

Abans de l'any 2000, els experts utilitzaven el Science Citation Index de l'Institute for Scientific Information (ISI), en la versió de CD-ROM, per a realitzar anàlisis especialitzades. El 2002, Thomson Reuters va llançar una plataforma web integrada que va fer accessible a un públic ampli la base de dades Web of Science. Després van aparèixer altres índexs de citacions que es van erigir en competència de Web of Science: Scopus d'Elsevier (2004) i Google Acadèmic (versió beta creada el 2004). Després es van introduir instruments basats en el web, com ara Incites (que fa servir Web of Science) i SciVal (que usa Scopus) i també programari per a analitzar perfils individuals de citacions basats en Google Acadèmic (Publish or Perish, que va aparèixer el 2007).

El 2005, Jorge Hirsch, un físic de la Universitat de Califòrnia a San Diego, va proposar l'índex-h, que va popularitzar el recompte de citacions d'investigadors individuals. L'interès en el factor d'impacte de les revistes acadèmiques va créixer incessantment

^a. Traducció al català de Hicks *et al.* (2015). «[The Leiden Manifesto for research metrics](#)». *Nature*, 520, 429-431.) ([pàgina web del manifest](#)).

^b. School of Public Policy, Georgia Institute of Technology, Atlanta, Estats Units d'Amèrica.

^c. Centre for Science and Technology Studies (CWTS), Universitat de Leiden, Països Baixos.

^d. Ingenio (CSIC-UPV), Universitat Politècnica de València, València, Espanya.

^e. Science Policy Research Unit (SPRU), Universitat de Sussex, Brighton, Regne Unit.

^f. Observatoire des Sciences et des Techniques (OST-HCERES), París, França.

des del 1995. Recentment, han aparegut mesures d'ús social i de comentaris en línia: F1000Prime es va establir el 2002; Mendeley, el 2008, i Altmetric.com, el 2011.

Com a investigadors de cièntometria, científics socials i gestors de recerca hem observat amb creixent preocupació un ús incorrecte generalitzat dels indicadors en l'avaluació de l'acompliment científic. Els següents són alguns dels nombrosíssims exemples possibles. A tot el món, les universitats s'han obsessionat amb la seva posició en els rànquings globals (com ara el rànquing de Xangai i la llista del *Times Higher Education*), quan aquestes llistes estan basades en el que a parer nostre són dades inexactes i indicadors arbitraris.

Algunes organitzacions demanen l'índex-h als candidats que es presenten a ofertes d'ocupació. Diverses universitats basen la promoció en valors llindar de l'índex-h i en el nombre d'articles en revistes d'«alt impacte». Els currículums s'han convertit en oportunitats de presumir d'aquestes «puntuacions», en particular en biomedicina. A tot arreu, els supervisors demanen prematurament als seus estudiants de doctorat que publiquin en revistes d'alt impacte i que aconseguixin finançament extern.

A Escandinàvia i a la Xina, algunes universitats distribueixen fons d'investigació o bonificacions sobre la base d'un nombre: per exemple, calculant puntuacions individuals d'impacte per a repartir «recursos d'acompliment», o donant als investigadors una prima per a publicacions en una revista amb un factor d'impacte superior a 15.²

Per aquestes raons, vam presentar el *Manifest de Leiden*, que rep aquest nom de la conferència on va cristal·litzar. Els seus deu principis no són cap novetat per a experts en cièntometria, però cap de nosaltres no seria capaç de recitar-los en la seva totalitat, ja que fins ara no s'havien codificat. Celebritats en cièntometria, com Eugene Garfield (fundador de l'ISI), ja han presentat de vegades alguns d'aquests principis,³ però no poden estar presents quan els avaluadors informen gestors universitaris que no són experts en la metodologia pertinent. Els científics que busquen literatura per a disputar o impugnar avaluacions només troben les informacions necessàries en el que són, per a ells, revistes opaques i de difícil accés.

Oferim aquesta síntesi de bones pràctiques en avaluació basada en indicadors mètrics perquè els investigadors puguin demanar comptes als avaluadors, i perquè els avaluadors puguin demanar comptes als indicadors.

DEU PRINCIPIS

1. L'avaluació quantitativa ha de donar suport a la valoració qualitativa per experts

Els indicadors poden corregir la tendència a perspectives esbiaixades que es donen en la revisió per parells i facilitar la deliberació. En aquest sentit, els indicadors poden enfortir l'avaluació per parells, ja que prendre decisions sobre col·legues és difícil sense diverses fonts d'informació. No obstant això, els avaluadors no han de cedir a la temptació de supeditar les decisions als nombres. Els indicadors no poden substituir els raonaments informats. Els decisors tenen plena responsabilitat sobre les seves avaluacions.

2. L'exercici ha de ser mesurat d'acord amb les missions de recerca de la institució, grup o investigador

Els objectius d'un programa d'investigació s'han d'especificar al principi, i els indicadors usats per a mesurar-ne l'acompliment han d'estar clarament relacionats amb aquests objectius. L'elecció i els usos dels indicadors han de tenir en compte els contextos socioeconòmics i culturals. Els científics tenen diverses missions d'investigació. La investigació per a avançar les fronteres del coneixement acadèmic és diferent de la recerca focalitzada a proveir solucions per a problemes socials. L'avaluació pot estar basada en mèrits rellevants per a la indústria, el desenvolupament de polítiques, o per als ciutadans en general, en comptes de mèrits basats en nocions acadèmiques d'excel·lència. No hi ha un model d'avaluació que es pugui aplicar en tots els contextos.

3. L'excel·lència en investigació de rellevància local ha de ser protegida

En moltes parts del món, excel·lència en investigació s'associa únicament amb publicacions en anglès. La llei espanyola, per exemple, explicita el desig i la conveniència que els acadèmics espanyols publiquin en revistes d'alt impacte. El factor d'impacte es calcula per a revistes indexades per Web of Science, que és una base de dades basada en els Estats Units i que conté una gran quantitat de revistes en anglès. Aquests biaixos són especialment problemàtics en les ciències socials i les humanitats, àrees en les quals la investigació està més orientada a temes regionals i nacionals. Molts altres camps científics tenen una dimensió nacional o regional, per exemple l'epidemiologia del VIH a l'Àfrica subsahariana.

Aquest pluralisme i la rellevància social tendeixen a ser suprimits quan es creen articles d'interès als guardians de l'alt impacte: les revistes en anglès. Els sociòlegs espanyols

molt citats a Web of Science han treballat en models abstractes o estudiant dades dels Estats Units. En aquest procés es perd l'especificitat dels sociòlegs amb alt impacte en les revistes en espanyol: temes com la llei laboral local, atenció mèdica per a gent gran o ocupació d'immigrants.⁴ Indicadors basats en literatura d'alta qualitat no anglesa servirien per a identificar i recompensar l'excel·lència en investigació localment rellevant.

4. Els processos de recopilació i anàlisi de dades han de ser oberts, transparents i simples

La construcció de les bases de dades necessàries per a avaluar ha de seguir processos establerts abans que la investigació sigui completada. Aquesta ha estat la pràctica comuna entre els grups acadèmics i comercials que han desenvolupat metodologies d'avaluació durant diverses dècades. Aquests grups van publicar els protocols de referència en la literatura revisada per parells. Aquesta transparència permet l'escrutini i el control dels mètodes. Per exemple, el 2010, un debat públic sobre les propietats tècniques d'un important indicador utilitzat per un dels nostres grups (el Centre for Science and Technology Studies (CWTS) de la Universitat de Leiden, als Països Baixos), es va saldar amb una revisió en el càlcul d'aquest indicador.⁵ Les noves empreses comercials en el camp han de respondre als mateixos estàndards. Ningú no ha d'acceptar avaluacions automàtiques sortides de caixes negres o processos impenetrables. La simplicitat és una virtut en un indicador perquè afavoreix la transparència. Però indicadors simplístics poden distorsionar l'avaluació (vegeu el principi 7). Els avaluadors s'han d'esforçar a trobar un equilibri: indicadors simples que siguin respectuosos amb la complexitat dels processos d'investigació descrits.

5. Les dades i les anàlisis han d'estar obertes a la verificació per part dels avaluats

Per tal d'assegurar la qualitat de les dades, els investigadors inclosos en estudis bibliomètrics han de poder comprovar que les seves contribucions han estat correctament identificades. Els responsables i gestors dels processos d'avaluació han de garantir l'exactitud de les dades usades mitjançant mètodes d'autoverificació o auditoria per terceres parts. Les universitats podrien implementar aquest principi en els seus sistemes d'informació. Aquest hauria de ser un principi rector en la selecció de proveïdors d'aquests sistemes. La compilació i el procés de dades d'alta qualitat, precisos i rigorosos, requereixen temps i costen diners. Els responsables han d'assignar pressupostos a l'altura d'aquestes necessitats de qualitat.

6. Les diferències en les pràctiques de publicació i citació entre camps científics s'han de tenir en compte

La millor pràctica en avaluació és proposar una bateria d'indicadors i deixar que els diferents camps científics escullin els indicadors que millor els representen. Fa uns quants anys, un grup d'historiadors va rebre una puntuació relativament baixa en una avaluació nacional de parells perquè escrivien llibres en comptes d'articles en revistes indexades per Web of Science. Aquests historiadors van tenir la mala sort de formar part del departament de psicologia. L'avaluació d'historiadors i científics socials requereix la inclusió de llibres i literatura en la llengua local; l'avaluació d'investigadors en informàtica necessita considerar les contribucions a les conferències.

La freqüència de citació varia segons els camps: les revistes més citades en rànquings de matemàtiques tenen un factor d'impacte al voltant de 3; en canvi, les revistes més citades en rànquings de biologia cel·lular tenen factors d'impacte al voltant de 30.

Per tant, es necessiten indicadors normalitzats per camp, i el mètode més robust de normalització està basat en percentils: cada publicació es pondera segons el percentil al qual pertany a la distribució de citacions del seu camp (per exemple, el percentil 1%, 10%, 20% més alt). Una única publicació altament esmentada millora una mica la posició d'una universitat en un rànquing basat en percentils, però pot propulsar la universitat d'un lloc mitjà a les primeres posicions en un rànquing basat en mitjanes de citacions.⁶

7. L'avaluació individual d'investigadors s'ha de basar en la valoració qualitativa del seu dossier d'investigació

L'índex-h augmenta amb l'edat de l'investigador, encara que aquest ja no publiqui. L'índex-h també varia per camps: els científics en les ciències de la vida poden arribar a 200; els físics, a 100, i els científics socials, a 20 o 30.⁷ És un índex que depèn de la base de dades: hi ha informàtics que tenen un índex-h de 10 a Web of Science, però de 20 o 30 a Google Scholar.⁸ Llegir i valorar el treball d'un investigador és molt més apropiat que confiar en un únic número. Fins i tot quan es comparen un gran nombre de científics, és millor adoptar un enfocament que consideri informació diversa sobre cada individu, incloent-hi el seu coneixement, experiència, activitats i influència.

8. S'ha d'evitar la concreció improcedent i la falsa precisió

Els indicadors de ciència i tecnologia tendeixen a l'ambigüitat conceptual i a la incertesa, i es fonamenten en hipòtesis que no estan universalment acceptades. Per aquesta raó, les bones pràctiques fan servir múltiples indicadors per a construir un retrat robust i plural. En la mesura que sigui possible quantificar-la, la informació sobre incertesa i error hauria d'acompanyar els valors dels indicadors publicats, per exemple usant barres d'error. Si això no fos possible, els productors d'indicadors haurien d'evitar, almenys, d'oferir un fals nivell de precisió. Per exemple, el factor d'impacte de revistes es publica amb tres decimals per a evitar empats. No obstant això, atesa l'ambigüitat conceptual i la variabilitat aleatòria de les citacions, no té sentit distingir entre revistes per petites diferències en el factor d'impacte. S'ha d'evitar la falsa precisió: només un decimal està justificat.

9. Cal reconèixer els efectes sistèmics de l'avaluació i els indicadors

Els indicadors canvien el sistema científic per mitjà dels incentius que estableixen. Aquests efectes haurien de ser anticipats. Això significa que una bateria d'indicadors és sempre preferible, ja que un sol indicador és susceptible de generar comportaments estratègics i substitució d'objectius (segons la qual la mesura es converteix en un fi). Per exemple, en els anys noranta, Austràlia va finançar investigació en universitats d'acord amb una fórmula basada sobretot en el nombre de publicacions d'un institut. Les universitats podien calcular el «valor» d'una publicació en una revista arbitrada; l'any 2000, el valor es va estimar en 800 dòlars australians (480 dòlars nord-americans), destinats a recursos d'investigació. Com era d'esperar, el nombre d'articles publicats per autors australians va augmentar, però en revistes menys citades, fet que suggereix que la qualitat dels articles va disminuir.⁹

10. Els indicadors s'han d'examinar i actualitzar periòdicament

Les funcions de la investigació i els objectius de l'avaluació canvien o es desplacen, i el sistema de recerca coevoluciona amb ells. Mesures que van ser útils en el seu dia passen a ser inadequades i apareixen nous indicadors. Per tant, els sistemes d'indicadors s'han de revisar i, potser, modificar. En adonar-se dels efectes de la seva fórmula simplista d'avaluació, el 2010 Austràlia va adoptar la iniciativa Excellence in Research for Australia, que és més complexa i posa èmfasi en la qualitat.

Passos següents

Essent fidel a aquests deu principis, l'avaluació de la recerca pot tenir un paper important en el desenvolupament de la ciència i les seves interaccions amb la societat. Els indicadors de recerca poden proporcionar informació crucial que seria difícil d'aglutinar o entendre a partir d'experiències individuals. Però no s'ha de permetre que la informació quantitativa es converteixi en un objectiu. Les millors decisions es prenen combinant estadístiques robustes sensibles als objectius i la naturalesa de la investigació avaluada. Tant l'evidència quantitativa com la qualitativa són necessàries; cadascuna és objectiva a la seva manera. Les decisions sobre la ciència s'han de prendre d'acord amb processos d'alta qualitat informats per dades de la qualitat més gran.

1. WOUTERS, P. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (ed. Cronin, B. & Sugimoto, C.), (MIT Press, 2014), p. 47-66.
2. SHAO, J. & SHEN, H. *Learned Publishing*, 24 (2011), p. 95-97.
3. SEGLEN, P. O. *Br. Med. J.*, 314 (1997), p. 498-502 . GARFIELD, E. *J. Am. Med. Assoc.*, 295 (2006) p. 90-93.
4. LÓPEZ PIÑEIRO, C. & HICKS, D. *Res. Eval.*, 24 (2015), p. 78-89.
5. VAN RAAN, A. F. J., VAN LEEUWEN, T. N., VISSER, M. S., VAN ECK, N. J. & WALTMAN, L. J. *Informetrics*, 4 (2010), p. 431-43.
6. WALTMAN, L. *et al. J. Am. Soc. Inf. Sci. Technol.*, 63 (2012), p. 2419-2432.
7. HIRSCH, J. E. *Proc. Natl Acad. Sci. USA* 102 (2005), p. 16569-16572.
8. BAR-ILAN, J. *Scientometrics*, 74 (2007), p. 257-271.
9. BUTLER, L. *Res. Policy*, 32 (2003), p. 143-155.