



Inici | Presentació | Nou | Instruccions autors | Call for papers | Indexada a | Arxiu

Cerca (<http://temaria.net/simple.php?origen=1575-5886&idioma=ca>)

Per què fallen els indicadors bibliomètrics: paràmetres inestables, models incorrectes i propietats irrelevantes

[English version]

① JORDI MOLAS-GALLART
INGENIO (CSIC-UPV)
Universitat Politècnica de València
jormoga@ingenio.upv.es
① ISMAEL RÀFOLS
INGENIO (CSIC-UPV)
Universitat Politècnica de València
i.rafols@ingenio.upv.es

Opcions

<meta />

Metadades

DOI: <https://dx.doi.org/10.1344/BiD2018.40.22>

Citació recomanada

Molas Gallart, Jordi; Ràfols, Ismael (2018). "Per què fallen els indicadors bibliomètrics : paràmetres inestables, models incorrectes i propietats irrelevantes". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 40 (juny) . <<http://bid.ub.edu/40/molas.htm>>. DOI: <http://dx.doi.org/10.1344/BiD2018.40.22> [Consulta: 21-06-2018].

Resum

La generalització dels indicadors bibliomètrics en l'avaluació de la recerca i en el disseny de polítiques ha anat acompanyada d'una percepció cada cop més estesa segons la qual l'ús d'aquestes eines sovint és problemàtic o inapropiat. En aquest article presentem un marc senzill que permet analitzar les condicions en què l'ús d'indicadors com a mètode d'avaluació pot fallar. Els motius d'aquests errors poden ser tres. En primer terme, pot ser que els paràmetres dels models que relacionen les propietats i els indicadors siguin inestables i, per tant, que els indicadors no siguin comparables en el temps o en l'espai. En segon lloc, pot ser que els models emprats siguin incorrectes. Finalment, pot ser que la propietat de l'indicador sigui irrellevant en relació amb l'objecte d'estudi. Aquest marc que proposem pot resultar especialment útil per a comprendre els problemes que representa l'ús d'indicadors convencionals en espais "perifèrics", és a dir, en àrees geogràfiques, llengües o disciplines que se situen en els marges del sistema científic.

Resumen

La generalización de los indicadores bibliométricos en la evaluación de la investigación y en el diseño de políticas ha ido acompañada de una percepción cada vez más generalizada según la cual el uso de estas herramientas a menudo es problemático o inapropiado. En este artículo presentamos un marco sencillo que permite analizar las condiciones en las que el uso de indicadores como método de evaluación puede fallar. Los motivos de estos errores pueden ser tres. En primer lugar, puede ocurrir que los parámetros de los modelos que relacionan las propiedades y los indicadores sean inestables y, por lo tanto, que los indicadores no sean comparables en el tiempo o en el espacio. En segundo lugar, puede pasar que los modelos empleados sean incorrectos. Finalmente, puede suceder que la propiedad del indicador sea irrelevante en relación con el objeto de estudio. Este marco que proponemos puede resultar especialmente útil para comprender los problemas que representa el uso de indicadores convencionales en espacios "periféricos", es decir, en áreas geográficas, lenguas o disciplinas que se sitúan en los márgenes del sistema científico.

Abstract

The spread of bibliometric indicators in research evaluation and policy has been accompanied by an increasing realisation that indicators' use is often problematic and/or inappropriate. In this article, we propose a parsimonious framework to analyse the conditions under which the use of indicators become problematic. We propose that indicators' use in evaluation can breakdown for three reasons. First, because the parameters of the models linking properties and indicators are unstable and, as a consequence, indicators cannot be compared across spaces or time. Second, because the underlying models are incorrect. Third, because the property of the indicator is irrelevant for the issue examined. This framework can be particularly helpful in fostering an understanding of the problems of conventional indicators in "peripheral" spaces – i.e. in geographical, linguistic or disciplinary areas that are in the margins of the science system.

Paraules clau: Avaluació científica, Indicadors bibliomètrics

Keywords: Scientific evaluation, Bibliometric indicators

1 Introducció

Les pràctiques de comunicació científica es troben en un estat d'evolució constant. El desenvolupament d'Internet ens ha permès accedir des de qualsevol lloc a una infinitat de fonts d'informació com mai abans no havia estat possible. En conseqüència, els mitjans de comunicació han canviat, alhora que també ho han fet, i de manera significativa, els usos que podem atribuir a aquests mitjans. Per exemple, la capacitat que tenim actualment d'accedir i analitzar milions d'articles de publicacions científiques ha fet de l'article electrònic, i del contingut informacional com a metadades, una font d'informació que podem utilitzar amb finalitats d'anàlisi i avaluació. La proliferació d'indicadors basats en publicacions i citacions han convertit el que abans era un mer canal de comunicació en els fonaments d'un complex sistema de mètriques d'avaluació. Així mateix, com que els articles i altres mitjans de publicació electrònics han passat a fer-se servir amb altres finalitats que la de la comunicació, especialment l'avaluació, la comunitat científica també ha adaptat les seves estratègies de publicació: els vehicles de comunicació científica s'adeqüen a què es mesura i com es mesura (De Rijcke et al., 2016).

Aquest article tracta de l'ús amb finalitats d'avaluació que es fa dels indicadors en canals de comunicació científica i analitza les limitacions i els reptes que amenacen aquesta pràctica. Els indicadors bibliomètrics són un recurs cada cop més utilitzat en l'avaluació de la recerca i la planificació estratègica. Gläser i Laudel (2007) suggereixen tres raons principals que expliquen aquest ús exponencial a l'hora d'avaluar la producció de coneixement i comparar el rendiment de les unitats de producció de coneixement, siguin persones o estats. En primer lloc, la creixent demanda d'avaluacions no es pot satisfer per mitjà de l'avaluació d'experts i, en aquest sentit, els indicadors bibliomètrics esdevenen un mètode d'avaluació més ràpid i més econòmic. En segon terme, atès que es basen en resultats de recerca validats per diversos experts en la matèria, els indicadors bibliomètrics són considerats un mètode d'avaluació "objectiu", més fiable que no pas l'avaluació d'experts, percebuda com una tècnica "subjectiva". El tercer motiu que donen és que els indicadors bibliomètrics són aparentment accessibles (intel·ligibles) per a polítics i empresaris i sembla que redueixin la necessitat d'haver de recórrer a personal científic especialitzat perquè avalui la recerca; així, en el finançament de la investigació, hi ha qui veu els indicadors com una drecera directa per a evitar el problema del principal i l'agent (van der Meulen, 1998). Mentre que l'avaluació d'experts exigeix un coneixement profund de les pràctiques i les característiques dels camps de recerca en qüestió, hi ha la creença que la interpretació de les quantitats generades en una anàlisi bibliomètrica no requereix aquest coneixement contextual especialitzat. En la presa de decisions, doncs, els indicadors bibliomètrics fan l'efecte de ser un mètode més senzill i directe que no pas, per exemple, l'avaluació qualitativa dels resultats de recerca. De manera semblant, en països que ocupen una posició perifèrica en el sistema científic es pot considerar que els indicadors bibliomètrics convencionals tenen una legitimitat internacional que l'avaluació d'experts, que és local, no té (Ràfols et al., 2016).

En aquest context, veiem que les decisions que es prenen en el disseny i l'aplicació de polítiques científiques es basen, i fins i tot venen determinades, cada cop més per l'ús comparatiu d'indicadors quantitatius: el *benchmarking*, els rànquings i la comparació directa d'indicadors simples (siguin sintètics o reflecteixin una sola dimensió) són exemples d'aquesta tendència. Això no obstant, l'ús poc curós dels indicadors implica un gran nombre d'inconvenients, i aquesta és la raó per la qual en els darrers anys han proliferat les crides a l'ús responsable d'aquestes eines (Hicks et al., 2015, Wilsdon et al., 2015).

En aquest article distingim tres tipus de problemes que presenta l'ús dels indicadors ciènciomètrics. Analitzem els casos en què aquests indicadors no permeten una aproximació adequada a les propietats que haurien de reflectir i examinem els motius pels quals sorgeixen aquests problemes. Així mateix, identifiquem aquelles àrees que els indicadors actuals no tenen en compte o que cobreixen de manera irregular. Atès que la recerca i la comunicació dels resultats científics es veuen afectats directament per l'avaluació que se'n fa per mitjà d'un grup reduït d'indicadors bibliomètrics, en el sentit que l'ús d'aquests indicadors altera els objectius originals de les activitats i les polítiques de recerca (Rijcke et al., 2016), els problemes que estudiem tenen una conseqüència immediata en les activitats de recerca com ara la consolidació de la naturalesa perifèrica d'alguns dels temes objectes d'avaluació (Chavarro et al., 2017, Vessuri et al., 2014). Aquesta qüestió la tractem a través d'una sèrie d'exemples extrets de la literatura actual.

2 Marc analític: propietats, mètriques i indicadors

Les societats i els individus valorem les propietats que no podem observar de manera directa. Aquest és el cas de molts dels objectius últims que persegueixen les polítiques públiques: el benestar, la salut o la seguretat són propietats que no són observables directament però que considerem importants. Així, a pesar que no les podem percebre directament, podem fer altres observacions i fer-les servir com a indicadors que ofereixin una representació útil i vàlida d'aquestes propietats. Considerem, posem per cas, que les citacions que es fan d'un article en reflecteixen la qualitat acadèmica.¹ Conseqüentment, si entenem les citacions com un indicador d'aquesta propietat, l'observació dels cops que aquest article ha estat citat ens donarà una aproximació de la qualitat acadèmica que té aquesta recerca científica concreta.

Per expressar-nos sense ambigüitats, aclarim tot seguit la terminologia que fem servir en aquest article. Quan parlem de **propietats** ens referim als atributs o característiques generals de la recerca que persegüim: la qualitat acadèmica, la intensitat de la col·laboració, la diversitat de coneixements o el paper de l'agent de coneixement, entre altres. Per **indicadors** entenem allò que podem observar empíricament i registrar de manera directa (citacions, nombre d'autors, nombre de disciplines, nombre de posicions d'enllaç, etc.) i que podem relacionar amb una propietat mitjançant un model teòric (vegeu la figura 1). En altres paraules, els indicadors són observacions (qualitatives o quantitatives) que a través d'un model conceptual o teòric representen una propietat d'un sistema. Finalment, els **paràmetres** fan referència als nombres que donen les referències amb les quals es poden traçar comparacions entre les observacions i els models. Com veurem més endavant, els paràmetres poden canviar segons el context.²

Volem recalcar que l'enllaç entre l'indicador i la propietat que aquest reflecteix reposa en un model teòric, sigui de manera explícita o implícita. Per exemple, en el cas de les citacions, Merton les entén com una forma de reconeixement i presenta un model teòric segons el qual aquestes són un indicador de qualitat. De la mateixa manera, la despesa en recerca i desenvolupament és una mostra de l'esforç posat en l'activitat científica i, en menor mesura, en la tecnològica. És a dir, si seguim el model lineal d'innovació, en què la recerca científica genera un coneixement que posteriorment és utilitzat per produir avenços tecnològics, que al seu torn condueixen a innovacions que milloren la salut i el benestar de la població, aleshores podem considerar la despesa en recerca i desenvolupament un indicador associat a (futures) millores en salut i benestar. Si els models són més complexos i tenen en compte, posem per cas, que la innovació es desenvolupa a través de processos de retroalimentació en què participen una sèrie d'actors i activitats, l'anàlisi d'aquesta innovació exigirà un nombre més elevat d'indicadors que reflecteixin un conjunt d'activitats més ampli, com pot ser el que es fa servir en l'*Enquesta sobre innovació comunitària (CIS)* i que es defineix en documents com ara el *Manual d'Oslo* (Smith, 2004).

Figura 1. Relació entre propietat i indicador. Els indicadors són alternatives que serveixen per a mesurar una propietat quan hi ha un model teòric que relaciona la propietat amb les observacions dels indicadors



D'aquesta problemàtica en deriven dues conseqüències bàsiques:

- La primera, que per més que sigui òbvia sovint se sol passar per alt, és que un indicador es relaciona amb una propietat d'interès, però són coses diferents. Això pren una importància rellevant quan els indicadors es fan servir com a base per al disseny de polítiques. Les propietats d'interès acostumen a variar segons el context en el qual s'emmarca cada política, i, per tant, la pertinència d'un indicador

concret pot ser que no sigui la mateixa en contextos diferents.

- La segona és que la relació que s'estableix entre un indicador i una propietat està regida per un model teòric. Un altre cop, la idoneïtat dels supòsits teòrics de la qual depèn la validesa de l'indicador pot canviar segons el context. És a dir, les relacions entre indicadors i propietats són vàlides només en unes condicions particulars i, així, la capacitat que té un indicador per reflectir una propietat d'interès depèn de si aquestes condicions es donen en el context específic en què s'aplica l'indicador.

Així, doncs, la validesa de qualsevol indicador depèn sempre del context: el que funciona en un cas pot no funcionar en un altre. Tanmateix, l'ús més freqüent dels indicadors com a mètode d'avaluació es fa amb finalitats comparatives. Per exemple, com ja hem dit més amunt, la majoria de vegades en què s'utilitzen els indicadors bibliomètrics com a sistema d'avaluació és perquè cal prendre decisions sobre opcions alternatives, sovint en matèria de distribució de recursos. En aquestes circumstàncies és essencial que el model que determina la relació entre indicador i propietat sigui el mateix en tots els casos en què s'apliquen els indicadors. En aquesta situació diem que el model ha de ser estable transversalment; això és, comparable en contextos diferents en un mateix moment. D'altra banda, els indicadors també es poden fer servir per fer estudis longitudinals o comparacions d'un mateix element en moments diferents, com seria el cas d'un estudi en què es volguessin identificar els efectes que ha tingut l'aplicació d'una política pública determinada comparant la situació abans i després de l'aplicació. En aquests casos diem que el model ha de ser estable longitudinalment, al llarg del temps.

La falta d'estabilitat en el model que enllaça indicador i propietat és solament una de les raons que pot dur l'indicador a representar una versió distorsionada de la realitat. En la secció següent analitzem els diversos motius pels quals els indicadors com a mètode d'avaluació poden ser inadequats en circumstàncies determinades.

3 Indicadors cienciomètrics: causes de la distorsió

Podem dividir les causes per les quals els indicadors ofereixen una mala interpretació de les propietats d'interès en un context específic en tres categories diferents:

Causes de la distorsió	Indicadors	Model	Propietats
Paràmetres inestables	INESTABLE	Correcte	Rellevant
Model incorrecte	Estable	INCORRECTE	Rellevant
Propietats irrelevantes	Estable	Correcte	IRRELLEVANT

Taula 1. Problemes que poden fer que els indicadors representin incorrectament les propietats d'interès

La resta de l'article se centra a estudiar la casuística descrita en aquesta taula. A continuació analitzem individualment les tres situacions en què una de les tres condicions deixa de ser vàlida:

- El model que vehicula la relació entre indicadors i propietats d'interès no és vàlid en totes les situacions en què s'ha d'aplicar. Encara que el model sigui correcte, els paràmetres del model canvien d'un cas a l'altre (columna esquerra).
- El model és incorrecte, per més que sigui estable i rellevant. És a dir, els indicadors i les propietats que aquests han de reflectir s'ubiquen en una àrea que és significativa per a l'objectiu que perseguim, i el model que relaciona propietats i indicadors és estable. Això no obstant, el model no és correcte, en el sentit que pren com a vàlids supòsits sobre la relació entre propietats i indicadors que no s'adeqüen a la nostra interpretació actual de la realitat (columna central).
- El model és estable, però les propietats tenen una rellevància desigual per als propòsits de l'estudi. Els indicadors ens fan fixar-nos en propietats que no són les que ens interessen per obtenir els resultats que volem (columna dreta).

4 Paràmetres inestables

Els indicadors s'apliquen a poblacions heterogènies, i consegüentment és possible que el model que enllaça indicador i propietat no sigui vàlid per a tota la població objecte d'estudi. És a dir, és possible que els paràmetres del model no siguin els mateixos per a tots els casos sotmesos a comparació. Aquesta inestabilitat podem trobar-la de manera transversal, en diferents elements analitzats en un mateix moment, o de manera longitudinal, en un mateix element analitzat en moments diferents. En aquestes situacions, cada observació que fem dels elements que mostren uns nivells similars de la propietat d'interès oferirà una lectura diferent dels indicadors.

Un exemple d'aquesta situació és la cobertura desigual que es fa de les diverses disciplines, llengües o classes de recerca científica en les bases de dades de publicacions científiques més emprades en l'anàlisi bibliomètrica: Web of Science i Scopus (Vessuri et al., 2014; Chavarro, Tang i Ràfols, 2017). En aquest cas el model vincularia les citacions rebudes amb la qualitat acadèmica de la investigació duta a terme per una persona o un grup de recerca. El plantejament teòric seria, doncs, que l'objectiu d'un investigador és difondre entre la comunitat científica els resultats de la seva recerca, la qual cosa es fa per mitjà de canals de comunicació que tenen una llarga tradició en aquest àmbit, les publicacions científiques. Una bona recerca ha de tenir les qualitats adients que la facin influent, i aquesta influència es tradueix en citacions de l'article en què s'ha publicat la recerca en qüestió, seguint la teoria del reconeixement de Merton. Així, doncs, les citacions esdevenen un indicador de qualitat acadèmica. A pesar de les diverses crítiques que ha rebut des de rigorosos estudis, com ara el de Martin i Irvine (1983), aquesta és una idea bastant estesa i es fa servir en tots els nivells d'anàlisi —des de l'avaluació del rendiment d'una persona concreta fins als estudis de

benchmarking del rendiment d'un estat—, sigui de manera directa, sigui per mitjà d'altres indicadors complementaris, com és l'índex d'impacte d'una publicació.³

Això no obstant, encara que acceptem que les citacions són realment un indicador de qualitat en diferents àmbits de recerca —com veurem més endavant, això no és sempre així—, els patrons de publicació i de citació difereixen d'un àmbit a l'altre i d'una disciplina a l'altra. Per exemple, els índexs de citació i d'impacte de les publicacions de biologia cel·lular i molecular són molt més elevats que no pas els de les publicacions especialitzades en matemàtiques (Althouse et al., 2009). Aquesta disparitat s'ha associat tradicionalment a una diferència entre disciplines en el percentatge de citacions d'articles d'altres publicacions indexades i en la mitjana de citacions que té un article d'altres publicacions. Així, doncs, "Those fields which cite heavily within the ISI data set, such as Molecular Biology or Medicine, buoy their own scores. Those fields which do not cite heavily within the ISI data set such as Computer Science or Mathematics have correspondingly lower scores" (Althouse et al., 2009).

D'aquí resulta que qualsevol comparació directa entre camps diferents que es faci a partir de citacions de publicacions que pertanyen a àmbits de recerca diferents serà una comparació errònia perquè el model no serà transversalment estable. El patró de generació de citacions varia d'una disciplina a l'altra i, per tant, la relació entre les citacions i la qualitat o l'impacte també és diferent: una citació en matemàtiques no significa el mateix que una citació en biologia.

Per solucionar aquest problema s'ha proposat normalitzar els indicadors per mitjà de l'operació següent: dividir el nombre de vegades que se cita un article per la mitjana de citacions que es fan en un camp concret. Per bé que seguim tenint el mateix indicador, ara el model ens assegura que les citacions reflecteixen la qualitat acadèmica comparada. Aquesta mesura normalitzadora té en compte la tendència a citar que hi ha en el camp de la publicació: sigui fent servir indicadors de citació ponderats o a partir de l'índex d'impacte de la publicació per comparació a altres publicacions de la mateixa disciplina. De fet, ja fa molts anys que hi ha un intens debat i s'han anat fent propostes de fórmules per ponderar citacions d'àmbits de recerca diferents que permetin fer comparacions multidisciplinàries a partir de citacions (Ophthof i Leydesdorff, 2010). Així i tot, continua havent-hi tendències de citació diferents segons el tipus de recerca —per exemple, entre la bàsica i l'aplicada—, per més que aquesta s'emmarqui en la mateixa categoria disciplinària del Web of Science (Van Eck et al. 2013). A més, a pesar de les nombroses opinions contràries (DORA, 2013; Hicks et al., 2015), encara és una pràctica freqüent la d'utilitzar l'índex d'impacte de la publicació en la qual s'ha publicat un article com a indicador de la qualitat de la recerca expressada en aquest article.

Aquesta disparitat també s'esdevé quan en l'estudi hi ha més d'una llengua involucrada. El fet que les bases de dades bibliomètriques més utilitzades (WoS i Scopus) tinguin una millor cobertura de publicacions escrites en anglès que en qualsevol altre idioma se sol considerar una limitació que afecta tots els indicadors que es basin en aquesta informació. Així, mentre que en l'àmbit de les ciències naturals l'ús de l'anglès com a llengua vehicular cada vegada està més establert, en les ciències socials, el dret i les humanitats l'ús de les llengües nacionals és important, amb la qual cosa la visibilitat internacional de les publicacions d'aquestes disciplines es veu minvada (Van Leeuwen, 2013, p. 1). Els idiomes nacionals també tenen força pes en àmbits professionals com és la medicina clínica.⁴ I aquí, un altre cop, de resultes d'aquesta cobertura irregular, el model que relaciona el nombre de citacions (o el nombre de publicacions) amb la qualitat de la recerca no és aplicable igualment a tots els casos en què els resultats de les investigacions es publiquen en idiomes diferents.

Finalment, la relació entre qualitat i citacions (o publicacions amb un alt nombre de citacions rebudes) no és la mateixa en una recerca que se centri en l'anàlisi empírica de temes d'importància local que en una recerca tracti, per exemple, temes teòrics de rellevància general (Chavarro, Tang i Ràfols, 2017). I el problema encara es complica més si ens fixem que també hi ha diferències en les investigacions sobre qüestions de rellevància local en països diferents. La presència, o més aviat l'absència, de determinats temes locals en la literatura acadèmica "de primer ordre" no és casual. En aquest sentit, sovint se sol fer referència a un comentari de l'editor de la cèlebre publicació especialitzada en medicina *The Lancet*, en què adverteix d'un "widespread systematic bias in medical journals against diseases that dominate the least-developed regions of the world" (Horton, 2003, p. 712). En realitat, aquestes paraules beuen d'un altre article, menys citat i actualment difícil de trobar, publicat a l'*European Science Editing*, una revista en si mateixa poc coneguda, en el qual s'analitzaven cinc de les publicacions de més renom en medicina que contenen articles avaluats per experts: *The Lancet*, *The BMJ*, *The New England Journal of Medicine*, *JAMA* i *Annals of Internal Medicine*. En aquest article s'explica que l'espai que aquestes publicacions dedicaven a articles sobre patologies especialment significatives a l'Europa Occidental, l'Amèrica Central i del Sud, el Carib, l'Àfrica del Nord i l'Orient Mitjà era molt reduït, tret dels que versaven sobre certes malalties infeccioses (VIH/sida, malària i tuberculosi), i la majoria eren escrits per autors de països amb ingressos elevats, o per autors de països amb ingressos baixos en col·laboració amb autors de països d'ingressos elevats. L'estudi conclouia que "[the] best chances of getting published in one of these journals is only through collaborating with a developed country scientist" (Obuaya, 2002).

Val a dir que aquesta inestabilitat transversal no és exclusiva del camp de la bibliomètrica, sinó que també és present en alguns dels indicadors d'innovació més utilitzats. Per exemple, el nombre d'empreses derivades del món universitari (empreses creades per acadèmics i estudiants que s'acaben de titular) fa temps que s'utilitza com a indicador de l'activitat emprenedora de la comunitat universitària. Així, doncs, una manera de mesurar l'emprenedoria acadèmica seria simplement comptar quantes d'aquestes empreses genera una universitat, una regió o un estat. Tanmateix, la connexió entre aquestes empreses derivades i les activitats i habilitats d'emprenedoria no és la mateixa en tots els contextos. D'entrada, en casos en què l'índex d'atur és elevat, l'autoocupació en forma de microempresa pot ser una de les poques vies que té el jovent que s'acaba de llicenciar per guanyar-se un sou. Així, en regions i països amb taxes d'atur juvenil disperses, tant la cultura i les habilitats d'emprenedoria com el nombre d'empreses derivades del món universitari seran diferents. De la mateixa manera, aquesta relació també pot canviar amb el pas del temps si en un mateix país l'índex d'atur juvenil canvia considerablement i, per tant, l'indicador pot ser que esdevingui inestable longitudinalment. En aquest sentit, cal afegir que s'ha demostrat que en contextos molt burocratitzats les empreses derivades de la universitat es fan servir per canalitzar la recerca acadèmica (Morales-Gualdrón et al., 2009). El problema, doncs, es repeteix també en aquests casos, ja que la manera en què es relaciona la quantitat d'empreses derivades i l'emprenedoria varia segons la rigidesa de la burocràcia acadèmica: partint d'un mateix nivell d'emprenedoria universitària, el nombre d'empreses derivades serà superior en els contextos molt burocratitzats.

5 Models incorrectes

El segon motiu pel qual els indicadors poden mostrar una visió errònia de la variable d'interès és que el model sigui incorrecte. Això pot passar fins i tot quan la variable és rellevant des d'un punt de vista normatiu i el model és estable. Per exemple, la idea que els resultats científics s'expressen en articles de publicacions pot no ser gens certa en determinats àmbits de recerca. Com hem dit més amunt, els paràmetres deixen de ser estables quan es comparen disciplines amb tendències i patrons de publicació diferents fent servir indicadors de publicació no ponderats. A més, quan una publicació concreta no es considera un mitjà rellevant per difondre els resultats d'una recerca, els indicadors de publicació basats en publicacions no són vàlids per a l'avaluació d'aquests resultats, encara que no hi hagi cap comparació implicada. Un cas semblant és quan les bases de dades emprades recullen només una fracció de les publicacions rellevants. Així, un article publicat el 2006 assegurava que el *Journal Citation Reports (JCR)* només contenia l'11% de les publicacions en matèria d'educació avaluades per experts (Togia i Tsigilis, 2006). En aquestes circumstàncies, hom podria argumentar que el *JCR* no oferia aleshores un reflex vàlid dels resultats científics perquè la idea (el model) segons la qual una mostra significativa dels resultats de la recerca feta en aquest àmbit es publicava al *JCR* no era vàlida.

Un altre cas en què es pot considerar que el model és incorrecte és quan els experts classifiquen les publicacions segons la seva qualitat, i

després aquestes classificacions s'utilitzen per a avaluar la qualitat d'organitzacions o d'investigadors a títol individual. En aquesta situació el model sosté que els experts que avaluen la recerca tenen la capacitat d'identificar la qualitat de les publicacions i classificar-les segons aquesta propietat, que aquesta qualitat manifesta la qualitat dels diversos articles que apareixen en aquestes publicacions, i que quan els resultats d'una recerca són "prou bons" per a publicar-los-hi els autors d'aquests estudis normalment escolliran aquestes publicacions per davant d'altres que es consideren menys prestigioses. Hi ha àmbits en què les classificacions de publicacions fetes per experts han esdevingut el pal de paller a partir del qual s'han construït indicadors de qualitat de recerca. Un bon exemple de classificacions d'aquest tipus és la de l'Association of Business Schools (ABS), del Regne Unit, que ha passat a fer-se servir a tot el món per a avaluar els resultats d'investigadors (Nedeva et al., 2012) i amb altres finalitats, com és ara seleccionar acadèmics "actius en la recerca" per a sotmetre'ls a l'avaluació anual britànica, el Research Excellence Framework. La recerca ha demostrat, però, que les classificacions no són neutrals pel que fa als tipus de recerca; dit d'una altra manera, les classificacions de publicacions tendeixen a afavorir determinats tipus de recerca. En conseqüència, la creença que els experts poden classificar les publicacions segons la seva qualitat d'una manera fiable i, doncs, neutral, no és certa. Les publicacions que ocupen els llocs superiors en les classificacions de l'ABS han demostrat tenir una cobertura de disciplines menys diversa que no pas altres publicacions que estan en posicions inferiors i, per tant, l'ús d'aquestes classificacions com a eina de mesura beneficia els grups de recerca que se centren en l'administració i la direcció d'empreses quan se'ls compara amb unitats d'innovació i recerca interdisciplinàries, tal com va demostrar l'estudi de Rafols et al. (2012, p. 1276). En aquesta anàlisi comparativa, els primers rebien uns resultats superiors si per a avaluar-los es feien servir les classificacions de l'ABS, però quan s'utilitzava una anàlisi a partir de citacions la situació es capgirava i eren les segones les que obtenien uns resultats millors. Els autors d'aquest estudi conclouen que les classificacions de l'ABS funcionen com un "general mechanism [...] [that] can suppress interdisciplinary research" (Rafols et al., 2012, p. 1262). En essència, el problema sorgeix perquè els supòsits teòrics que relacionen les classificacions de l'ABS amb la qualitat de recerca són incorrectes.

6 Propietats irrelevantes

Podem trobar-nos casos en què els indicadors representin les propietats de manera estable i adequada i que, tanmateix, aquestes propietats siguin irrelevantes per a un context normatiu concret o tinguin una rellevància desigual en diversos contextos normatius. En altres paraules, els indicadors no "il·luminen" totes les zones on hi ha problemes socials i econòmics de pes que es poden abordar des de la recerca. D'aquesta manera, podem dir que els indicadors estadístics convencionals tendeixen a il·luminar zones que són rellevants per als centres d'activitat econòmica, tecnològica i científica, que és on s'han desenvolupat.

Quan l'anàlisi i la normativa se centren només en els àmbits coberts pels indicadors més comuns i es perden de vista els problemes que les polítiques de recerca i desenvolupament haurien d'abordar, diem que es produeix un "efecte fanal", un problema força habitual en el camp de les ciències socials. Aquest terme fa referència a una vella història d'un borratxo que anava mirant per terra buscant les claus a la vora d'un fanal. Era conscient que les havia perdudes en algun altre lloc, però les buscava allà perquè és on hi havia llum. Els fanals, com alguns indicadors, no il·luminen totes les àrees d'importància.

Ja l'any 1964 Abraham Kaplan deia que gran part de la recerca conductual estava "vitiated [...] by the principle of the drunkards search" (Kaplan, 2009, p. 11). En el nostre àmbit d'investigació, els indicadors més comuns en l'estudi i l'avaluació de les polítiques en matèria de ciència, tecnologia i innovació (CTI) il·luminen només una fracció, que sovint no és la més important, de la realitat. En el cas del disseny i avaluació de polítiques, aquestes qüestió pot resumir-se en la figura 2, que mostra un ampli ventall de possibles problemes que podrien tractar-se des de la recerca científica, el desenvolupament tecnològic i la innovació. Tanmateix, com que els recursos que es poden destinar a aquestes activitats són finits, només uns quants d'aquests problemes seran atacats en un context temporal i geogràfic concret, per mitjà d'una sèrie de mecanismes de prioritització de polítiques. Els indicadors que hi ha disponibles, doncs, només il·luminen un subconjunt d'aquests problemes.

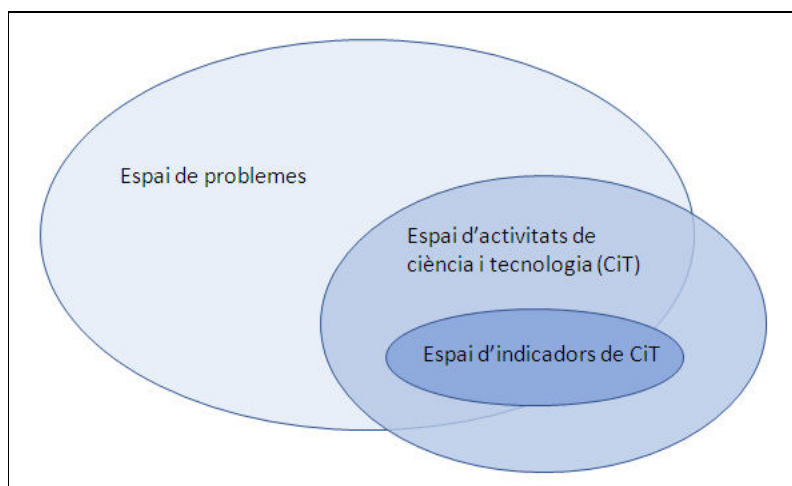


Figura 2. Problemes, activitats de CTI, indicadors i perifèries

La falta de correspondència entre l'espai dels problemes que les polítiques de CTI volen abordar, l'espai de les activitats de CTI que es duen a terme i l'espai de les activitats de CTI que estan cobertes pels indicadors estables dels quals hom disposa genera un efecte fanal: l'analista o el responsable de la presa de decisions es fixa només en l'espai dels indicadors. Per bé que els indicadors poden oferir una representació acurada de la realitat que volen representar, aquesta realitat és rellevant només parcialment o directament irrellevant segons els objectius legislatius que s'hagin marcat.⁵

En la recerca científica podem trobar diversos exemples de situacions en què els indicadors que s'utilitzen no reflecteixen activitats rellevants. Les publicacions científiques poden incórrer en l'error de representar activitats de recerca que tenen per objectiu principal afegir informació a repositoris o bases de dades. Així, una tasca habitual d'un grup de lingüistes que treballen en el desenvolupament i la millora de diccionaris pot ser fer aportacions a un diccionari que ja existeix en comptes d'escriure articles en publicacions, llibres o monografies. Malgrat que tasques com aquesta són importants, no queden reflectides pels indicadors actuals.

Altres exemples d'activitats de CTI que el conjunt d'indicadors més utilitzat deixa a l'ombra són les innovacions dutes a terme per comunitats

locals, les destinades a reduir considerablement els costos d'un servei o un producte o les pensades per satisfer les necessitats dels grups poblacionals més desfavorits. Aquestes innovacions poden contribuir de manera essencial al benestar social, sobretot en els països en vies de desenvolupament, i poden ser peces clau de la transformació tecnològica i social arreu del món. Així i tot, queden fora de l'abast dels indicadors que s'han fet servir tradicionalment per avaluar la innovació en ciència i alta tecnologia, i s'han considerat processos d'innovació que s'esdevenen "below the radar" (Pansera, 2013). De manera semblant, malgrat que l'impacte econòmic de les activitats de recerca i desenvolupament es podria avaluar, en principi, per mitjà de l'anàlisi de variables econòmiques com són l'atur o el PIB, l'avaluació de l'impacte que tenen en la societat i el medi ambient es veu limitada per la manca d'indicadors adequats.

7 Conclusió: conseqüències de l'ús problemàtic dels indicadors

Sempre que vulguem avaluar una activitat fent servir indicadors hem d'acceptar una sèrie de criteris i supòsits. Principalment, hem d'assegurar-nos que els indicadors reflecteixen idòniament les propietats que són rellevants per a la política que volem aplicar. Si prenem els indicadors de manera acrítica, les finalitats últimes de l'activitat sotmesa a avaluació i les propietats associades a aquests objectius queden definides implícitament, sense debat ni consideració. Així, és possible que els criteris que duu implícit l'ús d'indicadors deixin de tenir relació amb els objectius de l'estudi. Aquest problema concret pot generar un problema a escala normativa: pot ser que es formi una bretxa d'avaluació "between evaluation criteria and the social and economic functions of science" (Wouters, 2014).

Aquesta bretxa d'avaluació és crítica perquè les persones i les organitzacions tendeixen a adaptar la seva activitat als criteris amb els quals se'ls avaluarà. Quan aquesta avaluació gira entorn d'un conjunt d'identificadors clarament identificable, és habitual que qui és objecte d'avaluació miri de millorar el seu rendiment ajustant-lo a aquests indicadors. Aquesta conseqüència que té l'ús dels indicadors en el comportament de les persones o organitzacions sotmeses a avaluació ha estat anomenada "constitutive effects" (Dahler-Larsen, 2012).

Si els indicadors estan ben dissenyats i mantenen una relació adequada amb els objectius normatius que es volen assolir, aquests efectes constitutius són, de fet, positius, en el sentit que guien la recerca de qui se sotmet a avaluació en la direcció desitjada pels responsables del disseny de les polítiques. Quan els indicadors són incorrectes, però, aquesta guiatge performatiu passa a ser problemàtic: les activitats de CTI se centren en les àrees il·luminades pels indicadors o en àmbits que no tenen relació amb els objectius de la normativa. Aquesta performativitat pot acabar limitant l'abast de les activitats de recerca, tecnologia, desenvolupament i innovació o fent que vagin en la direcció "equivocada". S'ha dit, per exemple, que l'efecte fanal pot suposar la supressió de la diversitat científica i enfocar la recerca cap a una homogeneïtat més elevada (Rafols et al., 2012), en detriment dels mètodes de recerca crítics o heterodoxos. Aquest problema agafa una rellevància especial si tenim en compte la creixent pressió política per fer servir indicadors com a factors de suport, o fins i tot determinants, en el disseny de polítiques. La majoria d'experts i analistes són conscients d'aquesta problemàtica i fan un ús responsable i curós dels indicadors, però la pressió política ha conduït a un ús imprudent i indiscriminat d'indicadors inadequats amb finalitats avaluadores. L'obstinació actual de la classe política la mesura de resultats, la combinació de termes com ara *evidències*, *dades* i *mesures* en el discurs polític, i la popularitat de les classificacions i el benchmarking com a eines d'avaluació són factors que contribueixen a un ús indiscriminat dels indicadors, que es fan servir sense tenir en compte la rellevància ni la validesa que tenen en el context específic en el qual s'apliquen.

Aquesta situació exigeix repensar l'ús dels indicadors en l'avaluació. Els tres tipus de problemes que hem exposat en aquest article poden solucionar-se per mitjà de la contextualització dels indicadors; això és, per mitjà del disseny i l'adaptació d'indicadors a unes condicions d'avaluació concretes, específiques per tractar determinats casos i problemes. Cal adoptar un punt de vista més divers en el disseny i l'ús dels indicadors que pugui contribuir al disseny de polítiques de manera plural i adaptable a cada situació (Rafols, Ciarli et al., 2012). Als nostres ulls, aquests enfocaments, per bé que encara els queda molt camí, haurien de ser l'objecte central de la recerca en indicadors científics i tecnològics.

Bibliografia

- Althouse, B. M., West, J. D., Bergstrom, T. i Bergstrom, C. T. (2009). 'Differences in Impact Factors Across Fields and Over Time', *Journal of the American Society for Information Science*, 60(1), p. 27-34.
- Chavarro, D., Tang, P. i Rafols, I. (2017). 'Why researchers publish in non-mainstream journals: Training, knowledge-bridging and gap filling', *Research Policy*, 46, p. 1666-1680.
- Dahler-Larsen, P. (2012). *The Evaluation Society*. Traducció de Sampson, S. Stanford: Stanford University Press.
- DORA (2013). San Francisco Declaration on Research Assessment. Disponible a: <http://www.ascb.org/dora/> (<http://www.ascb.org/dora/>)
- Gläser, J. i Laudel, G. (2007). 'The Social Construction Of Bibliometric Evaluations', a: Whitley, R. & Gläser, J. (eds.). *The Changing Governance of the Sciences Sociology of the Sciences Yearbook*: Springer Netherlands, p. 101-123.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S. i Rafols, I. (2015). 'The Leiden Manifesto for research metrics', *Nature*, 520, p. 429-431.
- Opthof, Tobias; Leydesdorff, Loet (2010). Caveats for the journal and field normalizations in the CWTS ('Leiden') evaluations of research performance. *Journal of Informetrics*, 4(3), p. 423-430.
- Horton, R. (2003). 'Medical journals: evidence of bias against the diseases of poverty', *The Lancet*, 361(9359), p. 712-713.
- Kaplan, A. (2009). *The conduct of inquiry: methodology for behavioral science*. New Brunswick, New Jersey: Transaction Publishers. 4a reimpressió.
- Martin, B.R. i Irvine (1983). 'Assessing basic research. Some partial indicators of scientific progress in radio astronomy'. *Research Policy* 12, p. 61-90.
- Morales-Gualdrón, S. T., Gutiérrez-Gracia, A. i Roig Dobón, S. (2009). 'The entrepreneurial motivation in academia: a multidimensional construct', *International Entrepreneurship and Management Journal*, 5(3), p. 301-317.
- Nedeva, M., Boden, R. i Nugroho, Y. (2012). 'Rank and File: Managing Individual Performance in University Research', *Higher Education Policy*, 25, p. 335-360.
- Obuaya, C. (2002). 'Reporting of research and health issues relevant to resource-poor countries in high-impact medical journals', *European Science Editing*, 28(3), p. 72-77.
- Pansera, M. (2013). 'Frugality, Grassroots and Inclusiveness: New Challenges for Mainstream Innovation Theories', *African Journal of Science, Technology, Innovation and Development*, 5(6), p. 469-478.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P. i Stirling, A. (2012). 'How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business Management', *Research Policy*, 41(7), p. 1262-1282.

Rafols, I., Ciarli, T., van Zwanenberg, P. i Stirling, A. (2012). 'Towards indicators for 'opening up' science and technology policy'. *Proceedings of 17th International Conference on Science and Technology Indicators*, Vol. 2, p. 663-74. Disponible a: 2012.sticonference.org/Proceedings/vol2/Rafols_Towards_675.pdf

Rafols, I., Molas-Gallart, J., Chavarró, D. A., & Robinson-García, N. (2016). On the dominance of quantitative evaluation in 'peripheral' countries: Auditing research with technologies of distance. Disponible a: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2818335

Rijcke, S. d., Wouters, P. F., Rushforth, A. D., Franssen, T. P. i Hammarfelt, B. (2016). 'Evaluation practices and effects of indicator use: a literature review', *Research Evaluation*, 25(2), p. 161-169.

Smith, K. (2004). 'Measuring Innovation', a: Fagerberg, J., Mowery, D.C. & Nelson, R.R. (eds.). *The Oxford Handbook of Innovation*. Oxford: Oxford University Press, p. 148-177.

Stirling, A. (2012). *Sharing the Lathes of Knowledge: broadening out and opening up policy metrics, models and indices*. Obra no publicada.

Togia, A. i Tsigilis, N. (2006). 'Impact factor and education journals: a critical examination and analysis', *International Journal of Educational Research*, 45(6), p. 362-379.

Van der Meulen, B. (1998). 'Science policies as principal-agent games: Institutionalization and path dependency in the relation between government and science', *Research Policy*, 27(4), p. 397-414.

Van Leeuwen, T. (2013). 'Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship?', *Bibliometrie - Praxis und Forschungen*, (Band 2).

Vessuri, H., Guédon, J. C. i Cetto, A. M. (2014). 'Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development', *Current Sociology*, 62(5), p. 647-665.

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. i Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, London: HEFCE.

Wouters, P. (28 August 2014). 'A key challenge: the evaluation gap', *The Citation Culture*. Disponible a: <https://citationculture.wordpress.com/2014/08/28/a-key-challenge-the-evaluation-gap/> (Consulta: 18.08.2015).

Notes

¹ Pensem que *qualitat acadèmica* fa referència a un concepte massa vague i ambigu i, consegüentment, l'ús que se'n fa en l'avaluació científica pot resultar problemàtic. Suggerim, en canvi, fer servir *impacte científic*. En aquest article utilitzem la primera expressió, tanmateix, perquè és una propietat que se sol associar a un dels indicadors més freqüents: les citacions.

² Aquest és un marc volgutament senzill. En altres casos pot ser útil fer servir marcs més complexos que tinguin en compte els significats de mètrica o mesura, observació i variable (Stirling, 2012). En aquest article, però, pensem que un marc que inclogui propietats i indicadors té la solidesa suficient per explicar els reptes principals als quals s'enfronta l'avaluació de la recerca.

³ L'índex d'impacte d'una publicació és el nombre de vegades que els articles que s'hi han publicat s'han citat en altres publicacions durant un període de temps determinat després d'haver-se publicat, generalment dos anys. Malgrat que fa temps que es qüestiona l'ús d'aquesta mesura com a indicador de qualitat, sovint encara se sol creure que un article aparegut en una publicació que té un índex d'impacte elevat és un article d'alta qualitat.

⁴ For a interessant esbrinar com es tradueix aquesta cobertura idiomàtica desigual en una cobertura també irregular de les disciplines i els temes de recerca (Archambault et al., 2006; Piñeiro i Hicks, 2015).

⁵ El problema s'agreuja quan aquesta atenció selectiva no és intencional, sinó que de manera implícita i incorrecta s'assumeix que tot l'espai de les activitats de CTI queda cobert pels indicadors que es fan servir per avaluar-les. Utilitzar indicadors que il·luminen només zones determinades d'un conjunt d'activitats per mesurar-ne el rendiment global és una pràctica habitual.

