

Bibliometria: Leidenský manifest štatistického hodnotenia vedeckého výskumu

Diana Hicks, Paul Wouters, Ludo Waltman, Sarah de Rijcke a Ismael Rafols

Používajte týchto desať princípov pri hodnotení vedeckého výskumu, vyzývajú Diana Hicks, Paul Wouters a ich spoluautori.

Posudzovanie vedeckého výskumu sa v čoraz väčšej miere opiera o štatistické dáta. Kedysi platilo, že hodnotenie výskumu bolo neformálne, občasná a vykonávali ho kolegovia. Dnes závisí na štatistických kritériách a vykonáva sa pravidelne.¹ Tento spôsob hodnotenia však nestačí. Dnešným problémom je, že hodnotenie sa zakladá na číslach namiesto úsudku. Štatistické postupy sa rozmohli a väčšinou sa používajú s dobrými úmyslami, no nie vždy aj kompetentne a často nesprávne. Hrozí nám tak, že práve tie nástroje, ktoré boli navrhnuté, aby zlepšili akademické prostredie, ho poškodia. Hodnotenie vedy je čoraz častejšie v rukách inštitúcií, ktorým chýbajú znalosti, ako štatistické údaje správne používať a interpretovať, alebo nevedia, na koho sa v tomto ohľade obrátiť.

Do roku 2000 používali odborníci na bibliometriu Science Citation Index vydávaný na CD Inštitútom pre vedecké informácie (Institute for Scientific Information, ISI). V roku 2002 spustilo vydavateľstvo Thomson Reuters integrovanú online platformu, čím bola ich databáza [Web of Science](#) sprístupnená celej vedeckej obci. Odozvou bol vznik ďalších konkurenčných citačných indexov: [Scopus](#) vydavateľstva Elsevier (spustený v roku 2004) a [Google Scholar](#) (beta verzia spustená v roku 2004). Objavili sa prvé online nástroje, ktoré umožňovali porovnávať produktivitu vedeckých inštitúcií a ich impakt, napríklad [InCites](#) (používajúci Web of Science) a [SciVal](#) (používajúci Scopus). Nasledoval vznik programov umožňujúcich analýzu jednotlivých citačných profilov, na základe databázy Google Scholar ([Publish or Perish](#) spustený v roku 2007).

V roku 2005 navrhol Jorge Hirsch, fyzik z Kalifornskej univerzity v San Diegu, takzvaný [h-index](#), vďaka ktorému sa rozšírila prax počítania citovaní jednotlivých bádateľov. Záujem o [impakt faktor](#) časopisov vzrastal kontinuálne od roku 1995 (pozri obrázok Posadnutosť impakt faktorom).

Najnovším trendom tohto vývoja sú štatistické nástroje, týkajúce sa spoločenského dosahu a online komentárov – [F1000Prime](#) (spustený v roku 2002), [Mendeley](#) (od roku 2008) a [Altmetric.com](#) vytvorený v roku 2011 a podporovaný vydavateľstvom Macmillan Science and Education, vlastníka Nature Publishing Group.

My, bibliometrici, odborníci na spoločenské vedy a administrátori vedeckého výskumu, pozorujeme so stále väčším znepokojením, ako sa štatistické kritéria určené na hodnotenie vedeckého výstupu bežne používajú nevhodným spôsobom. Uvedme len niekoľko príkladov. Univerzity na celom svete sa stali posadnuté tým, na akom mieste sa nachádzajú v rebríčkoch hodnotenia (napríklad v [Šanghajskom rebríčku](#) a v zozname vypracúvanom [Times Higher](#)

[Education](#)), hoci tieto zoznamy sú založené, podľa nás, na nepresných dátach a subjektívnych indikátoroch.

Niektorí zamestnávateľia vyžadujú od potenciálnych zamestnancov konkrétne hodnoty h -indexu. Na viacerých univerzitách závisí váš kariérny postup na tom, či ste dosiahli určitú hodnotu h -indexu a či máte dostatočný počet článkov v časopisoch s vysokým impakt faktorom. Životopisy výskumníkov, predovšetkým v oblasti biomedicíny, sa premenili na priestor, kde je potrebné sa chváliť svojimi dosiahnutými hodnotami. Vo všetkých disciplínach dnes vyžadujú vedúci dizertačných prác od svojich doktorandov, aby publikovali v časopisoch s vysokým impakt faktorom a zabezpečili externé finančné prostriedky ešte predtým, než dokončia doktorát.

Niektoré univerzity v Škandinávii a v Číne pridelujú financie alebo bonusy na základe jediného čísla, napríklad tým, že počítajú impakt jednotlivých výskumníkov a na jeho základe udeľujú niekomu "príplatky za produktivitu", alebo tým, že výskumníci, ktorí publikujú v časopise s impakt faktorom vyšším ako 15, majú nárok na bonus.²

V mnohých prípadoch stále platí, že výskumníci aj ich hodnotitelia sú schopní dospieť k vyváženému hodnoteniu. Zneužívanie metrických kritérií sa však stalo natoľko rozšíreným, že ho nemôžeme ignorovať.

Preto predkladáme Leidenský manifest, nazvaný podľa konferencie, kde bol zostavený (<http://sti2014.cwts.nl>). Desať pravidiel, ktoré tu uvádzame, nie je žiadnou novinkou pre bibliometrikov, hoci žiaden z nás by ich nedokázal odcitovať všetky, keďže až doteraz neboli presne sformulované. Priekopníkom v našom odbore, ako napr. Eugenovi Garfieldovi, zakladateľovi ISI, treba pripísať zásluhu za publikovanie niektorých týchto pravidiel.^{3,4} Žiaľ, vedúci predstavitelia univerzít, zvyčajní adresáti správ hodnotiacich komisií, takéto odborné články nečítajú a nie sú odborníkmi na bibliometriu. Podobne výskumníci, ktorí chcú vzniesť námietky voči výsledkom svojho hodnotenia, musia hľadať štúdie, na ktoré by sa mohli odvolať, roztrúsené v odborných časopisoch, ktoré sú pre nich neprehľadné a ku ktorým nemajú prístup.

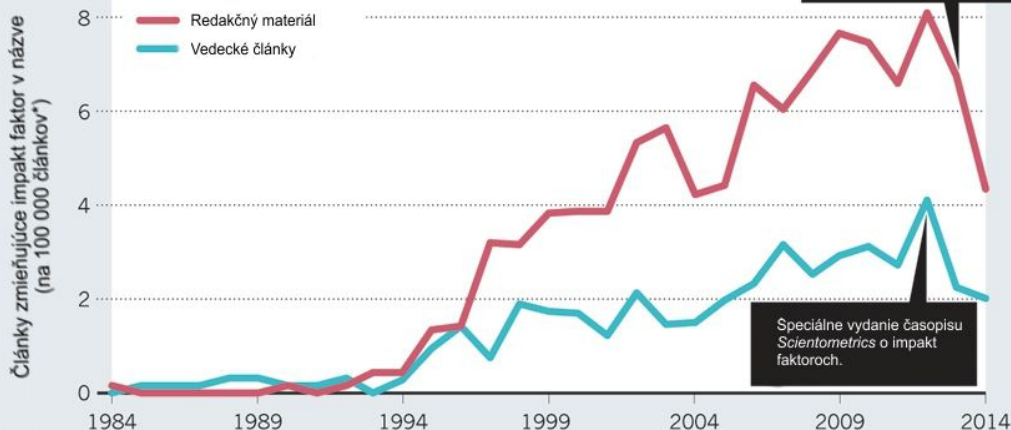
Práve preto v tomto manifeste poskytujeme zhrnutie osvedčených postupov v štatistickom vyhodnocovaní výskumu. Výskumníci na ich základe budú môcť posudzovať prácu hodnotiteľov a hodnotitelia kriticky zvážiť význam rôznych indikátorov.

POSADNUTOSŤ IMPAKT FAKTOROM

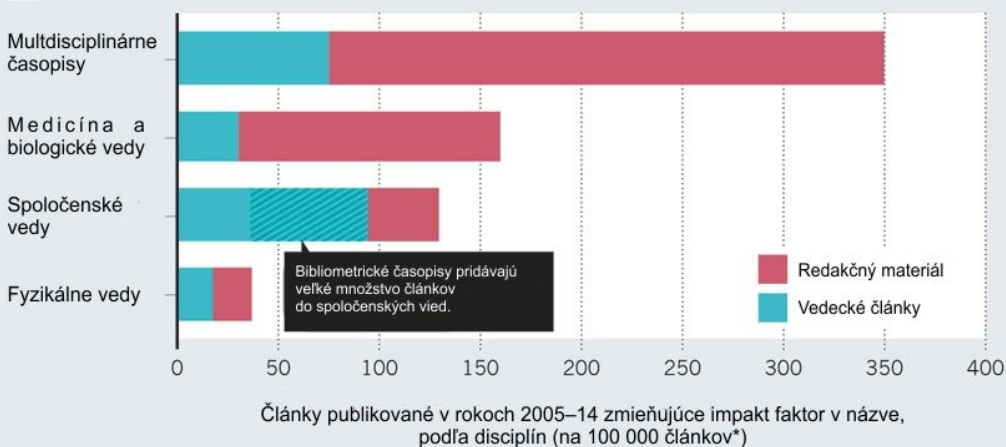
Obrovský nárast záujmu o jediný hrubý ukazovateľ – priemerný počet citácií publikačných jednotiek vydaných v časopise za posledné dva roky – ilustruje krízu v hodnotení vedeckého výskumu.

DORA* deklarácia upozornila, že kvalita výskumu sa nerovná impakt faktoru časopisu.

1 ČLÁNKY UVÁDZAJÚCE IMPAKT FAKTOR V NÁZVE



2 KTO JE NAJVIAC POSADNUTÝ?



* Články indexované vo Web of Science. † DORA, Sanfranciská deklarácia hodnotenia výskumu

Desať pravidiel

1) **Kvantitatívne hodnotenie by malo dopĺňať kvalitatívne hodnotenie odborníkov.** Kvantitatívne, štatistické ukazovatele dokážu zabrániť zaujatosti v recenznom procese a pomôcť pri rozhodovaní. Ich použitie by ideálne malo zlepšiť recenzný proces, keďže posudzovanie druhých výskumníkov je zložité bez širokej škály rôznych informácií. Hodnotitelia však nesmú podľahnúť pokušeniu zameniť vlastný úsudok za čísla. Indikátory nesmú nahrádzať informovaný úsudok. Hodnotiteľ má taktiež vždy zodpovednosť za udelené hodnotenie.

2) **Posudzujte vedecký výkon na základe stanovených cieľov inštitúcie, skupiny alebo vedca.** Ciele konkrétneho výskumného programu by mali byť stanovené na začiatku výskumu

a indikátory použité pri hodnotení výkonu by mali byť jasne nastavené vzhľadom na tieto ciele. Výber indikátorov a spôsob, akým sú použité, by mali zahŕňať širšie spoločensko-ekonomické a kultúrne kontexty výskumu. Ciele výskumníkov sa líšia. Výskum, ktorého cieľom je posunúť hranice vedeckého poznania, sa odlišuje od výskumu usilujúceho sa vyriešiť spoločenské problémy. Hodnotenie sa môže zakladať na prospešnosti daného výskumu pre štátnu sféru, privátny sektor alebo verejnosť, nie nevyhnutne na akademickú excelenciu. Neexistuje žiaden hodnotiaci model, ktorý by bolo možné aplikovať na všetky vedecké projekty a disciplíny.

3) Ochráňte vynikajúci výskum s regionálnym významom. V mnohých častiach sveta je akademická excelencia spájaná s publikovaním v anglickom jazyku. Napr. podľa španielskeho zákona by mali španielski výskumníci publikovať články v časopisoch s vysokým impakt faktorom. Impakt faktor sa pritom počíta pre časopisy vedené v databáze Web of Science sídliacej v USA, ktorá obsahuje prevažne anglické odborné periodiká. Takáto predpojatosť je obzvlášť problematická v prípade spoločenských a humanitných vied, pre ktoré platí, že je výskum vo väčšej miere zameraný regionálne alebo má význam pre konkrétne štáty. Mnohé predmety výskumu majú štátny alebo regionálny rozmer - napríklad epidemiológia HIV v subsaharskej Afrike.

Výskumníci sú motivovaní písať články zaujímavé pre časopisy s vysokým impakt faktorom, teda časopisy v anglickom jazyku, na úkor diverzity výskumu a jeho spoločenského významu. Španielski sociológovia, ktorí vo Web of Science dosahujú vysoké citačné skóre, preto pracujú buď na abstraktných modeloch alebo na dátach týkajúcich sa USA. Stráca sa tak výskum sociológov špecifický pre Španielsko a publikovaný v kvalitných časopisoch v španielskom jazyku, napríklad na témy ako miestne zákonníky práce, zdravotná starostlivosť o seniorov v rodinách alebo zamestnanosť prísťahovalcov.⁵ Bibliometrie založené na kvalitných neanglických publikáciách by mali slúžiť na identifikáciu a oceňovanie excelencie v regionálne významných výskumoch.

4) Zber a analýza hodnotiacich dát musia byť otvorené, transparentné a jednoduché. Tvorba databáz používaných na hodnotenie by mala podliehať jasným pravidlám, ktoré boli stanovené pred dokončením výskumu. Tento postup sa udomácnil u akademických i súkromných skupín podieľajúcich sa na vytvorení bibliometrickej hodnotiacej metodiky počas posledných desaťročí. Tieto skupiny vychádzali z protokolov publikovaných v recenzovanej literatúre. Transparentnosť týchto protokolov umožňovala ich kontrolu. V roku 2010 napríklad verejná diskusia o technických vlastnostiach istého indikátora, používaného jednou z našich skupín (Strediskom pre štúdium prírodných vied a techniky na Leidenskej univerzite v Holandsku) viedla k prehodnoteniu spôsobu, akým sa počítal.⁶ Noví komerční hodnotitelia, ktorí sa chcú zapojiť do hodnotiaceho procesu, by mali dodržiavať tie isté kritéria, nikto by nemal akceptovať netransparentné hodnotenie.

Pri ukazovateľoch je ich jednoduchosť prednosťou, keďže tá dovoľuje väčšiu transparentnosť. Naopak príliš zjednodušujúca štatistika môže skresliť dáta (pozri bod 7). Hodnotitelia sa musia

usilovať o strednú cestu - jednoduché ukazovatele, dostatočne vystihujúce zložitosť vedeckého procesu.

5) Umožnite hodnoteným overiť dáta aj analýzu. V záujme dodržania kvality je potrebné, aby výskumníci, ktorí sú predmetom bibliometrických výskumov, mali možnosť potvrdiť, že ich vedecké výstupy boli spracované korektne. Vykonávatelia hodnotiaceho procesu, by mali garantovať, že dáta, s ktorými pracujú, sú presné, buď overovaním zo strany hodnotených výskumníkov alebo auditom prostredníctvom nezávislej, tretej strany. Univerzity by mohli zaviesť takýto postup overovania v svojich interných informačných systémoch. Prítomnosť nástroja umožňujúceho overovanie by dokonca mala zaväzovať pri objednávkach univerzitných informačných systémov. Zber a analýza presných a kvalitných dát vyžadujú čas a peniaze. Počítajte s nimi v rozpočtoch.

6) Berte ohľad na rozdiely v publikačnej a citačnej praxi v rôznych vedných odboroch. Najlepšou metódou je ponúknuť škálu ukazovateľov, ktoré môžu byť aplikované pri hodnotení, a umožniť jednotlivým odborom, nech si z nich vyberú. Pred pár rokmi sa napríklad stalo, že skupina európskych historikov získala relatívne nízke skóre v hodnotení výsledkov na národnej úrovni, pretože publikovali monografie namiesto článkov v časopisoch zaradených do Web of Science. Títo historici mali jednoducho smolu, že boli súčasťou katedry psychológie. V prípade historikov a spoločenskovedných výskumníkov je potrebné zahrnúť medzi ich výstupy monografie a publikácie v národných jazykoch. V prípade informatikov je zasa potrebné započítavať konferenčné príspevky.

Miera citácií sa líši podľa vedného odboru. V prípade najkvalitnejších časopisov v oblasti matematiky sa ich impakt faktor pohybuje okolo 3, v cytológii je to zhruba 30. Ukazovatele hodnotenia je potrebné štandardizovať, najspoľahlivejšou metódou je v tomto prípade štandardizácia na základe percentilov: každému článku je potrebné prisúdiť hodnotu na základe percentilu, ku ktorému sa radí v rámci citácií v danom odbore (napr. najlepšie 1%, 10% alebo 20%). Jediná vplyvná a často citovaná štúdia len nepatrne ovplyvní rebríčkové postavenie univerzity v hodnotení, založenom na percentiloch, ale môže vystreliť univerzitu zo stredných priečok na vrchol, pokiaľ sa počítajú citačné priemery.⁷

7) Hodnotenie jednotlivých vedcov by malo byť založené na kvalitatívnom posúdení ich vedeckého výkonu. Čím ste starší, tým vyšší je váš *h*-index, hoci nové články nepublikujete. Tento ukazovateľ sa navyše líši podľa odboru: v medicíne a biologických vedách je vrcholovým skóre 200, vo fyzike 100 a v spoločenských vedách 20–30.⁸ Jeho hodnota napokon závisí na databáze použitej pri jeho generovaní: v informatike napríklad môžete mať podľa Web of Science *h*-index okolo 10, zatiaľ čo podľa Google Scholar to je 20-30.⁹ Preto je oveľa vhodnejšie prečítať a posúdiť prácu výskumníka ako zamerať sa na jedno číslo. Aj v prípadoch, kedy je potrebné vyhodnotiť prácu veľkého počtu výskumníkov, najlepší prístup zohľadňuje informácie o odborných znalostiach, skúsenostiach, aktivitách a vplyve vedcov z viacerých zdrojov.

8) **Vyhýbajte sa nevhodnej korektnosti a falošnej presnosti.** Ukazovatele používané vo vede sú z definície náchylné k neurčitosti a nepresnosti a ich výklad vyžaduje prijímať určité zásadné predpoklady, ktoré nie sú všeobecne akceptované. Už dlho sa napríklad vedie debata o výpovednej hodnote počtu citácii. Práve preto je najosvedčenejšou metódou používať viacero ukazovateľov, pretože poskytujú spoľahlivejší a dôkladnejší prehľad. Pokiaľ je možné vyjadriť nepresnosť kvantitatívne, napríklad pomocou intervalu spoľahlivosti, takáto informácia by mala byť uvedená spolu s hodnotami použitých ukazovateľov. Pokiaľ kvantifikácia nie je možná, tí, ktorí stanovujú hodnoty ukazovateľov, by sa mali vyhnúť falošnej presnosti. Hodnoty impakt faktorov odborných časopisov sa napríklad uvádzajú s presnosťou na tri desatinné miesta, aby sa predišlo tomu, že viaceré časopisy majú rovnaké skóre. Kvôli prirodzenej miere neurčitosti a úlohe náhodných premenných pri počítaní počtu citácii však nedáva zmysel rozlišovať medzi časopismi na základe podobne malých rozdielov v impakt faktore. Vyhnite sa falošnej presnosti: len jedno desatinné miesto možno považovať za presné.

9) **Uvedomte si vplyvy hodnotenia a indikátorov na systém vedeckého výskumu.** Ukazovatele menia systém tým, že vytvárajú incentívy. Tieto dôsledky treba predpokladať. Znamená to, že súbor ukazovateľov je vždy lepší ako ukazovateľ jediný takého nastavenie totiž priam vyzýva hodnotených, aby svoju prácu vnímali ako hru, ktorej cieľom je získať čo najvyššie skóre (pričom skutočné ciele výskumu zostávajú bokom). V deväťdesiatych rokoch napríklad financovala Austrália univerzitný výskum na základe vzorca založeného na počte článkov publikovaných jednotlivými ústavmi. Univerzity tak mohli vyčíslit' "cenu" jedného článku v recenzovanom časopise, v roku 2000 to bolo 800 austrálskych dolárov (suma rovná zhruba 470 eur) v prepočte na financie pridelované na výskum. Ako sa dalo očakávať, počet článkov publikovaných austrálskymi výskumníkmi sa zvýšil, ale zároveň vychádzali v časopisoch, ktoré mali nižšiu mieru citovanosti, čím zrejme poklesla kvalita článkov.¹⁰

10) **Pravidelne prehodnocujte indikátory.** Ciele výskumu aj úloha hodnotiacich procesov sa menia. Akademické prostredie sa neustále vyvíja. Štatistické ukazovatele, ktoré boli kedysi užitočné, zastarávajú a popritom vznikajú nové indikátory. Preto musia byť kritéria hodnotenia preverované a prípadne pozmenené. V Austrálii si uvedomili dôsledky aplikácie príliš jednoduchého vzorca a v roku 2010 zaviedli iniciatívu nazvanú Excelencia vedeckého výskumu pre Austráliu.

Ako ďalej

Pokiaľ sa bude riadiť týmito desiatimi pravidlami, hodnotenie vedy môže zohrávať významnú úlohu pri vývoji vedy a jej interakciách so spoločnosťou. Bibliometria vedeckého výskumu dokáže poskytnúť zásadné informácie, ktoré je ťažké získať a interpretovať len na základe individuálnych odborných znalostí. Nesmieme však dovoliť, aby sa kvantifikovateľné informácie zmenili z pomocného nástroja na cieľ vedeckého výskumu.

Najlepšie rozhodnutia o tom, ako smerovať vedu a výskum, sa dosahujú spojením spoľahlivej štatistiky s ohľadom na ciele a povahu hodnoteného vedeckého výskumu. Na to potrebujeme kvantitatívne aj kvalitatívne dáta; pričom obe skupiny sú objektívne odlišným spôsobom. Rozhodovanie o podobách vedy musí spočívať na hodnotiacich procesoch najvyššej kvality založených na kvalitných dátach.

*Preklad z anglického originálu pripravili [Martin Odler](#) a [Evina Steinová](#). Pôvodný anglický text bol publikovaný v Nature **520**, 429–431 (23. apríl 2015), doi:10.1038/520429a*

¹ Wouters, P. in *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact* (eds Cronin, B. & Sugimoto, C.) 47–66 (MIT Press, 2014).

² Shao, J. & Shen, H. *Learned Publ.* 24, 95–97 (2011).

³ Seglen, P. O. *Br. Med. J.* 314, 498–502 (1997).

⁴ Garfield, E. J. *Am. Med. Assoc.* 295, 90–93 (2006).

⁵ López Piñero, C. & Hicks, D. *Res. Eval.* 24, 78–89 (2015).

⁶ van Raan, A. F. J., van Leeuwen, T. N., Visser, M. S., van Eck, N. J. & Waltman, L. J. *Informetrics* 4, 431–435 (2010).

⁷ Waltman, L. *et al.* *J. Am. Soc. Inf. Sci. Technol.* 63, 2419–2432 (2012).

⁸ Hirsch, J. E. *Proc. Natl Acad. Sci. USA* 102, 16569–16572 (2005).

⁹ Bar-Ilan, J. *Scientometrics* 74, 257–271 (2008).

¹⁰ Butler, L. *Res. Policy* 32, 143–155 (2003).