

## Voiko tieteellistä kekseliäisyyttä mitata?

**JORMA TOPPARI** esitteli kolumnissaan Duodecimin numerossa 1/2007 h-indeksin, uuden menetelmän tutkijoiden paremmuuden selvittämiseksi (Hirsch 2005). Kun tutkijan julkaisu laitetaan järjestykseen siteerausten perusteella lähtien eniten siteeratusta, h-indeksi on sen julkaisun järjestysnumero, jonka kohdalla siteerausmäärä on mahdollisimman lähellä järjestysnumeroa. Topparin mielestä h-indeksi on hyvä keino tutkijoiden tuotannon vertailemiseen, ja hän laski useille suomalaisille tutkijoille indeksin arvon. Suurimman arvon 130 sai Erkki Ruoslahti. Indeksin laskemisessa on kuitenkin samat keskeiset puutteet kuin julkaisujen ja siteerausten kokonaislukumäärien laskemisessa.

Tieteessä on erilaisia tavoitteita, kuten 1) kerätä tietoa ihan vain tietomäärän lisäämiseksi, 2) käytännön hyöty tai 3) uusien käsitteiden ja uusilla periaatteilla toimivien menetelmien kehittäminen. Tyypin 1 tutkimuksesta on esimerkkinä epidemiologisten korrelaatioiden laskeminen ja ihmisen genomien sekvenssin määrittäminen. Tyypin 2 tutkimuksesta on esimerkkinä kontrolloidut lääketutkimukset ja tyypin 3 tutkimuksesta helikobakteerin keksiminen (Rautelin ja Kosunen 2005).

Lääketieteessä tarvitaan kaikkia näitä tutkimustavoitteita, mutta tyypin 3 tutkimusta pidetään tieteen ylimpänä tavoitteena, ja juuri tämän ryhmän tutkimuksista jaetaan Nobelin palkintoja.

Jos h-indeksi olisi luotettava mittari parhaan tutkimuksen tunnistamiseksi, niin sen pitäisi olla suuri Nobelin palkinnon saaneilla – jos hyväksymme ajatuksen, että nobelistit edustavat tieteen korkeinta eliittiä.

Taulukkoon 1 laskin viimeisten neljän vuoden lääketieteen nobelistien h-indeksit. Lähes kaikilla h-indeksi sijoittuu Topparin laskemalle Turun lääketieteen professorien vaihteluvälille (4–44), ja lähes kaikki indeksit ovat pienempiä kuin Topparin listaamat suomalaisten huippututkijoiden h-indeksit. Topparin väite, että h-indeksi 98 olisi Nobel-luokkaa, on perusteeton, sillä tuoreiden nobelistien arvot ovat kaikki sen alapuolella (taulukko 1). Nobelin palkintoja ei jaeta sen perusteella, että tutkija on tehnyt suuren määrän hyvää työtä, vaan sen

perusteella, että tutkija on tehnyt mullistavan tieteellisen keksinnön.

Toinen h-indeksin keskeinen ongelma on se, ettei kirjoittajien lukumäärää huomioida. Alkuperäisessä kirjoituksessaan Hirsch (2005) mainitsi tämän ongelman pohdiskelematta sitä syvällisemmin, mutta Toppari jätti asian mainitsematta.

Kun 20 omenaa jaetaan tasan kymmenelle lapselle, jokainen saa kaksi omenaa. Tämän opimme koulun alaluokilla. Mutta kun 20 julkaisua jaetaan kymmenelle tutkijalle, jokainen saa 20 julkaisua, jos he ovat julkaisujen yhteisiä kirjoittajia. Kun kymmenen hengen tutkimusryhmä tekee 20 julkaisua, tehokkuus on kuitenkin aivan sama kuin yhden tutkijan laatiessa kaksi julkaisua. Julkaisuluetteloa vertailtaessa jokainen

**TAULUKKO 1.** Lääketieteen Nobelin palkinnon vuosina 2002–2006 saaneiden tutkijoiden h-indeksit. Nobelistit on poimittu Nobel-säätiön sivustolta (<http://nobelprize.org/>) ja julkaisu sekä niiden siteerausmäärät Science Citation Index -tietokannasta (Terveystieteiden keskus kirjaston Web of Science -liittymä). Jotkin tutkijoista ovat käyttäneet julkaisuissaan sekä yksi- että kaksikirjaimisia etunimien lyhenteitä, ja taulukkoa varten kävin läpi molemmat vaihtoehdot. Kirjoittajien lukumäärä koskee h-indeksin julkaisuja. Kaikilla nobelisteilla kirjoittajamäärän mediaani on välillä 2–5.

Tutkija	Nobelin palkinto vuonna	h-indeksi	Kirjoittajia keskimäärin
Fire AZ	2006	45	4,4
Mello CC	2006	28	5,4
Marshall BJ	2005	39	4,3
Warren JR	2005	24	4,3
Axel R	2004	81	4,8
Buck LB	2004	28	3,5
Lautenbur PC	2003	42	2,6
Mansfield P	2003	43	5,4

ison ryhmän jäsen on kymmenen kertaa niin tehokas kuin yksinäinen tutkija, mutta tämä on virheellisestä laskutavasta johtuva harhakuva. Jos kymmenen hengen ryhmä tuottaa 13 julkaisua samassa ajassa kuin yksittäinen tutkija tuottaa, niin suuri ryhmä on tehokkaampi – jos tutkimusta mitataan julkaisujen lukumäärällä. On tietysti mahdollista, että ison ryhmän tutkimus on parempaa ja siihen viitataan enemmän, mutta siteerausten lukumäärä pitäisi korjata tekijöiden lukumäärällä. Kun kymmenen kirjoittajan julkaisua siteerataan 200 kertaa, se on keskiarvona kirjoittajille aivan sama kuin jos yhden kirjoittajan julkaisua siteerataan 20 kertaa.

Kirjoittajien lukumäärän vaihtelu voidaan huomioida käytämällä painokertoimia, jotka summautuvat ykköseen. Painokerrointen avulla voidaan laskea julkaisujen ja siteerausten korjattuja kokonaismääriä sekä h-indeksejä, jotka ovat sopusoinnussa alaluokkien matematiikan opetuksen kanssa niin, että omenoiden tai julkaisujen lukumäärä ei itsestään lisääny, kun niitä jaetaan pienempiin ryhmiin. Painokerrointen valinta puolestaan on sopimusasia.

Taulukossa 2 tarkastelen Topparin valitsemissa suomalaisia huippututkijoita. Kävin läpi kaikkien h-indeksiin sisältyvät julkaisut ja laskin kirjoittajien lukumäärän kussakin julkaisussa ja poimin

tiedon, onko kyseinen tutkija ensimmäinen tai viimeinen kirjoittaja. Kirjoittajien lukumäärissä on valtava vaihtelu. Ruoslahden h-indeksijulkaisuissa on kahdeksan sellaista, joissa hän on ainoa kirjoittaja, ja 32 on sellaisia, joissa hän on toinen kahdesta kirjoittajasta. Toisena ääripäänä on Tuomilehto, jolla on 17 julkaisua yli 20 hengen ryhmissä. On matemaattisesti kyseenalaista verrata tutkijoita ottamatta huomioon sitä, että kirjoittajamäärien keskiarvo vaihtelee yli nelinkertaisesti.

Laskin Topparin valitsemissa tutkijoille korjatut h-indeksit kahdella tavalla. Ensiksi jaoin julkaisun puoliksi ensimmäisenä ja viimeisenä mainituille tekijöille (EV), jotka yleensä kantavat suu-

**TAULUKKO 2.** Suomalaisten tutkijoiden h-indeksien korjaaminen kirjoittajien lukumäärillä. Julkaisu ja niiden siteerausmäärät on poimittu Science Citation Index -tietokannasta. Kirjoittajien lukumäärä koskee h-indeksin julkaisuja. Indeksit on korjattu seuraavalla tavalla. EV: ensimmäisen ja viimeisen kirjoittajan painokerroin on 0,5 ja muiden 0,0, TP: painokerroin on jokaiselle kirjoittajalle samansuuruinen  $1/N$  siten, että  $N$  on kirjoittajien lukumäärä. Ruoslahden julkaisuissa kirjoittajien lukumäärän mediaani on neljä ja Kallioniemen julkaisuissa yhdeksän ja muiden tutkijoiden mediaanit ovat näiden välissä.

	Korjaamaton h-indeksi	Kirjoituksia, joissa kirjoittajien lukumäärä					Kirjoittaja		Korjattu h-indeksi	
		1	2-3	4-10	11-20	> 20	keskimäärin	maksimi	EV	TP
Ruoslahti E	133	8	58	65	2	0	3,9	13	56,5	45,8
Alitalo K	99	0	13	70	16	0	7,3	16	31,0	17,6
Tuomilehto J	83	0	2	51	13	17	17,2	117	9,0	10,5
Laakso M	73	4	6	49	11	3	9,3	166	24,0	16,9
Salonen JT	72	2	9	59	2	0	5,9	14	27,5	15,9
Kallioniemi OP	69	1	2	42	17	7	11,0	43	15,0	9,7
Palotie L	69	0	5	43	16	5	11,2	78	17,5	9,8
Virtanen I	68	0	17	51	0	0	4,8	10	20,0	16,2
Suurimman ja pienimmän suhde	2,0						4,4		6,3	4,7

rimman vastuun työstä. Toiseksi jaoiin julkaisun kaikille kirjoittajille tasapuolisesti (TP) niin, että kukin kirjoittaja sai samansuuruisen osuuden julkaisusta. Oikeudenmukainen painotus on jossakin näiden ääripäiden välissä.

Korjatut h-indeksit ovat jakautuneet aivan eri tavalla kuin korjaamaton (taulukko 2). Suurimman ja pienimmän ero on yli kaksinkertaistunut verrattuna korjaamattomaan h-indeksiin, ja tutkijoiden järjestyskin on muuttunut. On mahdollista, että korjatun h-indeksin ääripäiden väliin löytyy taulukon ulkopuolisia suomalaisia huippututkijoita, ja siinä suhteessa Topparin henkilövalinta voi olla harhaanjohtava.

Oletus, että siteerausmäärä kertoo suoraan julkaisun merkittävyyden, on h-indeksin perusta. Tuo oletus ei kuitenkaan pidä paikkaansa, mitä voidaan havainnollistaa tarkastelemalla kahta viime vuosisadan poikkeuksellisen merkittävää julkaisua. Vuonna 1951 Lowry ym. julkaisivat menetelmän proteiinipitoisuuden määrittämiseksi, ja siitä tuli maailman eniten siteerattu julkaisu: yli 100 000 siteerausta vuoteen 2006 mennessä. Vuonna 1953 Watson ja Crick julkaisivat DNA:n rakenteen, joka on yksi viime vuosisadan tärkeimmistä tieteellisistä keksinnöistä ja vuoden 1962 Nobelin palkinnon aihe. Vuoteen 1962 mennessä DNA:n rakenteen kuvaavaa julkaisua oli siteerattu vain 393 kertaa mutta Lowryn julkaisua jo 2 982 kertaa. Viime vuosina Watsonin ja Crickin tutkimukseen

on viitattu yhä enemmän (Strasser 2003), mutta silti vain 2 955 kertaa vuoteen 2006 mennessä. Jos siteerausmäärä olisi luotettava mittari tutkimuksen tärkeyden arviointiin, niin proteiinimäärittysmenetelmän kehittäminen olisi huomasti tärkeämpi asia kuin DNA:n rakenteen selvittäminen.

Monet asiat vaikuttavat siihen, kuinka paljon kuhunkin julkaisuun viitataan, ja tutkimuksen syvälinen merkitys on vain yksi asia monien joukossa (Cantell 1994, Tuomisto 1999). Suuri siteerausmäärä ei todista, että työ olisi mullistavaa ja uusia uria avaavaa, ja pieni ei todista, että tutkimus olisi huonoa ja merkityksetöntä. Päätelmä yksittäisen tutkimuksen laadusta täytyy tehdä lukemalla tutkimus ja miettimällä sen merkitystä kyseisellä tutkimuskentällä (Raivio 2004).

Jos on karkea käsitys niistä asetelmista, joissa todella suuret tieteelliset keksinnöt on synnytetty (Watson 1968, Crick 1990, Mullis 1990, Hakala 1996 ja 2002, Marshall 2006, Warren 2006), niin julkaisuluettelon pituus, siteerausten kokonaismäärä ja h-indeksi vaikuttavat kovin huonoilta mittareilta tutkimuksen merkityksen arvioimiseksi. Tuollaisiin mittareihin perustuva arviointia kuitenkin tehdään paljon, ja Topparin kuvaama h-indeksikin julkaistiin arvostetussa lehdessä.

Tieteessä pyritään tarkkoihin mittauksiin ja mahdollisten virhelähteiden huolelliseen pohdintaan. Tämän johdosta voisi olettaa, että keskinäisessä vertailus-

saan tutkijat edellyttäisivät samanlaista kriittisyyttä, johon he ovat tottuneet omassa päivätyössään. Siksi on hämmästyttävää, että tutkijoita vertaillaan laskemalla korjaamattomia julkaisujen kokonaismääriä ja h-indeksejä. Tällaisia menetelmiä käytettäessä pitäisi käyttää painokertoimia, jotka summautuvat ykköseen, niin että samaa julkaisua ei lasketa moneen kertaan eri ihmisten ansioksi.

Korjaamattomien laskentatapojen käyttö on isojen tutkimusryhmien johtajien edun mukaista, koska he ovat tehokkaasti verkostuneita ja voivat saada pienillä panoksilla nimensä lukuisiin julkaisuihin. Vaikka Suomen tieteelle on tärkeitä, että meillä on tehokkaasti verkostuneita senioritutkijoita, tuloksellisuus pitäisi laskea matemaattisesti moitteettomalla tavalla. Jos panos on pieni, niin tuloksesta pitäisi laskea vain pieni osuus kyseisen tutkijan ansioksi.

Korjaamattomat laskutavat antavat pienten ryhmien tehokkuudesta vähättelevän kuvan. Taulukko 1 kertoo, että kaikkien tuoreiden nobelistien julkaisuissa on keskimäärin alle kuusi kirjoittajaa (mediaanit välillä 2–5), eli he ovat toimineet pienissä ryhmissä kutakin julkaisua kohti. Suuri kirjoittajamäärä ei siis ole edellytys hyvälle tutkimukselle, ja taulukon 1 perusteella tilanne voi olla pikemminkin päinvastainen. Luovuus on rohkeutta toisinajatteluun, ja uusi idea syntyy yleensä vain yhdessä päässä (Cantell 1994, Huttunen 1995). Sen johdosta tutkimusryhmien

suuri koko ei saisi olla tavoitteena itseisarvo. Siinä suhteessa olen Topparin kanssa samaa mieltä, että varmaankaan Suomen lääketiede ei siitä hyötyisi, jos huippu-tutkimusryhmille siirrettäisiin vielä entistäkin suurempi osuus tieteeseen suunnatuista resursseista.

Tämän kirjoituksen tarkoitus ei ole etsiä suomalaisten huippu-tutkijoiden paremmuusjärjestystä. Yksinkertaisella indeksillä tutkijat saadaan nopeasti ja helposti järjestykseen, mutta sellainen järjestys voi olla täysin merkitsetön. Vakavasti otettava tutkijoiden arviointi pitäisi perustaa tutkimuksen sisältöön eikä pelkkään bibliometriaan, kuten tässä lehdessä on aiemminkin esitetty (Raivio 2004).

\* \* \*

Kiitän Pekka Louhialaa, Ossi Rahkosta ja Heikki Vuorista tämän kirjoituksen kriittisestä lukemisesta ja parannusehdotuksista.

### Kirjallisuutta

- Cantell K. Mietteitä tieteen mittareista. *Duodecim* 1994;110:2098–103.
- Crick FJ. What mad pursuit: a personal view of scientific discovery. Lontoo: Penguin, 1990.
- Hakala J. Lääketieteen nobelstit, luovuus ja tieteellinen löytäminen. *Suom Lääkäril* 1996;51:962–6.
- Hakala JT. Luova prosessi tieteessä. Helsinki: Gaudeamus, 2002.
- Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci* 2005;102:16569–72.
- Huttunen M. Impaktia, laatua vai mietiskelyä? *Duodecim* 1995;111:1815–17.
- Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein measurement with the Folin phenol reagent. *J Biol Chem* 1951;193:265–75.

- Marshall BJ. *Autobiography*. 2006. [http://nobelprize.org/nobel\\_prizes/medicine/laureates/2005/](http://nobelprize.org/nobel_prizes/medicine/laureates/2005/)
- Mullis KB. The unusual origin of the polymerase chain reaction. *Sci Am* 1990;262(4):56–65.
- Raivio K. Tuhoako impaktiluku vertaisarvioinnin? *Duodecim* 2004;120:1195–6.
- Rautelin H, Kosunen TU. Nobelin palkinto helikobakteerin tutkijoille. *Duodecim* 2005;121:2631–3.
- Strasser BJ. Who cares about the double helix? Collective memory links the past to the future in science as well as history. *Nature* 2003;422:803–4.
- Tuomisto J. Sitaatio- ja impaktiluvut – isäntä vai renki? *Duodecim* 1999;115:142–5.
- Warren JR. *Autobiography*. 2006. [http://nobelprize.org/nobel\\_prizes/medicine/laureates/2005/](http://nobelprize.org/nobel_prizes/medicine/laureates/2005/)
- Watson JD, Crick FHC. Molecular structure of nucleic acids – a structure for deoxyribose nucleic acid. *Nature* 1953;171:737–8.
- Watson JD. *Kaksoiskierre*. Helsinki: Weilin Göös, 1969.

**HARRI HEMILÄ, dosentti, yliopistonlehtori**  
harri.hemila@helsinki.fi  
Helsingin yliopiston  
kansanterveystieteen laitos  
PL 41, 00014 Helsingin yliopisto