

Om bruk av bibliometri på individnivå

En veiledning for forskningsinstitusjoner og forskningsledere

Fra UHR-Publisering, mai 2020.

Innledning

Publiseringssindikatoren er utviklet for å gi oversikt over forskningen på aggregert nivå og er en del av finansieringsmodellen for forskning i universitets- og høyskolesektoren, instituttsektoren og helseforetakene. Den gir årlig statistikk med nasjonal sammenlignbarhet og bidrar med informasjon til evalueringene av norsk forskning.

Siden innføringen i 2005 har indikatoren også vært brukt og misbrukt til andre formål lokalt og på individnivå. Evalueringen av indikatoren i 2014¹ anbefalte derfor at bruk av indikatoren burde diskuteres mellom forskningsinstitusjonene for å fremme god forskningsledelse og gjensidig læring. UHR-Publisering fulgte opp på flere måter, blant annet ved å publisere et notat med råd om lokal bruk av indikatoren.

I mai 2020 har vi revidert notatet, både med en kort versjon² og med denne utvidete versjonen, hvor vi i tillegg til publiseringssindikatoren også omtaler *Journal Impact Factor (JIF)* og *H-indeks* og ser de tre indikatorene i lys av til *DORA-erklæringen* og *Leiden-manifestet*, som begge er internasjonalt kjente kilder til råd om bruk av bibliometri på individnivå, og som våre råd er i tråd med. Vi presenterer rådene først.

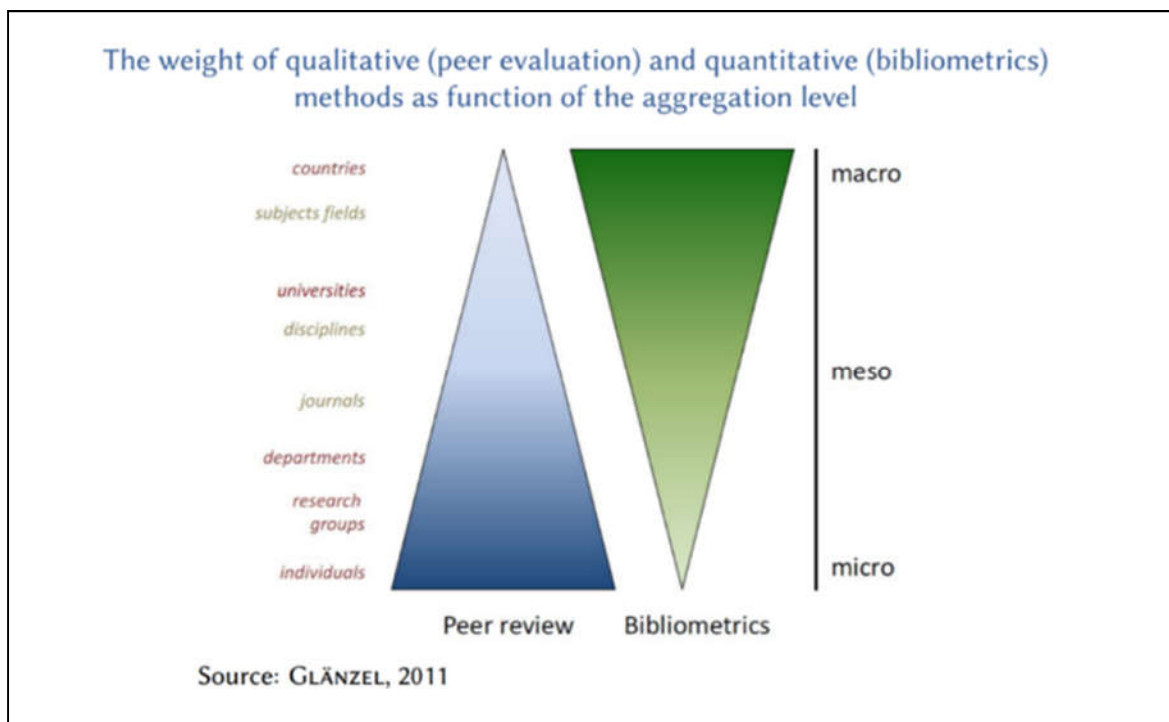
Fire råd om bruk av bibliometri på individnivå

- 1. Bibliometri skal ikke brukes alene.** Ansettelse, opprykk, karriereoppfølging og fordelinger av oppgaver og ressurser skal bygge på helhetsvurderinger. I de fleste tilfeller skal ikke bare forskningen vurderes. Men også i vurderinger av forskningen har bibliometriske indikatorer begrenset verdi fordi de kun er tilbakeskuende, ikke tar hensyn til forskjeller i forutsetninger og ikke kan erstatte et beslutningsansvar.
- 2. Bibliometri ser ikke framover.** Bibliometri peker bakover mot tidligere forskningsaktivitet. Vurderinger i forbindelse med ansettelse, opprykk eller tildeling av ressurser skal også se framover og vurdere mulighetene for å innfri utlyste krav og forventninger.
- 3. Bibliometri forstår ikke forskjeller i forutsetninger.** Forskning og faglig aktivitet foregår i mer eller mindre aktive faser alt etter hvilke andre oppgaver man er engasjert i, hvilke ressurser som er tilgjengelige til enhver tid, og hvilken type prosjekter og samarbeid man er med i. Forskningsledere har ansvar for å forstå denne variasjonen ved fordeling av ressurser og ved ansettelse og opprykk.
- 4. Bibliometri kan ikke beslutte.** Når søknadene er mange og det er behov for å foreta et utvalg (screening) før nærmere vurderinger av relevante kvalifikasjoner, kan bibliometri eventuelt være ett av flere hensiktsmessige hjelpemidler. Men verdien av å bruke bibliometri vil være avtakende jo nærmere beslutningen man kommer – se figuren, som viser den avtakende relevansen av kvantitative versus kvalitative metoder i vurdering av forskning på ulike nivåer i forskningssystemet.³

¹ https://npi.nsd.no/dok/eval2014/Evaluering_af_den_norske_publiceringsindikator_2014.pdf

² Lenke til kort versjon

³ Etter Wolfgang Glänzel: Introduction to scientometrics, European Summer School for Scientometrics, 2011.



Vi understreker at dette kun er ment som råd. Vurdering av forskere på individnivå er et faglig ansvar for institusjonene og må tilpasses institusjonenes innretning og fagområdeforskjeller.

Et internasjonalt perspektiv

Bruk av bibliometri på individnivå i forskningen blir stadig mer utbredt – og omdiskutert – overalt i forskningens verden. *DORA-erklæringen* og *Leiden-manifestet*, som begge er internasjonalt kjente kilder til råd om bruk av bibliometri på individnivå, gir svar til denne situasjonen som vi ønsker å gjengi her.

I Norge brukes vanligvis disse tre indikatorene på individnivå: *Journal Impact Factor (JIF)*, *H-indeks* og *Publiseringspoeng*. De to første har internasjonal utbredelse og brukes særlig i naturvitenskap, teknologi og medisin. Den tredje er skapt i Norge og brukes mest i Norden, særlig i humaniora og samfunnsvitenskap.

I denne veiledningen omtaler vi først *DORA-erklæringen* og *Leiden-manifestet*. Deretter gir vi i et innblikk i hvordan de tre indikatorene beregnes og forklarer hvorfor de er problematiske å bruke på individnivå før vi konkluderer med de fire rådene.

DORA-erklæringen

The Declaration on Research Assessment (DORA),⁴ som flere organisasjoner innen norsk forskning har signert eller er i ferd med å signere, har følgende hovedanbefaling:

Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

⁴ <https://sfdora.org/>

Anbefalingen begrunnes blant annet med:

--- the need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations;

--- the need to assess research on its own merits rather than on the basis of the journal in which the research is published;

---the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

DORA-erklæringen nevner eksplisitt *Journal Impact Factor*, men publiseringspoeng kan også kalles en 'journal-based metrics'. Dette kommer vi tilbake til. Man kan også merke seg at DORA-erklæringen fokuserer på nettopp de sammenhengene hvor forskere blir vurdert individuelt: ved ansettelse, opprykk og ved søknader om forskningsmidler.

DORA-erklæringen har 11 referanser til vitenskapelige artikler. Den eldste av dem, som er fra 1997 og stod i *British Medical Journal*, er av en norsk forsker, Per O. Seglen: 'Why the impact factor of journals should not be used for evaluating research'.⁵

Leiden-manifestet

Leiden-manifestet⁶ (leidenmanifesto.org) ble opprinnelig publisert som en artikkel i *Nature* i 2015. Den er signert noen av verdens ledende bibliometriske forskere og basert på diskusjoner hvor hele det internasjonale forskningsfeltet deltok (STI-konferansene). Leiden-manifestet er altså den bibliometriske forskningens egne råd om bruk av bibliometri på individnivå. Manifestet har nå en egen nettside med en video som framstiller de ti rådene som gis.⁷

Det første, generelle rådet advarer mot å bruke bibliometri alene til å vurdere eller ta beslutninger i forskningen:

Quantitative evaluation should support qualitative, expert assessment. *Quantitative metrics can challenge bias tendencies in peer review and facilitate deliberation. This should strengthen peer review, because making judgements about colleagues is difficult without a range of relevant information. However, assessors must not be tempted to cede decision-making to the numbers. Indicators must not substitute for informed judgement. Everyone retains responsibility for their assessments.*

Det syvende rådet er i tillegg interessant i vår sammenheng fordi det handler om bruk av bibliometri på individnivå, og fordi *h-indeks* er spesielt nevnt:

Base assessment of individual researchers on a qualitative judgement of their portfolio. *The older you are, the higher your h-index, even in the absence of new papers. The h-index varies by field: life scientists top out at 200; physicists at 100 and social scientists at 20–30. It is database dependent: there are researchers in computer science who have an h-index of around 10 in the Web of Science but of 20–30 in Google Scholar. Reading and judging a researcher's work is much more appropriate than relying on one number. Even when comparing large numbers of researchers, an approach that considers more information about an individual's expertise, experience, activities and influence is best.*

⁵ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf>

⁶ <http://www.leidenmanifesto.org/>

⁷ <https://vimeo.com/133683418>

Journal Impact Factor

Journal Impact Factor (JIF) ble lansert i begynnelsen av 1960-årene av Eugene Garfield (1925-2017), grunnleggeren av *Science Citation Index* (nå: *Web of Science*), som en måte å måle viktigheten av vitenskapelige tidsskrifter. JIF var ment som et redskap for bibliotekene når de skulle avgjøre innkjøp. Formålet var altså et helt annet enn å vurdere forskning på individnivå.

JIF måler den gjennomsnittlige siteringshyppigheten i et bestemt år. Standardmetoden er å beregne gjennomsnittet for artikler publisert gjennom to år med utgangspunkt i antall siteringer de har mottatt ved utgangen av det tredje året. JIF for et tidsskrift i 2019 er derfor antallet siteringer som artiklene fra 2017 og 2018 har mottatt ved utgangen av 2019 dividert på antallet artikler fra 2017-18.

I en artikkel i *Nature* i 2001 skrev Garfield følgende om å bruke JIF til å vurdere den enkelte artikkel i vedkommende tidsskrift:⁸

It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the journal impact factor as a surrogate. The latter practice is fraught with difficulties, as Seglen and others have pointed out.

Seglen hadde vist at antallet siteringer til artiklene i et tidsskrift er sterkt skjevfordelt: Noen få artikler mottar størstedelen av siteringene, mens størstedelen av artiklene mottar få eller ingen siteringer. Funnet er senere bekreftet av en større studie av norsk forskning.⁹

Gjennomsnittet for tidsskriftet kan altså ikke brukes som en antakelse om hvor ofte den artikkelen man vurderer er sitert. JIF sier bare noe om tidsskriftets betydning. En vurdering på individnivå innebærer dessuten ofte en sammenligning med andre forskere. JIF varierer mellom fagområder og forskningsfelt på samme måte som siteringshyppighet generelt. Bruk av JIF på individnivå kan derfor forstyrre en rettferdig sammenligning og føre til prioritering av bestemte former for forskning.

Det er verdt å merke seg at DORA-erklæringen ikke innebærer en kritikk av JIF i seg selv. Så lenge den brukes til det formålet den er beregnet på, å måle et tidsskrifts vitenskapelige innflytelse i sitt fagfelt, er dette verken i strid med DORA-erklæringen eller med internasjonal bibliometrisk forskning.

H-indeks

H-indeks er den eneste av de tre indikatorene som er utviklet for å måle forskning på individnivå. Den er også den eneste som er utviklet utenfor forskningsfeltet bibliometri. Den ble forslått i en artikkel i tidsskriftet PNAS i 2005 av fysikeren Jorge Eduardo Hirsch som en forenklet måling av en forskers resultatoppnåelse.¹⁰ Den kombinerer produktivitet (antall artikler) med gjennomslagskraft (antall siteringer) ved å defineres som det høyeste antall artikler man har utgitt med et tilsvarende høyt antall siteringer.

En del av problemene med h-indeksen er nevnt i sitatet ovenfor fra det syvende rådet i Leiden-manifestet. I tillegg kan nevnes at beregningen forutsetter at enhver artikkel kun har én forfatter mens forskeres produktivitet i stor grad avhenger av det gjennomsnittlige antallet medforfattere. Denne og flere begrensninger er diskutert av Tor-Arne Hagve i en artikkel i Tidsskrift for Den norske legeförening i 2011.¹¹

⁸ <https://www.nature.com/articles/35079156>

⁹ <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174205>

¹⁰ <https://www.pnas.org/content/102/46/16569>

¹¹ <https://tidsskriftet.no/2011/12/kronikk/h-indeks-et-godt-mal-pa-forskningsaktivitet>

Hvis h-indeks brukes direkte i vurderinger av ansettelse, opprykk eller søknader om forskningsmidler, vil man systematisk gi muligheten til de forskerne som er kommet lengst i karrieren og som er knyttet til fag hvor både siteringshyppigheten og antall medforfattere er høyt. Unge forskere vil ikke ha en sjanse.

Publiseringsindikatoren

Publiseringsindikatoren ble utviklet på oppdrag fra Kunnskapsdepartementet av Universitets- og høyskolerådet i samarbeid med NIFU i 2003-2004. Indikatoren skulle bidra til å representere og stimulere forskningsaktivitetene i finansieringssystemet for universiteter og høyskoler. I dag brukes den også i finansieringssystemet for forskning i instituttsektoren og helsesektoren. Hensikten var aldri å representere eller måle forskningen på individnivå.

Publiseringspoeng beregnes på institusjonsnivå med utgangspunkt i andelen av publikasjonens forfattere som krediterer vedkommende institusjon i publikasjonen. (Hvis en forfatter krediterer mer enn én institusjon, øker det samlede antallet forfatterandeler.) For hver publikasjon som kan tilknyttes institusjonen, multipliseres kvadratroten av institusjonens samlede forfatterandeler med et poeng som er avhengig av publikasjonens form (artikkel i serie eller tidsskrift, artikkel i antologi, monografi) og nivået på publiseringskanalen (nivå 1 eller 2). Denne beregningsmåten har vist seg å balansere godt mellom de ulike fagtradisjonenes publiseringsmønstre. Kvadratroten gjør det dessuten omstendelig å beregne indikatoren på individnivå.

I mange fag er den enkelte forskers bidrag til en publikasjon gjenspeilet i rekkefølgen av forfattere. Publiseringsindikatoren tar ikke hensyn til rekkefølgen. Dette er ikke nødvendig på institusjonsnivå, men blir et problem hvis indikatoren brukes på individnivå.

Publiseringsindikatoren er heller ikke egnet til å måle verdien av den enkelte publikasjon. Problemet er det samme som med Journal Impact Factor (JIF): Publiseringskanalen kan ikke si noe direkte om publikasjonen. Men som nevnt ovenfor, kan JIF brukes til si noe om et tidsskrifts vitenskapelige innflytelse i sitt fagfelt. Derfor er JIF en del av den informasjonen som fagorganene bruker når de vurderer hvilke tidsskrifter som skal være på nivå 2 i Norge. Men i siste instans skal vurderingen være kvalitativ: Nivå 2 i et fagfelt skal representere de mest prestisjetunge tidsskriftene fra et internasjonalt perspektiv og gjenspeile en ønskelig retning for publiseringsmønstret i fagfeltet. Norge er ett av svært få land som forankrer slike vurderinger i fagmiljøene. De fleste land bruker kun JIF.

På ulike nivåer i norsk forskning brukes publiseringsindikatoren som ledelsesinformasjon eller i kriteriene for tildeling av ressurser. Dette kan skje på en forsvarlig måte, også ved bruk på individnivå, men bruken bør kunne diskuteres, helst på tvers av institusjoner og fakulteter og i det minste mellom ansatte og ledelse. Bruk individnivå følger uansett ikke av myndighetenes anbefalinger og er alltid den enkelte institusjons eget ansvar. I rapporten *Vekt på forskning* (UHR, 2004), hvor publiseringsindikatoren ble introdusert, sies det at (s. 59):¹²

... å fordele alle forskningsressurser etter antallet publikasjoner et bestemt år i fortiden er det samme som å sette til side et kvalitativt skjønn, en framtidsrettet forskningsstrategi og institusjonsledelsens ansvar for å ha innsikt i at forskningen blant de ansatte vil foregå i mer eller mindre aktive faser alt etter hvilke andre oppgaver de er engasjert i, hvilke ressurser som ellers er tilgjengelige til enhver tid, og hvilken type prosjekter det arbeides med.

¹² https://npi.nsd.no/dok/Vekt_pa_forskning_2004.pdf

Referanser

Garfield, E. (2001). Impact factors, and why they won't go away. *Nature*, 411(6837):522.

Hagve, T.A. (2011). H-indeks – et godt mål på forskningsaktivitet? *Tidsskr Nor Legeforen*, 131, 2494-6.

Hicks, D., Wouters, P., Waltman, L., Rijcke, S. D., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431. doi:10.1038/520429a

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ*, 314(7079), 497-497.

Vekt på forskning. Nytt system for dokumentasjon av vitenskapelig publisering. Innstilling fra faglig og teknisk utvalg til UHR. Oslo: Universitets- og høyskolerådet, 2004.

Zhang, L., Rousseau, R., Sivertsen, G. (2017). Science deserves to be judged by its contents, not by its wrapping: Revisiting Seglen's work on journal impact and research evaluation. *PLoS ONE*, 12(3): e0174205. doi:10.1371/journal.pone.0174205