

Bruk av siteringer og tidsskriftimpaktfaktor til forskningsevaluering

Af Per O. Seglen

Dagens forskningspolitiske utvikling følger omtrent samme mønster i alle land: reduserte bevilgninger til fri grunnforskning, økt vektlegging av forskningens samfunnsmessige nytte, ønsker om sterkere politisk styring, og krav om dokumentasjon av forskningskvalitet i en form som kan forstås og brukes av et stadig voksende forskningsbyråkrati. Den tradisjonelle vurderingen av forskningskvalitet ved hjelp av fagekspertter (peer review) er svært tids- og arbeidskrevende, og gir resultater i en form som kan være vanskelig å behandle administrativt og statistisk. Mange har derfor kastet sine øyne på alternative evalueringmetoder, først og fremst på såkalte *bibliometriske indikatorer* som kan beskrive forskningsresultater på en kvantitativ måte. Bibliometrien henter sin informasjon fra internasjonale databaser, og benytter dels opptelling av antall artikler som et mål for produktivitet, dels opptelling av siteringer til publiserte tidsskriftartikler som et mål for forskningskvalitet. På grunnlag av disse basisparameterne kan det konstrueres diverse relative indikatorer

som f.eks. kan beskrive tilstand og utviklingstrender innen forskjellige fagområder i et bestemt land eller i en bestemt institusjon, i forhold til verden forøvrig. I enkelte tilfelle brukes også bibliometriske indikatorer til evaluering av enkeltforskere og forskningsgrupper. Da dette kan få bevilgningsmessige konsekvenser for den enkelte forsker, er det viktig å være klar over hva en bibliometrisk evaluering innebærer. I det følgende vil vi forsøke å sette kritisk søkelys på siteringsbasert evaluering, og reise spørsmålet om siteringsbaserte indikatorer i det hele tatt kan brukes til evaluering av enkeltforskere og forskningsgrupper.

Siteringshyppighet: mange feilkilder ved registrering

Siteringshyppigheten (siteringsraten) for tidsskriftartikler kan en finne i *Science Citation Index* (SCI) som utgis hvert år av Institute for Scientific Information (ISI) i Philadelphia, og som er tilgjengelig både som trykksak, som kompakt-disk, og ved direkte tilgang til databasen. Indeksen er en liste

over artikler som ble sitert (dvs. brukt som referanser) forrige år, evnt. en større periode, innen det fagområdet som dekkes av indeksen.

Tabell 1. Måletekniske problemer ved innhenting af siteringsdata

1. Ufullstendig dekningsgrad i databasen (nasjonalt; fagfelt)
2. Variabelt tidsskrift-utvalg
3. Sen registrering
4. Registrering bare på første forfatter
5. Språkproblemer (f.eks. æ, ø, å)
6. Trykkfeil (opptil 25%)
7. Synonymi (samme artikkel registrert i flere varianter)
8. Homonymi (flere forfattere med samme navn; spesielt problem i Japan)
9. Valg av tidsvindu (korttids/langtids-indeks; årlig/kumulativt)
10. Flerforfatter-kreditering (dele på antall forfattere?)
11. Kostnader (database-søk er dyre)

Innhenting av siteringsdata fra SCI er beheftet med en rekke feilkilder (Tabell 1). Databasen dekker ca. 3500 av de mest brukte tidsskriftene innen naturvitenskap og medisin, noe som bare er en brøkdel av det antallet tidsskrifter som virkelig finnes innen disse fagområdene. Dekningsgraden er altså ufullstendig, og varierer sterkt fra fag til fag. For et internasjonalisert fag som kjemi kan det regnes med at ca. 90% av aktuelle tidsskrifter finnes i SCI-databasen. Dekningsgraden for biologi, som i langt større grad beskjeftiger seg med nasjonale og regionale problemstillinger, er imidlertid bare ca. 30%. Bøker registreres ikke, dvs. forskere og fag som i vesentlig grad benytter bøker som publikasjonsmedium blir diskriminert. Databasen er sterkt dominert av engelskspråklige tidsskrifter, noe som diskriminerer land med et betydelig innslag av nasjonalspråklige tidsskrifter (Tyskland, Frankrike, Russland).

Tidsskrift-utvalget kan variere fra år til år, noe som kan gi et feilaktig inntrykk av brå forandringer i

siterthet. En tilsynelatende tilbakegang i norsk kjemisk forskning for en del år tilbake, viste seg f.eks. å bero på at det nasjonale tidsskriftet »Kjemi« ikke lenger ble registrert. Registreringen kan av og til være såpass forsinket at artikler publisert mot slutten av et kalenderår først kommer med på neste års statistikk. Siteringer registreres bare under den siterte artikkelens førsteforfatter, hvilket vanskeliggjør innsamling av data vedrørende enkeltforskere, medmindre det tas utgangspunkt i forfatternes egne publikasjonslister.

Spesialbokstaver, som vi har i de skandinaviske alfabetene (æ, ø, å), brukes ikke i ISI-databasen, og translitterasjonen er ikke konsekvent: ø blir noen ganger til o, andre ganger til oe. Regulære trykkfeil er et annet betydelig problem: i ett tilfelle fant jeg at en og samme referanse var registrert i 70 forskjellige trykkfeils-varianter, altså en høy grad av *synonymi*. Det kan være trykkfeil både i navn og bibliografisk referanse: sammensatte etternavn og spesialtegn er særlig vanskelige, og ikke engang forfatterne selv er helt konsekvente m.h.t. navnebruk. Et annet problem er *homonymi*, dvs. at flere forfattere har samme navn. Homonymien gjør seg særlig gjeldende i Japan: K. Suzuki står f.eks. registrert med flere hundre artikler pr. år. Alle disse måletekniske problemene gjør sitt til at siteringsdata lett kan bli upålitelige.

Et spesielt problem i evalueringssammenheng er valget av tidsvindu for registrering av siteringer. En korttidsindeks (basert på f.eks. de første 2-3 årene etter publikasjon) kan gi skjevheter i forhold til tiden på året artikkelen er publisert, og vil favorisere de sentrale, engelskspråklige tidsskriftene som raskt når fram til flertallet av lesere. Nyhetspregede artikler med kort levetid vil videre favoriseres i forhold til artikler med en mer varig bruksverdi. En langtidsindeks (8-10 år) vil prinsipielt være å foretrekke, men er naturligvis svært upraktisk i evalueringssammenheng.

Et annet evalueringsteknisk problem er hvordan kreditering av flerforfatterarbeider skal foretas. Skal alle forfattere tilskrives alle siteringer (noe

som vil innebære en total overkreditering, og en voldsom favorisering av publikasjonsmessig samarbeid), skal de fordeles likt på forfatterne, eller bør det benyttes en fordelingsnøkkel som f.eks. tilskriver førsteforfatter og prosjektleder større andeler enn de øvrige? Hva med samarbeid mellom institusjoner? Disse spørsmålene er ikke avklart på noen entydig måte, og håndteres forskjellig fra evaluering til evaluering (i den grad de i det hele tatt reises).

Endelig bør det nevnes at datasøk kan bli kostbare, hvis en skal utnytte de mulighetene som gis ved direkte aksess til SCI-databasen. For visse typer bibliometriske studier kan CD-ROM-versjoner av 5-årslige SCI-data kjøpes, men her ligger både de siterte artiklene og siteringsvinduet innenfor samme 5-års-intervall, noe som gjør dataene dårlig egnet til evalueringsformål.

Artikler siteres ikke bare fordi de er gode

Et hovedproblem ved bruk av siteringsdata er spørsmålet: hva er egentlig siteringen et mål for? Alle som har publisert en del selv kan svare på det: vi refererer først og fremst til artikler vi har bruk for i sammenheng med det vi skriver om, dvs. artikler som dokumenterer eller belyser metoder, opplysninger eller argumenter brukt i vår egen artikkel. Siteringer er altså primært et uttrykk for *bruksverdien innen forskning*. Sekundært kommer en lang rekke forhold inn i bildet (Tabell 2) før det eventuelt blir aktuelt å vurdere kvaliteten eller originaliteten av en artikkel (1). Noen forskere er nok kvalitetsbevisste i valget av referanser, men de fleste er det helt åpenbart ikke. Siteringshyppighet måler nok noe som kan tillegges verdi i forskersamfunnet (påvirkningskraft, innflytelse, impakt), men er ikke noe mål for vitenskapelig kvalitet.

Tabell 2. Problemer vedrørende valg av referanser

1. Primærkriterier er bruksverdi innen forskning, ikke kvalitet
2. Ufullstendig siteringsbelegg (begrensning av antall referanser medfører at bare en brøkdel av bakgrunnsstoffet siteres)

3. Allmennkunnskap siteres ikke (glemsel ved inkorporering)
4. Bruk av sekundære kilder (oversiktsartikler o.a.)
5. Argumentativ sitering (ofte som støtte til egne argumenter)
6. Smiger (sitering av redaktører, potensielle referee'er og andre autoriteter)
7. Show-of (sitering av de nyeste "in"-artiklerne)
8. Referanse-kopiering (avskrift av andres referanser)
9. Konvensjoner: metoder siteres; reagenser ikke
10. Selvsitering
11. Kollegial sitering (reflekterer ofte informasjonsoverføring ved direkte kontakt)

Et vitenskapelig arbeid er ofte basert på en kunnskapsbakgrunn *bestående* av hundrevis av relevante artikler, men av plasshensyn kan bare et begrenset utvalg av disse inkluderes i referanselisten. Bare en brøkdel av den benyttede kunnskapen belegges altså med siteringer (1). Utvalget kan være basert på svært blandede kriterier: erfarne, kunnskapsrike forskere vil ofte prioritere originalitet og kvalitet, og sitere dem som æres bør, mens uerfarne forskere gjerne foretar mer tilfeldige valg, og har en tendens til å bruke oversiktsartikler og andre sekundære referanser. Det er dessverre relativt vanlig at kopiere referanser fra andres artikler uten at disse blir lest, noe som i enkelte tilfelle kan avsløres ved at bestemte trykkfeilsvarianter opptrer hyppigere enn den korrekte bibliografiske referansen. Både referansekopiering og litteratursøk via tidligere artikler vil favorisere de referansene som allerede er kommet seg inn i litteraturen, noe som kan konservere en siteringsmessig ubalanse.

Siteringsadferden har en tendens til å følge visse konvensjoner som kan ha stor *betydning* for siteringshyppigheten. I biokjemisk litteratur er det eksempelvis en tradisjon for å referansebelegge benyttede metoder, men ikke benyttede reagenser. Mye brukte metoder kan oppnå svært høye siteringshyppigheter; f.eks. er både verdens og Norges mest siterte artikler metodologiske arbeider. I gjennomsnitt er derimot ikke metodologiske arbeider

noe oftere sitert enn andre vitenskapelige arbeidere: metoder som ikke brukes av andre, siteres ikke, uansett hvor geniale de måtte være.

En spesiell konvensjon er at fundamental, etablert kunnskap vanligvis ikke belegges med referanser («det er alminnelig kjent at ...»). Dette fenomenet kalles «glemsel ved inkorporering», og innebærer at de aller beste og viktigste vitenskapelige arbeidene paradoksalt nok sjelden siteres.

En stor del av referansene i et vitenskapelig arbeid er knyttet til diskusjon av resultatene. Ideelt sett bør denne være en balansert *pro et contra*-diskurs, men i praksis argumenterer forfatterne gjerne til støtte for sine egne resultater og konklusjoner (ofte en nødvendighet for å få arbeidet publisert), og velger referanser deretter. Originale innfallsvinkler og resultater som bryter med etablerte konvensjoner blir derfor ikke nødvendigvis belønnet med de siteringer de kunne fortjene. Negative siteringer forekommer i praksis sjelden; uenighet møtes som regel med ignorering snarere enn med motargumentasjon. Enkelte unntak forekommer dog; eksempelvis ble Pons og Fleischmanns «påvisning» av kald fusjon, som ansees for feilaktig, relativt høyt sitert, hovedsakelig negativt.

Det synes utbredt å inkludere referanser til personer som antas å kunne komme til å fungere som referee for arbeidet. Dette skjer gjerne ved at det innledningsvis føres opp generelle henvisninger til de ledende gruppene innen feltet. Hvis redaktør eller redaksjonsmedlemmer har utført relevant arbeid, vil de som regel være sikret sitering. Det er også mange som anser det for flatterende å inkludere referanser til «siste nytt» innen feltet, for å antyde at en er med i forskningsfronten.

En vitenskapelig artikkel bygger svært ofte på forfatterens tidligere publiserte arbeider, og selvsiteringer er derfor, vitenskapelig sett, høyst relevante. For de lavt siterte artiklens vedkommende kan imidlertid selvsiteringer utgjøre flertallet av siteringene, og disse har naturligvis liten verdi som evaluering grunnlag. SCI skiller imidlertid ikke mellom selvsiteringer og andre siteringer, og siden bare førsteforfatteren er gjengitt i database-indeksen,

kan det for utenforstående være svært vanskelig å identifisere selvsiteringer (prosjektlederen er ofte sisteforfatter).

Det er relativt vanlig å sitere nære kollegaer, f.eks. ved samme avdeling eller institutt, ofte som en konsekvens av at en mottar informasjon eller metodeopplæring gjennom direkte kontakt. Siteringen kan derfor være relevant nok, men gjenspeiler ikke nødvendigvis hverken originalitet eller kvalitet i den siterte forskningen. Avtaler om gjensidig sitering vil kunne ha en stor effekt på siteringshyppighet, men det er ikke dokumentert at slike avtaler forekommer.

Variabilitet: siteringshyppigheten varierer sterkt, selv for artikler fra en og samme forsker

Selv om en aksepterer siteringshyppighet som en interessant og relevant parameter i evalueringssammenheng, finnes det flere tekniske problemer knyttet til anvendelsen enn dem som er angitt i Tabell 1 og 2. Ett problem er den store variabiliteten i siteringshyppighet, som gjør det vanskelig å påvise statistisk signifikante forskjeller mellom forskere og forskningsgrupper. Selv artikler publisert av en og samme forsker viser en lovmessig, negativ eksponensialfordeling m.h.t. siteringshyppighet, noe som innebærer at det alltid vil være en høy grad av overlapping mellom forskjellige forskere (2,3). Selv om en forsker i gjennomsnitt siteres dobbelt så hyppig som en annen, må en gjerne opp i et materiale på 50-60 artikler for å kunne fastslå om forskjellen virkelig er statistisk signifikant. Ved de fleste evalueringer er det ønskelig å vurdere forskningsaktivitet gjennom et begrenset tidsrom, f.eks. de siste 4-5 år, og da er det ikke mange som når opp i en så høy produksjon at en forsvarlig statistisk behandling blir mulig.

Felt-effekter: siteringshyppigheten avhenger av forskningsfeltet

Den viktigste innvendingen mot bruk av siteringsdata er at siteringshyppigheten er så sterkt påvirket av siteringsdynamikken innen forskningsfeltet (Tabell 3). Dette er forhold som ikke har noe med

den enkelte artikkel å gjøre, og som sterkt vil kunne favorisere visse typer forskning. Eksempelvis inneholder biokjemiske artikler gjennomsnittlig dobbelt så mange referanser som matematiske artikler, hvilket betyr dobbelt så mange siteringer. På den annen side har matematiske artikler mye lengre levetid, dvs. referanselistene inneholder relativt sett færre nye artikler enn de biokjemiske tidskriftene. I sum innebærer disse to forholdene at en gjennomsnitts biokjemiker, ved bruk av en korttids-indeks (f.eks. siteringer de siste tre årene), automatisk blir sitert fire ganger så ofte som en gjennomsnitts matematiker (4).

Tabell 3. Forskningsfelt-effekter som påvirker siteringshyppighet

1. Antall referanser pr. artikkel (gjennomsnitt innen fagfeltet)
2. Referanse-aktualitet (i forhold til tidsvinduet for målingen)
3. Utviklingsdynamikk (ekspanderende eller skrumpende fagfelt)
4. Fagfeltets størrelse (påvirker spennvidden, men ikke gjennomsnittet)
5. Mellomfaglige relasjoner (f.eks. basal vs. anvendt)
6. Forskjeller mellom subdisipliner og spesialiteter (mikroheterogenitet)

Et annet viktig forhold er forskningsfeltets utviklingsdynamikk: innen et ekspanderende forskningsfelt vil det være forholdsvis mange flere som siterer enn som blir sitert, dvs. de som allerede befinner seg i feltet vil bli høyt sitert (pyramide-spill-effekt) (5). I et skrumpende forskningsfelt vil det være motsatt, uavhengig av kvaliteten på den gjenværende forskningen. Den absolutte størrelsen på forskningsfeltet spiller derimot ingen rolle for feltets *gjennomsnittlige* siteringshyppighet, men *maksimumsverdiene* kan naturligvis bli større i et stort felt.

De siteringsdynamiske forholdene *innen* et forskningsfelt er altså relativt lovmessige. Av større betydning er imidlertid siteringsrelasjonene *mel-*

lom ulike, men tilstøtende forskningsfelt. Det er de artiklene som sprenger grensene for sitt eget fagområde som kan oppnå de virkelig høye siteringshyppighetene. Spesielt illustrerende er relasjonen mellom basal og klinisk medisinsk forskning: den kliniske forskningen gjør utstrakt bruk av basale forskningsresultater, mens det omvendte knapt er tilfelle. Resultatet er at basale medisinske artikler i gjennomsnitt siteres 3-5 ganger hyppigere enn kliniske (6,7). Siteringsbaserte parametere egner seg altså svært dårlig til å evaluere medisinsk forskning, som ofte befinner seg i skjæringspunktet mellom klinikk og basalforskning. Jo mer klinisk orientert forskningen er, jo dårligere vil den komme ut av evalueringen.

Det finnes holdepunkter for at felt-effekter kan gi store forskjeller i siteringshyppighet også mellom nærbeslektede spesialområder, f.eks. mellom enmegrupper innen et tidsskrift (2). Det kan altså synes som om hver enkelt forsker/forskningsgruppe gjennom sin aktivitet definerer sin egen siteringsdynamiske sfære, som fastlegger sannsynligheten for å oppnå en bestemt siteringshyppighet. Denne *mikroheterogeniteten* kan innebære at det i praksis er umulig å korrigere for felt-effekter: det er ikke tilstrekkelig å firedoble matematikernes siteringstall i forhold til biokjemikernes. Ikke desto mindre gjøres det innen scientometrien til stadighet forsøk på å foreta slike korreksjoner ved å konstruere *relative* siteringsindikatorer, der siteringshyppigheten til en forskergruppe (dvs. en samling artikler) søkes uttrykt i forhold til forventet siteringshyppighet. Innenfor et fagområde kan dette f.eks. gjøres ved å dividere det reelle antall siteringer med et veid siteringsgjennomsnitt for de tidskriftene som er benyttet, basert på tidsskriftenes *impaktfaktorer* (siteringsgjennomsnitt; se nedenfor). Dette vil imidlertid straffe forskere som publiserer i tidsskrifter med høyt impakt, så en utbredt anvendelse av denne metoden ville nok raskt ha ført til endret publiseringsadferd. En annen metode er å utarbeide standardiserte korreksjonsfaktorer for de enkelte fagområdene, men også disse vil være tidsskriftbaserte, og vil ikke kunne fange opp mikroheterogenitet og mellomfaglige relasjoner

som springer ut av forskningens faktiske innhold. I praksis kan det være vanskelig å få vite hva en siteringsbasert evaluering egentlig bygger på, enten fordi de benyttede indikatorene ikke offentliggjøres, eller fordi de gis en så komplisert matematisk form at de blir vanskelige å kontrollere (særlig hvis en ikke har kjennskap til fallgrubene på dette området).

Tidsskrift-impakt

En enklere, og derfor mer benyttet måleparameter enn siteringshyppighet, er den såkalte *tidsskriftimpakt-faktor*, dvs. den gjennomsnittlige siteringshyppigheten for alle artiklene i et tidsskrift, som kan finnes i årlige *Journal Citation Reports* fra ISI (Tabell 4). Impaktfaktorene er relativt stabile over tid (8), og i de tilfelle der det finnes flere tidsskrifter innenfor samme fagområde, er det en relativt klar sammenheng mellom tidsskriftets impaktfaktor og dets prestisje i fagmiljøet. Impaktfaktoren kan altså se ut til å være en brukbar indikator på tidsskriftets kvalitet. Hvis en antar at tidsskriftgjennomsnittet er noenlunde representativt for enkelt-artiklene, burde det derfor tilsynelatende være mulig å få et tilnærmet mål for kvaliteten på en vitenskapelig artikkel simpelthen ved å plukke den tilsvarende tidsskriftimpakt-faktoren ut av tabellen.

Tabell 4. Eksempler på tidsskriftimpakt-faktorer (1991)

Cell	30,2
New England J. Med.	23,2
Science	19,6
Nature.....	19,3
Lancet.....	15,9
J. Clin. Invest.....	8,2
J. Biol. Chem.....	6,7
Biochim. Biophys. Acta	2,5
Acta Anaest. Scand	0,97
Acta Chir. Scand.	0,41
Acta Chir. Belg.....	0,13
Eur. J. Surg. Oncol.	0,00
Gjennomsnitt på verdensbasis.....	1,6

Mange av problemene knyttet til bruk av siteringshyppighet er imidlertid fortsatt tilstede ved bruk av tidsskriftimpakt (Tab. 5): det gjelder f.eks. ufullstendig dekningsgrad i databasen, og valget av tidsvindu. Tidsskriftimpakt-faktor beregnes på grunnlag av siteringer de første to årene etter publisering, altså en ekstrem korttids-indeks som favoriserer kortlivede, nyhetspregede artikler; jfr. de høye verdiene for *Nature* og *Science*. En særegenhet ved beregning av faktoren gir også skjeve utslag: det telles opp siteringer både til ordinære artikler, lederartikler, nyhetsoppslag og diskusjonsinnlegg, men de tre sistnevnte kategoriene regnes ikke med når siteringene deles på antall artikler! Tidsskrifter med mange slike bidrag, som ofte siteres hyppig, kan altså få uberettiget høye impaktfaktorer - opp til 75% for høye (9).

Tidsskrifter som ikke er inkludert i SCI tildeles ikke noen impaktfaktor, og diskrimineringen av ikke-engelskspråklige tidsskrifter blir spesielt sterk: disse siteres hovedsakelig av andre tidsskrifter med samme språk, som i liten grad er inkludert i SCI. Det tar også relativt lang tid før informasjon fra disse når fram til det engelskspråklige forskersamfunnet (ofte via sekundære kilder eller oversettelser), noe som fører til at siteringer til dem faller utenfor det korte tidsvinduet som brukes til beregning av impaktfaktor (10).

Felt-effektene gjør seg også i høy grad gjeldende for impaktfaktorene. De fleste tidsskriftene dekker bestemte fagområder, og får følgelig en impaktfaktor som er bestemt av fagområdets siteringsdynamikk: basale tidsskrifter har høyere impakt enn kliniske; tidsskrifter innen ekspansive fagområder har høyere impakt enn innen stasjonære områder; fagområdets siteringspraksis (referanseantall, referanse-aktualitet) gjenspeiles i tidsskriftimpakt osv. Selv om både siteringshyppighet og tidsskrift-impakt i gjennomsnitt er uavhengige av fagområdets størrelse, vil et stort fagområde gi rom for større stratifisering og derved for tidsskrifter med et høyere absolutt impakt enn innen et mindre fagområde. Eksempelvis har det ledende tidsskriftet for oversiktsartikler innen biokjemi (*Ann. Rev. Biochem.*) 5-6 ganger så høy impaktfaktor som det

tilsvarende tidsskriftet innen mikrobiologi (*Ann. Rev. Microbiol.*). Forskere innen mindre fagområder kan altså rett og slett mangle tilgang på høyimpakt-tidsskrifter, noe som vanskeliggjør en rettferdig bruk av impaktfaktorer (7).

Tabell 5. Problemer forbundet med bruk av tidsskriftimpakt-faktorer

1. Impaktfaktorer beregnes bare for et begrenset antall tidsskrifter
2. Ikke-engelskspråklige tidsskrifter diskrimineres
3. Impaktfaktoren favoriserer tidsskrifter med kortlivede artikler
4. Beregningsgrunnlaget for impaktfaktoren er feilaktig
5. Impaktfaktoren er avhengig av forskningsfeltet
6. Ikke alle fagområder har tidsskrifter med høyt impakt
7. Impaktfaktoren er ikke representativ for enkeltartiklene i tidsskriftet
8. Impaktfaktoren korrelerer dårlig med reell siteringshyppighet
9. En høy impaktfaktor gir ikke artiklene noen gratis siteringer
10. Tidsskriftvalg ved publisering er (var) ikke primært basert på impaktfaktor

Den største innvendingen mot bruk av tidsskriftimpakt-faktorer til evaluering er imidlertid at impaktfaktoren ikke er representativ for enkelt-artiklene i tidsskriftet. Ved måling av reell siteringshyppighet for enkelt-artiklene viser det seg at det er en stor og skjevfordelt (negativt eksponensiell) spredning i verdiene (3,8). De 15% mest siterte artiklene svarer for halvparten av siteringene, og den mest siterte halvparten av artiklene er, i gjennomsnitt, omtrent 10 ganger så høyt sitert som den minst siterte halvparten. Ved å tilskrive alle artikler samme verdi - dvs. tidsskrift-impaktfaktoren - tilsløres disse forskjellene: de minst siterte artiklene blir vurdert like høyt som de mest siterte. En oppnår altså det stikk motsatte av det som burde være hensikten med en evaluering.

Et stort, representativt utvalg av artikler fra et tidsskrift vil naturligvis vise en gjennomsnittlig siteringshyppighet som svarer til tidsskriftets impaktfaktor, og det samme vil ofte være tilfelle når større antall artikler evalueres samlet i forhold til sine respektive tidsskrifter. Ser en på enkeltforskere eller forskningsgrupper, er det imidlertid ofte svært dårlig korrelasjon mellom tidsskrift-impaktfaktor og artiklenes reelle siteringshyppighet (11). Impaktfaktorene er bare representative når den evaluerte forskningen er helt gjennomsnittlig, og på slike premisser er en evaluering egentlig helt overflødig.

Forskjellen mellom høyt siterte og mindre siterte forfattere i sin alminnelighet er langt på vei tidsskrift-uavhengig, og det er ikke noe som tyder på at en får noen gratis-siteringer ved å publisere i et høyt sitert tidsskrift (11). Det er heller ingen selvfølge at alle forskere streber etter å publisere i tidsskrifter med høyest mulig impakt: valget av tidsskrift er først og fremst bestemt av tidsskriftets faglige profil (12).

Bruk av tidsskriftimpakt-faktorer vil altså tendere til å tilsløre istedenfor å avsløre reelle forskjeller mellom forskere, dvs. det stikk motsatte av hva en ønsker å oppnå ved en evaluering.

Konklusjon

Som det vil framgå av ovenstående, er siteringsbaserte indikatorer dårlig egnet til evaluering av forskningskvalitet. Bruk av objektive, kvantitative indikatorer kan imidlertid overfladisk sett virke overbevisende, og spesielt bruken av tidsskriftimpakt-faktorer er tiltrekkende ved at den gjør det svært lett å foreta noe som kan likne på en evaluering. Impaktfaktorer er blitt benyttet til evaluering i en rekke europeiske land, og blir med jevne mellomrom forsøkt introdusert også i Norden, så det kan være grunn til å være på vakt (13).

Hva er så alternativet? Det finnes neppe noen vei utenom den direkte: en må evaluere *innholdet* i forskningen, ikke innpakningen eller andre former for vandring rundt grøten. Evalueringen må base-

res på publiserte forskningsresultater, som *leses* av eksperter med tilstrekkelig innsikt i forskningsfeltet til å kunne foreta en rettferdig evaluering, m.a.o. »peer review«. Mye kan imidlertid gjøres for å bedre kvaliteten på ekspertvurderingen, gjennom standardiserte kriterier, rasjonelle rutiner, klare målsettinger og en formålstilpasset evaluering. Det burde være mulig å kategorisere de forskjellige kvalitetskomponentene i forskningen (originalitet, soliditet, klarhet, viktighet, forskningsintern bruk-verdi, forskningsekstern relevans osv.), og karaktersette dem på grunnlag av en skala og et regelsett som det er oppnådd konsensus om i forskersamfunnet. Hvis ekspertene får anledning til å konsentrere seg om ren kvalitetsvurdering istedenfor tidsødende og ufruktbart komitéarbeid, vil kostnadene i tid og penger kunne bli akseptable i forhold til de store investeringene som ligger bak hver enkelt vitenskapelig publikasjon. En yrkesgruppe som stiller så høye kvalitetskrav i sitt arbeid, bør også kreve et minimum av kvalitet når dette arbeidet skal evalueres.

Referanser

1. MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *J. Am. Soc. Information Sci.* 1989;40:342-349.
2. Seglen PO. Evaluation of scientists by journal impact. In *Representations of Science and Technology*, Weingart P, Sehringer R, Winterhager M, eds., DSWO Press, Leiden, 1992; p. 240-252.
3. Seglen PO. The skewness of science. *J. Am. Soc. Information Sci.* 1992;43:628-638.
4. Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ. The application of bibliometric indicators: important field-and time-dependent factors to be considered. *Scientometrics* 1985;8:177-203.
5. Hargens LL, Felmlee DH. Structural determinants of stratification in science. *Am. Sociol. Rev.* 1984;49:685-697.
6. Folly G, Hajtman B, Nagy JI, Ruff I. Some methodological problems in ranking scientists by citation analysis. *Scientometrics*

- 1981;3:135-147.
7. Seglen PO. Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. *Tidsskr. Nor. Laegeforen.* 1989;31:3229-3234.
8. Seglen PO. How representative is the journal impact factor? *Research Evaluation* 1992;2:143-149.
9. Moed HF, Van Leeuwen TN. Impact factors can mislead. *Nature* 1996;381:186.
10. Sorokin NI. Impact factors of Russian journals. *Nature* 1996;380:578.
11. Seglen PO. Causal relationship between article citedness and journal impact. *J. Am. Soc. Information Sci.* 1994;45:1-11.
12. Gordon MD. How authors select journals: A test of the reward maximization model of submission behaviour. *Social Studies of Science* 1984;14:27-43.
13. Drettner B, Seglen PO, Sivertsen G. Inverkanstal som fördelningsinstrument: Ej accepterat av tidskrifter i Norden. *Läkartidningen* 1994;91:744-745.

Per O. Seglen, professor, dr. philos.

Norsk Institutt for forskning og høyere utdanning (NIFU), Hegdehaugsveien 31, 0352 Oslo og Institutt for Kreftforskning, Det Norske Radiumhospital, Montebello, 0310 Oslo.