

Kwantitatieve geletterdheid voor verantwoord onderzoeksbeleid

Mijnheer/mevrouw de rector magnificus, geacht faculteitsbestuur, zeer gewaardeerde toehoorders,

Het moet ergens begin juni 2009 zijn geweest, tien jaar geleden dus. Ik werkte nog geen twee weken bij het Centrum voor Wetenschap- en Technologie Studies, kortweg CWTS, en ik had aangegeven dat ik graag een korte presentatie zou willen geven voor mijn nieuwe collega's. De week ervoor had ik me verdiept in de berekening van de zogeheten kroonindicator van CWTS. Dit is een getal dat aangeeft hoe vaak de artikelen van een onderzoeker, een onderzoeksgroep of een universiteit gemiddeld genomen zijn geciteerd. Zo een getal kan behulpzaam zijn bij het evalueren van wetenschappelijk onderzoek. Het bijzondere aan de kroonindicator is dat deze indicator rekening houdt met verschillen tussen vakgebieden in de manier waarop onderzoekers citeren, waardoor een citatie in de wiskunde bijvoorbeeld anders wordt gewogen dan een citatie in de celbiologie. Het idee is dat op deze manier wetenschappelijke prestaties in verschillende gebieden met elkaar kunnen worden vergeleken.

Nadat ik me had verdiept in de berekening van de kroonindicator, kreeg ik een ongemakkelijk gevoel. Is dit wel echt de juiste manier om vergelijkingen tussen vakgebieden te maken, vroeg ik me af. Zouden de wiskundige formules achter de kroonindicator niet net op een iets andere manier moeten worden geformuleerd om tot een eerlijker vergelijking tussen vakgebieden te komen? De kroonindicator was in de jaren 90 door CWTS geïntroduceerd¹ en het centrum had er veel bekendheid mee gekregen. Ton van Raan, destijds directeur van CWTS, reisde de wereld over om de kroonindicator te promoten. Maar klopte de berekening van de indicator eigenlijk wel?

In mijn tweede week bij CWTS hield ik een korte presentatie waarin ik mijn twijfels over de berekening uitsprak en een alternatieve berekening suggereerde. Ik was best wel zenuwachtig. De kroonindicator betekende veel voor CWTS. Hoe zou er op mijn presentatie worden gereageerd? Hoeveel ruimte zou er zijn voor interne discussie over dit kroonjuweel van CWTS? Na mijn presentatie was het in eerste instantie stil. Niemand reageerde. Toen nam Ton van Raan het woord en gaf hij aan dat hij gevoelig was voor mijn argumenten en dat CWTS hier serieus naar moest gaan kijken. Dit was voor mij het begin van een periode die tot nu toe tien jaar duurt en waarin ik veel heb geleerd over het gebruik van kwantitatieve informatie ter ondersteuning van onderzoeksbeleid. In de rest van deze rede wil ik graag een aantal inzichten die ik in de loop van de jaren heb opgedaan met u delen.

Kwantitatief wetenschapsonderzoek

Mijn leerstoel richt zich op kwantitatief wetenschapsonderzoek. We leven in een wereld vol data, en het wetenschapssysteem is daar geen uitzondering op. We hebben data over de artikelen die wetenschappers publiceren. We weten ook hoe vaak er naar die artikelen gerefereerd wordt, niet alleen in andere artikelen maar bijvoorbeeld ook op sociale media, in blogs, in beleidsdocumenten, in medische richtlijnen en in patenten. We hebben ook steeds vaker toegang tot de volledige tekst van artikelen, wat een veel rijker beeld geeft dan wanneer we alleen de metadata zoals de titel en de

¹ De Bruin, R.E., Kint, A., Luwel, M., & Moed, H.F. (1993). A study of research evaluation and planning: The University of Ghent. *Research Evaluation*, 3(1), 25–41.

Moed, H.F., De Bruin, R.E., & Van Leeuwen, T.N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, 33(3), 381–422.

namen van de auteurs tot onze beschikking hebben. Verder zijn er statistieken beschikbaar over de uitgaven die landen aan wetenschappelijk onderzoek doen en over de subsidies die onderzoekers krijgen om hun werk te financieren. Ook hebben we steeds meer data over de manier waarop onderzoekers elkaars werk beoordelen, de zogeheten peer review.

Deels gaat kwantitatief wetenschapsonderzoek over het analyseren van deze grote hoeveelheden gegevens over het wetenschapssysteem. Dit wordt vaak scientometrie genoemd, ofwel het meten van de wetenschap. Scientometrie helpt ons om te begrijpen hoe het wetenschapssysteem functioneert. Het uiteindelijke doel is, wat mij betreft, om tot inzichten te komen die ons ondersteunen om het systeem zo goed mogelijk in te richten. Scientometrie kan ons bijvoorbeeld helpen bij het evalueren van onderzoek en bij het verdelen van schaarse middelen. Scientometrie biedt ook toegang tot informatie die behulpzaam kan zijn bij het ontwikkelen van onderzoeksbeleid, bijvoorbeeld om het wetenschappelijk carrièresysteem beter in te richten of om de wetenschap opener te maken.

Ik wil echter waarschuwen voor te hooggespannen verwachtingen van de scientometrie als een puur data-gedreven onderzoeksgebied. Voor veel vraagstukken is de beschikbaarheid van voldoende data absoluut essentieel, maar de waarde van data is beperkt als we niet ook de beschikking hebben over een theoretisch kader om de data te interpreteren. Nog te vaak ontbreekt het hieraan, waardoor data op simplistische manieren wordt geïnterpreteerd, wat soms zelfs tot misleidende conclusies leidt.

Kwantitatief wetenschapsonderzoek moet zich daarom naar mijn idee niet alleen bezighouden met het meten van de wetenschap, maar net zozeer met het ontwikkelen van theoretische kaders. Enerzijds kunnen deze theoretische kaders vanuit de scientometrie van een empirisch fundament worden voorzien en anderzijds kunnen ze bijdragen aan een meer genuanceerde duiding van scientometrische observaties. Kwantitatieve modellen en computersimulaties zijn veelbelovende methodes om op deze manier een brug te slaan tussen theorie en empirie. Ik kom hier straks nog even kort op terug.

Scientometrie en onderzoeksevaluatie

Ik neem u nu graag weer even mee terug naar de kroonindicator. Na mijn presentatie in juni 2009 ben ik samen met een aantal collega's bij CWTS aan de slag gegaan met een diepgaande vergelijking tussen de traditionele kroonindicator waar CWTS bekend mee was geworden en de alternatieve kroonindicator zoals ik die voor me zag. Om dit zorgvuldig te doen, namen we hier uitgebreid de tijd voor. Als we inderdaad zouden overstappen op de alternatieve berekening, zou de argumentatie voor die overstap kraakhelder moeten zijn.

Begin 2010 was ik hier met verschillende CWTS collega's nog altijd mee bezig. Op dat moment werden we verrast door een artikel van Loet Leydesdorff, een gewaardeerd collega uit Amsterdam.² In dit artikel werd precies dezelfde overstap op de alternatieve berekening van de kroonindicator bepleit. De argumentatie die Leydesdorff gaf was van een andere aard dan de argumentatie die we bij CWTS aan het ontwikkelen waren,³ maar de conclusie was hetzelfde: De traditionele kroonindicator moet worden vervangen door de alternatieve variant.

² Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics*, 4(3), 423–430.

³ Van Raan, A.F.J., Van Leeuwen, T.N., Visser, M.S., Van Eck, N.J., & Waltman, L. (2010). Rivals for the crown: Reply to Opthof and Leydesdorff. *Journal of Informetrics*, 4(3), 431–435.

Het is op dit moment niet zo interessant om in te gaan op de verschillende argumenten die in de kroonindicator discussie over tafel zijn gegaan. Wat ik interessanter vind, is hoe de discussie is ontstaan. CWTS verricht veel contractonderzoek. Als onderdeel daarvan, had CWTS de kroonindicator berekend voor onderzoekers van het AMC in Amsterdam. Sommige van deze onderzoekers waren het niet eens met de manier waarop de indicator werd berekend. Samen met Loet Leydesdorff hebben ze CWTS hier toen op aangesproken. De discussie over de kroonindicator was dus veel meer dan een academisch debat tussen een klein groepje specialisten. Het was een discussie met gevolgen voor de manier waarop onderzoekers en onderzoeksgroepen, niet alleen in Nederland maar ook elders in de wereld, werden geëvalueerd. Carrières van onderzoekers hingen misschien wel af van de uitkomst van deze discussie. Voor sommige onderzoekers van het AMC was de discussie zo belangrijk dat ze zich er persoonlijk mee zijn gaan bemoeien.

Betere indicatoren

De discussie over de kroonindicator en het besef van de mogelijke gevolgen van deze discussie voor carrières van onderzoekers waren voor mij en andere collega's bij CWTS reden om veel energie te steken in het verbeteren van de scientometrische indicatoren waar we bij CWTS mee werken. Dit heeft geresulteerd in iets waar ik best een beetje trots op ben, namelijk een systeem van indicatoren die op de meest logische en meest consistente manier zijn gedefinieerd.⁴

Deze indicatoren vormen bijvoorbeeld de basis voor de CWTS Leiden Ranking, een veel gebruikte vrij beschikbare databron met scientometrische indicatoren voor universiteiten.⁵ Naast indicatoren die aangeven hoe vaak de artikelen van een universiteit zijn geciteerd, bevat de Leiden Ranking een breed scala aan andere indicatoren. Er zijn indicatoren die aangeven hoe vaak universiteiten met elkaar en met het bedrijfsleven gezamenlijk publiceren. Recentelijk zijn ook indicatoren toegevoegd die laten zien in hoeverre universiteiten de artikelen die ze publiceren open beschikbaar maken en hoe de balans tussen mannen en vrouwen eruit ziet onder de auteurs van wetenschappelijke artikelen.⁶

De inzichten die we bij CWTS hebben ontwikkeld laten ook zien waarom bepaalde indicatoren beter niet kunnen worden gebruikt. De Hirsch index, of simpelweg de h-index, is hier een voorbeeld van. Zoals de meesten van u weten, wordt de h-index veel gebruikt om onderzoekers te evalueren. Tegelijkertijd is er ook veel kritiek op de h-index. Deze kritiek vind ik niet altijd terecht. Een deel van de kritiek is tamelijk oppervlakkig en laat niet zien wat er op een meer fundamenteel niveau verkeerd is aan de h-index. Bij CWTS hebben we kunnen laten zien dat de h-index inconsistent is in hoe de index onderzoekers evalueert.⁷ Wanneer onderzoeker Jansen volgens de h-index in eerste instantie bijvoorbeeld beter presteert dan onderzoeker De Vries, kan dit een tijdje later plotseling andersom zijn, ook al hebben de twee onderzoekers in de tussentijd precies dezelfde prestatie geleverd.

Populaire universitaire ranglijsten, die veel zichtbaarheid krijgen in de media, hebben vergelijkbare inconsistenties. Een fundamentele fout die deze ranglijsten maken is dat ze indicatoren die afhangen

⁴ Waltman, L., Van Eck, N.J., Van Leeuwen, T.N., Visser, M.S., & Van Raan, A.F.J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37–47.

⁵ Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E.C.M., Tijssen, R.J.W., Van Eck, N.J., ... Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419–2432.

⁶ Waltman, L., & Van Eck, N.J. (2019, May 15). CWTS Leiden Ranking 2019 provides indicators of open access publishing and gender diversity [Blog post]. Retrieved from <https://www.cwts.nl/blog?article=n-r2w2b4>

⁷ Waltman, L., & Van Eck, N.J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406–415.

van de omvang van een universiteit en indicatoren die daar juist onafhankelijk van zijn op één hoop gooien. Stel dat ik u vraag welk land in de wereld het rijkst is. Een veel gebruikte maat is het bruto binnenlands product. Wanneer we deze maat gebruiken, zien we dat de Verenigde Staten in absolute zin het rijkst zijn. Echter, wanneer we relatief per hoofd van de bevolking kijken, is Luxemburg het rijkst. Beide perspectieven, absoluut en relatief, hebben een zekere relevantie. Afhankelijk van wat je precies wilt weten, moet je ofwel het absolute ofwel het relatieve perspectief nemen. Het is echter totaal onlogisch om te proberen om de twee perspectieven op een of andere manier te middelen. Dit zou iets opleveren met een volstrekt onduidelijke interpretatie. In populaire universitaire ranglijsten is het echter heel normaal om op deze manier absolute en relatieve statistieken met elkaar te vermengen. Dit maakt het zo onduidelijk wat deze ranglijsten ons vertellen. Deels vertellen ze ons iets over de grootte van universiteiten. Universiteiten met veel geld en veel onderzoekers doen het doorgaans goed. Deels vertellen de ranglijsten ons ook iets over de prestaties van universiteiten relatief ten opzichte van hun grootte. Deze twee perspectieven, absoluut en relatief, worden echter op één hoop gegooid, waardoor niemand echt begrijpt wat populaire universitaire ranglijsten ons nou eigenlijk vertellen.

Neurochirurgie vs. neurologie

Ik heb nu een aantal kritische kanttekeningen geplaatst bij veel gebruikte scientometrische indicatoren, zoals universitaire ranglijsten en de h-index. De discussie over de berekening van de kroonindicator laat zien dat ook indicatoren van CWTS niet gevrijwaard zijn van kritiek. Om duidelijk te maken hoe lastig het is om goede scientometrische indicatoren te definiëren, wil ik hier nu nog een keer op terugkomen.

Nadat de berekening van de kroonindicator op basis van nieuwe inzichten was aangepast, kwam er op een bepaald moment toch weer een klacht binnen, dit keer nota bene vanuit onze eigen universiteit. CWTS had de kroonindicator berekend voor onderzoeksgroepen bij het LUMC, het Leids Universitair Medisch Centrum. We werden benaderd door Wilco Peul, neurochirurg bij het LUMC. Ook voor de onderzoeksgroep van Peul hadden we de kroonindicator berekend. De uitkomst was niet al te best. De score voor de groep van Peul was relatief laag. Volgens Peul was dit echter geen eerlijke weergave van de wetenschappelijke impact van het werk van zijn groep. Peul meende dat de lage score een consequentie was van een bepaalde methodologische keuze in de berekening van de kroonindicator. Hij kon dit in detail uitleggen.

In de kroonindicator worden citatiestatistieken binnen vakgebieden direct met elkaar vergeleken. Citatiestatistieken in verschillende vakgebieden, bijvoorbeeld in de wiskunde en in de celbiologie, worden niet direct vergeleken. Door alleen binnen vakgebieden directe vergelijkingen te maken, voorkom je dat je appels met peren vergelijkt. Dat is in ieder geval het idee. Er blijft echter één heel lastig punt over: Wat is precies een vakgebied? Hoe definieer je dit?

Precies op dit punt meende Peul dat de kroonindicator een methodologisch probleem heeft. Peul is neurochirurg. Hij wees ons er echter op dat neurochirurgie in de berekening van de kroonindicator niet als een zelfstandig vakgebied wordt gezien. In plaats daarvan is het onderdeel van een groter gebied, waar naast neurochirurgie ook neurologie onder valt. Peul vertelde dat neurochirurgie en neurologie heel verschillende vakgebieden zijn, die grotendeels los van elkaar opereren. Hij meende dat het niet eerlijk is om onderzoek in deze twee vakgebieden direct te vergelijken, vooral omdat neurologisch onderzoek in de regel veel meer geciteerd zou worden dan neurochirurgisch onderzoek.

Peul had stevige kritiek. De gedetailleerdheid van zijn kritiek bood ons bij CWTS de mogelijkheid om stap voor stap na te gaan of de kritiek terecht was. Hiervoor maakten we gebruik van speciale visualisatietechnieken, waar ik straks nog wat meer over zal zeggen. Met behulp van deze visualisatietechnieken konden we vaststellen dat de kritiek van Peul tot in detail klopte.⁸ Neurochirurgie en neurologie werden op één hoop gegooid, terwijl onze visualisaties aangaven dat het hier inderdaad om twee op zichzelf staande vakgebieden gaat. Door onze visualisaties op een speciale manier in te kleuren, konden we ook inzicht krijgen in citatiepatronen. Wederom werd het verhaal van Peul bevestigd. Neurologische artikelen worden gemiddeld genomen minstens twee keer zo veel geciteerd als neurochirurgische artikelen. Directe vergelijkingen tussen de twee vakgebieden zijn daarom problematisch, vooral als je bedenkt dat het voortbestaan van onderzoeksgroepen mede van dit soort vergelijkingen kan afhangen.

Betere algoritmes

De casus van Peul heeft bij CWTS bijgedragen aan allerlei technische innovaties. We denken vandaag de dag zorgvuldiger na over hoe we vakgebieden definiëren. Steeds vaker kiezen we een aanpak waarin vakgebieden middels een algoritmische bottom-up benadering worden gedefinieerd op basis van citatierelaties tussen wetenschappelijke artikelen.⁹ Artikelen die in termen van citatierelaties sterk aan elkaar zijn gekoppeld worden tot hetzelfde vakgebied gerekend, terwijl artikelen die nauwelijks citatierelaties hebben aan verschillende vakgebieden worden toegewezen. Een voordeel van deze aanpak is dat vakgebieden op een veel hoger detailniveau kunnen worden gedefinieerd en dat definities van vakgebieden mee-evolueren met ontwikkelingen in de wetenschap. Nieuwe opkomende vakgebieden worden automatisch geïdentificeerd.

Deze algoritmes hebben we de afgelopen jaren steeds verder doorontwikkeld. Recentelijk heb ik samen met mijn collega's Vincent Traag en Nees Jan van Eck het zogeheten Leiden algoritme uitgebracht.¹⁰ Dit is een sterk verbeterde variant van het Louvain algoritme, een populair algoritme dat wereldwijd in duizenden wetenschappelijk studies, in allerlei vakgebieden, is gebruikt.¹¹ Het Leiden algoritme lost bepaalde tekortkomingen van het Louvain algoritme op en daarnaast is het ook nog eens vele malen sneller. Het identificeren van vakgebieden in de hele wetenschap, op basis van tientallen miljoenen artikelen en honderden miljoenen citatierelaties, hoeft met het Leiden algoritme niet meer dan een half uur te duren.

Ook grote commerciële bedrijven hebben interesse in de algoritmes die we bij CWTS ontwikkelen. Elsevier gebruikt onze algoritmes bijvoorbeeld in research analytics software die aan veel onderzoeksinstellingen wordt verkocht.¹² De snelle adoptie van onze algoritmes in dit soort veelgebruikte producten laat het belang van ons werk zien en leidt hopelijk tot steeds hoogwaardiger scientometrische analyses.

⁸ Van Eck, N.J., Waltman, L., Van Raan, A.F.J., Klautz, R.J.M., & Peul, W.C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLoS ONE*, 8(4), e62395.

⁹ Waltman, L., & Van Eck, N.J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378–2392.

¹⁰ Traag, V.A., Waltman, L., & Van Eck, N.J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9, 5233.

¹¹ Blondel, V.D., Guillaume, J.L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10, P10008.

¹² <https://www.elsevier.com/solutions/scival/releases/topic-prominence-in-science>.

Contextuele scientometrie

Het algoritmische werk waar ik nu iets over heb verteld biedt een gedeeltelijke oplossing voor de problemen waar Wilco Peul met zijn kritiek op de kroonindicator de aandacht op vestigde. Je kunt de oplossing ook in een andere richting zoeken. Ik noemde al even kort de visualisatietechnieken die we bij CWTS hebben gebruikt om de kritiek van Peul te verifiëren. We maakten daarbij gebruik van de VOSviewer software voor scientometrische visualisatie.¹³ Deze software heb ik de afgelopen tien jaar samen met Nees Jan van Eck ontwikkeld. De software is vrij beschikbaar en wordt wereldwijd op grote schaal gebruikt. Er zijn meer dan 500 artikelen verschenen waarin gebruik is gemaakt van de software. Velen van u zullen bekend zijn met de visualisaties, of wetenschapskaarten, die met VOSviewer kunnen worden gemaakt. Voor wie hier niet bekend mee is, zou ik nu eigenlijk een voorbeeld van zo een visualisatie moeten laten zien. Ik heb een voorbeeld bij me, aan de binnenkant van de toga die ik draag. Ik denk dat dit niet helemaal het juiste moment is om de toga te openen, maar na afloop van deze oratie laat ik u een en ander graag zien.

Met VOSviewer visualisaties kon de kritiek van Wilco Peul in detail worden geverifieerd. Zoals ik zojuist heb uitgelegd, hebben we de inzichten die dit oplevert gebruikt om de kroonindicator te verbeteren. Je kunt jezelf echter een meer fundamentele vraag stellen. Als visualisaties in staat zijn om de tekortkomingen van de kroonindicator bloot te leggen, zouden ze dan niet ook een alternatief kunnen bieden voor deze indicator? Met andere woorden, moeten we er überhaupt wel naar streven om wetenschappelijke prestaties in eendimensionale numerieke indicatoren te vatten? Zou het niet veel beter zijn om een meer pluriform, multidimensionaal en gecontextualiseerd beeld te geven van wetenschappelijke prestaties?

Dit is een heel actuele discussie, bijvoorbeeld in de context van het erkennen en waarderen debat dat op dit moment in de Nederlandse onderzoekswereld wordt gevoerd.¹⁴ Als CWTS staan we pal voor zo een meer pluriforme benadering. Het invloedrijke Leiden Manifesto, waar ik een van de auteurs van ben, pleit hier bijvoorbeeld voor.¹⁵ Het bieden van een meer pluriforme benadering betekent in de eerste plaats dat onderzoek niet uitsluitend op basis van scientometrische parameters moet worden geëvalueerd. Daarnaast betekent het wat mij betreft dat scientometrie op een andere manier moet worden ingezet dan nu veelal gebeurt. Dit is vooral belangrijk wanneer scientometrie wordt gebruikt op het niveau van individuele onderzoekers of onderzoeksgroepen. Hier pleit ik voor een aanpak die ik contextuele scientometrie noem.

In de contextuele scientometrie is het altijd mogelijk om dieper in de scientometrische data te duiken. Bij ieder getalletje moet je bijvoorbeeld kunnen zien wat er achter het getalletje zit. Wat is de onderliggende data? Hoe is het getalletje tot stand gekomen op basis van die data? Visualisaties zoals die met de VOSviewer software kunnen worden gemaakt bieden een manier om dieper in de data te duiken. Aan de ene kant geven dit soort visualisaties diepere inzichten dan numerieke indicatoren, zoals bijvoorbeeld blijkt uit de casus van Wilco Peul. Aan de andere kant bieden dit soort visualisaties

¹³ Van Eck, N.J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

Van Eck, N.J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320).

¹⁴ <https://www.scienceguide.nl/2019/05/elke-wetenschapper-die-te-lang-aan-de-universiteit-wordt-gehouden-is-er-eeen-te-veel/>.

¹⁵ Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429–431.

nog altijd een vereenvoudigde weergave van de onderliggende data. Daarmee helpen ze gebruikers, voor wie de onderliggende data wellicht te onoverzichtelijk en te gedetailleerd is, om toch patronen te zien in de data.

Het idee van contextuele scientometrie wijkt op essentiële punten af van het gangbare perspectief op scientometrische indicatoren. In de contextuele scientometrie is er geen a priori onderscheid tussen valide en niet-valide indicatoren. Waar het om gaat is dat de gebruiker in staat wordt gesteld om een indicator tot op een diep niveau te doorgronden. Iedere gebruiker kan dan zelf vaststellen of een bepaalde indicator voor een bepaalde vraag wel of niet relevant is. Gebruikers kunnen ook zelf bepalen hoe een indicator het best kan worden geïnterpreteerd. Er worden in de scientometrie eindeloze discussies gevoerd over wat voor interpretaties wel en niet aan citatiestatistieken kunnen worden gegeven. Zeggen citatiestatistieken iets over de kwaliteit of excellentie van onderzoek, duiden ze de wetenschappelijke impact van onderzoek aan, of misschien het gebruik ervan? Vergelijkbare discussies spelen bij zogeheten altmetrics, zoals statistieken over het aantal keer dat wetenschappelijke artikelen op Twitter zijn genoemd. Is dit een indicator van de maatschappelijke impact van onderzoek, of is dat onzin?

In de contextuele scientometrie erkennen we dat het niet mogelijk en ook niet wenselijk is om tot eenduidige antwoorden op dit soort vragen te komen. Begrippen zoals kwaliteit, excellentie en impact zijn nogal ambigu en hebben voor iedereen een andere betekenis. Deze begrippen moeten daarom niet te veel door scientometrici worden ingevuld. Dit kan beter door de gebruiker van scientometrische analyses worden gedaan, op een manier die aansluit bij de doelen die de gebruiker heeft. In de contextuele scientometrie moeten scientometrici gebruikers op een slimme gelaagde manier, bijvoorbeeld met interactieve visualisaties, zo goed mogelijk toegang geven tot scientometrische data. Scientometrici moeten zich niet te veel met de duiding van de data bezighouden. Wanneer scientometrici erin slagen om de data op een intuïtieve en begrijpelijke manier aan gebruikers aan te bieden, kan de duiding aan de gebruiker zelf worden toevertrouwd.

Het idee van contextuele scientometrie, zoals ik dat zojuist heb geschetst, biedt ook een manier om de kloof te overbruggen tussen kwantitatieve scientometrie aan de ene kant en kwalitatief expertoordeel aan de andere kant. Naar mijn idee bestaat deze kloof niet echt. We hebben deze kloof met z'n allen zelf gecreëerd door op een hele nauwe manier tegen scientometrie aan te kijken. Hier moeten we zo snel mogelijk van afstappen.

Open citaties

Contextuele scientometrie brengt niet alleen een nieuwe manier van denken over het gebruik van scientometrie met zich mee. Het creëert ook een behoefte aan open scientometrische databronnen. Zolang scientometrie vooral bestaat uit statische rapportages van numerieke indicatoren, is openheid van de onderliggende data niet per se essentieel. In het geval van contextuele scientometrie moet het altijd mogelijk zijn om dieper in de data te duiken. Dit kan alleen wanneer er wordt gewerkt met databronnen die voldoende open zijn.

Veelgebruikte databronnen zoals Web of Science en Scopus bieden deze openheid op dit moment niet. De producenten van deze databronnen zijn huiverig om veel openheid te bieden, omdat dit wellicht hun commerciële belangen zou kunnen schaden. Helaas is het daarom lastig om op basis van Web of Science of Scopus data de filosofie van contextuele scientometrie volledig te verwezenlijken.

Gelukkig vinden er op dit moment belangrijke ontwikkelingen plaats richting open databronnen. Het meest in het oog springt het Initiative for Open Citations, een groep die lobbyt voor het open beschikbaar maken van referenties in wetenschappelijke artikelen.¹⁶ Dit initiatief wordt door CWTS ondersteund en gaat zich in de nabije toekomst naast referenties hopelijk ook op andere soorten metadata richten, bijvoorbeeld op abstracts en affiliaties. Het Initiative for Open Citations is al behoorlijk succesvol geweest. Binnen een jaar hebben de meeste grote wetenschappelijke uitgevers zich erbij aangesloten. Daarmee is nu ongeveer de helft van alle referenties open beschikbaar. Dat zijn ongeveer een half miljard referenties.

Helaas zijn er ook enkele grote uitgevers die het idee van open citaties nog niet ondersteunen. Het belangrijkste zijn de American Chemical Society, Elsevier en IEEE. Het lijkt erop dat deze uitgevers niet meedoen met open citaties omdat ze naast hun activiteiten als uitgever ook andere financiële belangen hebben die ze willen beschermen. Elsevier heeft met Scopus bijvoorbeeld aanzienlijke financiële belangen. Door niet mee te doen met open citaties, gebruiken organisaties hun rol als uitgever om hun belangen in de markt voor informatieproducten te beschermen. Organisaties lijken hier uitsluitend gericht te zijn om hun eigen succes, terwijl ze hun bredere maatschappelijke verantwoordelijkheid uit het oog verliezen.

Ik geloof dat we ons allemaal moeten inspannen om uitgevers aan hun maatschappelijke verantwoordelijkheid te herinneren. Zelf heb ik eerder dit jaar actie ondernomen door, met steun van mijn hele redactie, af te treden als hoofdredacteur van het Elsevier tijdschrift *Journal of Informetrics*.¹⁷ Het tijdschrift was in veel opzichten succesvol, ook mede dankzij de inspanningen van Elsevier, maar de weigering van Elsevier om referenties open beschikbaar te maken ging regelrecht in tegen de belangen van de scientometrische gemeenschap. Met de redactie van *Journal of Informetrics* hebben we een nieuw tijdschrift, *Quantitative Science Studies*, opgericht. Dit tijdschrift is, via de ISSI society, eigendom van de scientometrische gemeenschap. We kunnen daardoor zelf bepalen hoe we het tijdschrift organiseren, waarbij we bijvoorbeeld hebben gekozen voor fair open access publishing en voor ondersteuning van open citaties.

Scientometrie en onderzoeksbeleid

Ik heb het een en ander gezegd over het gebruik van scientometrie in onderzoeksevaluaties. Ik wil de discussie nu verbreden en ook andere aspecten van onderzoeksbeleid meenemen. In het bijzonder wil ik iets zeggen over de ambitie om onderzoeksbeleid van een wetenschappelijke onderbouwing te voorzien.

Ik ga u drie stellingen op het grensvlak van scientometrie en onderzoeksbeleid voorleggen. Sommigen van u zullen wellicht een duidelijke mening over deze stellingen hebben.

Stelling 1. Citaties zijn doorgaans heel scheef verdeeld over de artikelen in een tijdschrift en om die reden is het statistisch gezien onjuist om de journal impact factor te gebruiken om individuele artikelen en hun auteurs te beoordelen.

Stelling 2. De zwakke correlatie tussen de beoordeling van artikelen op basis van peer review en op basis van citatietellingen bewijst dat peer review niet kan worden vervangen door citatieanalyse.

¹⁶ <https://i4oc.org/>.

¹⁷ <https://www.cwts.nl/news?article=n-r2v294>.

Stelling 3. Analyses van publicatie- en citatiestatistieken bewijzen dat er in de wetenschap een gender bias ten nadele van vrouwen bestaat.

Deze drie stellingen worden door veel scientometrici geaccepteerd. De stellingen hebben bovendien een aanzienlijke invloed op onderzoeksbeleid. De drie stellingen zijn echter allemaal gebaseerd op onderzoek met serieuze conceptuele en methodologische tekortkomingen. Ik zal dit voor ieder van de stellingen kort toelichten.

Stelling 1. Er zijn gegronde redenen om bezorgd te zijn over de sterke invloed van de journal impact factor in onderzoeksevaluaties.¹⁸ Het is echter niet correct om te claimen dat dit gebruik van de impact factor statistisch gezien onjuist is.¹⁹ De impact factor is inderdaad geen goede voorspeller van het aantal citaties van een individueel artikel. Op grond hiervan kunnen we echter niet concluderen dat de impact factor geen behulpzame indicator kan zijn van bepaalde aspecten van de kwaliteit van een artikel. Wanneer we een dergelijke conclusie zouden trekken, zouden we citaties en kwaliteit ten onrechte op één lijn plaatsen.

Stelling 2. Op het niveau van individuele artikelen is er inderdaad een zwakke correlatie tussen peer review en citatietellingen.²⁰ Hier moet echter bij worden aangetekend dat de interne consistentie van peer review ook laag is. Bovendien sluiten zwakke correlaties op het niveau van individuele artikelen niet uit dat er op hogere niveaus van aggregatie sterkere correlaties bestaan. Om deze redenen zijn zwakke correlaties op het niveau van individuele artikelen op zichzelf geen reden om het gebruik van citatieanalyse op hogere aggregatieniveaus af te wijzen.²¹

Stelling 3. Er zijn studies met een gerandomiseerde opzet die bewijs leveren dat er in de wetenschap verontrustende gender biases ten nadele van vrouwen bestaan.²² Studies van publicatie- en citatiestatistieken worden ook vaak als bewijs aangevoerd, maar dit is onjuist. Het betreft hier observationele studies die alleen correlaties laten zien, geen causale verbanden. Vaak maken deze studies correcties voor versturende factoren, maar het is onmogelijk om voor alle verstoringen te corrigeren. Deze studies vinden inderdaad verschillen tussen mannen en vrouwen, maar ze tonen niet aan dat deze verschillen veroorzaakt worden door biases ten nadele van vrouwen.

Kwantitatieve geletterdheid

De tekortkomingen van een deel van het kwantitatieve wetenschapsonderzoek en de invloed die dit heeft op onderzoeksbeleid vind ik zorgwekkend. Deze tekortkomingen zijn zeker niet uniek voor het wetenschapsonderzoek. Ze spelen in allerlei domeinen waar beleidsrelevant onderzoek wordt gedaan. Om iets aan deze tekortkomingen te doen, wil ik voor meer kwantitatieve geletterdheid pleiten. Dit betekent niet per se dat we meer kwantitatieve analyses moeten gaan uitvoeren of dat

¹⁸ Wouters, P., Sugimoto, C.R., Larivière, V., McVeigh, M.E., Pulverer, B., De Rijcke, S., & Waltman, L. (2019). Rethinking impact factors: Better ways to judge a journal. *Nature*, 569, 621–623.

¹⁹ Waltman, L., & Traag, V.A. (2017). *Use of the journal impact factor for assessing individual articles need not be wrong*. arXiv:1703.02334.

²⁰ Wilsdon, J.R., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., ... & Tinkler, J. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. Retrieved from <https://doi.org/10.13140/RG.2.1.4929.1363>

²¹ Traag, V.A. & Waltman, L. (2019). Systematic analysis of agreement between metrics and peer review in the UK REF. *Palgrave Communications*, 5, 29.

²² Moss-Racusin, C.A., Dovidio, J.F., Brescoll, V.L., Graham, M.J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, 109(41), 16474–16479.

deze analyses geavanceerder moeten worden. Ik bedoel vooral dat we een beter begrip moeten krijgen van kwantitatieve analyses en van de conclusies die uit dit soort analyses kunnen worden getrokken.

Concreet stel ik een aanpak voor die op drie pijlers berust: Beter kwantitatief onderzoek, realistische verwachtingen van het onderzoek en hoogwaardig onderwijs. Ik zal deze pijlers nu een voor een bespreken.

Beter kwantitatief onderzoek

Ik begin met de eerste pijler: beter kwantitatief onderzoek. Helaas gaan er te vaak zaken op een redelijk basaal niveau mis. Een voorbeeld is de problematische focus op statistische significantie, waardoor de resultaten van onderzoek op een onnatuurlijke binaire manier, dus wel of niet significant, worden gepresenteerd en waardoor de praktische significantie van effecten die worden gevonden vaak totaal ondergesneeuwd raakt.²³ Praktische significantie moet veel meer aandacht krijgen en er moet niet op een onnatuurlijke binaire manier naar statistische onzekerheid worden gekeken.²⁴ Net zoals het naïef is om te denken dat scientometrische indicatoren als een objectieve maat kunnen worden gebruikt om goede en minder goede onderzoekers van elkaar te onderscheiden, is het evenzeer naïef om te denken dat statistische significantietoetsen op een objectieve manier kunnen aangeven welke effecten er wel en niet toe doen.

Een ander probleem is gelegen in het verwarren van correlatie en causaliteit. Gerandomiseerde experimenten worden vaak gezien als de geëigende methode om causaliteit vast te stellen. De moeilijkheid is echter dat het vaak praktisch onmogelijk is om zulke experimenten uit te voeren of dat experimenten alleen mogelijk zijn in een kunstmatige opzet die ver van de werkelijkheid afstaat. De meeste scientometrische studies hebben daarom een observationele opzet. De resultaten van deze studies worden vaak zonder goede onderbouwing in causale termen geïnterpreteerd. Dit is bijzonder problematisch. Er zijn systematische methodes om op basis van een observationele onderzoeksopzet tot causale interpretaties te komen,²⁵ maar helaas maken scientometrici hier nauwelijks gebruik van.

Een derde probleem heeft te maken met de manier waarop onderzoekers hun theoretische ideeën formaliseren. In veel gevallen gebeurt dit met simpele lineaire modellen. Dit soort modellen bieden vaak echter geen nauwkeurige weergave van de theoretische ideeën die onderzoekers hebben. Er is daarom behoefte aan meer op maat gemaakte modellen om theoretische ideeën op een goede manier te formaliseren. In andere gevallen vindt er helemaal geen formalisatie plaats. Dit hoeft ook niet altijd, maar het leidt wel tot het risico dat onderzoekers logische denkfouten maken die tot onjuiste conclusies leiden. Om dit te voorkomen, moeten onderzoekers hun theoretische ideeën vaker in formele modellen vertalen. Er zou bijvoorbeeld meer gebruik moeten worden gemaakt van agent-based modeling en andere simulatiebenaderingen.

²³ Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567, 305–307.

Wasserstein, R.L., Schirm, A.L., & Lazar, N.A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73(S1), 1–19.

²⁴ Cumming, G., & Calin-Jageman, R. (2016). *Introduction to the new statistics: Estimation, open science, and beyond*. Routledge.

²⁵ Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books.
Hernán, M.A., & Robins, J.M. (forthcoming). *Causal inference*. Chapman & Hall/CRC.

Realistische verwachtingen

Ik heb nu geschetst hoe kwantitatief wetenschapsonderzoek op een hoger niveau kan worden getild. Het zou echter naïef zijn om te denken dat beter onderzoek alle problemen oplost. Dit brengt me bij de tweede pijler: Realistische verwachtingen van het onderzoek.

We verwachten veel van het wetenschapsonderzoek. Sommigen van ons hebben misschien wel stiekem de ambitie om tot 100% evidence-based onderzoeksbeleid komen. Dit is echter onrealistisch. Het is al helemaal onrealistisch om te denken dat de vele complexe vragen waar we voor staan van eenduidige kwantitatieve antwoorden kunnen worden voorzien. Dit is precies het probleem van de drie eerder genoemde stellingen. In deze stellingen wordt op simpele eendimensionale kwantitatieve gronden een antwoord gegeven op een heel complexe vraag. Een gedegen antwoord op een dergelijke vraag moet echter een verscheidenheid aan relevante overwegingen in ogenschouw nemen. Sommige zaken kunnen worden gekwantificeerd, terwijl dat bij andere niet of nauwelijks mogelijk is. De diverse overwegingen moeten uiteindelijk allemaal worden meegewogen, waarbij er altijd ruimte blijft voor discussie over het belang dat aan de verschillende overwegingen moet worden toegekend. Ik geloof dat we af moeten van het idee dat dit soort complexe vragen met een simpel 'ja' of 'nee' kunnen worden beantwoord.

Onderzoeksbeleid moet uiteraard waar mogelijk gebruik maken van inzichten uit het wetenschapsonderzoek. Waar we gebruikers van scientometrische indicatoren soms verwijten dat ze te veel belang hechten aan de 'getalletjes', moeten we zelf echter niet een vergelijkbare fout maken door complexe vragen van simpele eendimensionale kwantitatieve antwoorden te voorzien. Verantwoord onderzoeksbeleid wordt zo goed mogelijk ondersteund door een combinatie van kwantitatief en kwalitatief wetenschapsonderzoek, maar komt ook deels tot stand op basis van meer anekdotische impressies en politieke afwegingen. Door realistisch te zijn over wat het wetenschapsonderzoek wel en niet te bieden heeft, maken we het wellicht ook makkelijker voor onszelf om de verleiding te weerstaan om te sterke conclusies aan ons onderzoek te verbinden.

Hoogwaardig onderwijs

De derde pijler, hoogwaardig onderwijs, is misschien wel de belangrijkste om een hoger niveau van kwantitatieve geletterdheid te bereiken. In allerlei domeinen, van wetenschap tot gezondheidszorg, criminaliteitsbestrijding en bedrijfsleven, worden kwantitatieve methodes gebruikt om beleid te ontwikkelen. Universiteiten en hogescholen bieden allerlei verschillende opleidingen aan waarin studenten een diepgaande kennis van kwantitatieve methodes opdoen. Met name de opkomst van multidisciplinaire opleidingen, zoals de data science opleidingen die diverse universiteiten en hogescholen hebben opgezet, is naar mijn idee een belangrijke ontwikkeling.

De toepassing van kwantitatieve data-gedreven methodes in de specifieke context van beleidsontwikkeling lijkt in deze opleidingen doorgaans echter nog niet de systematische aandacht te krijgen die het verdient. Ik geloof dat opleidingen zich hier veel specifiekker op moeten richten. Dit betekent dat onderwijs in kwantitatieve methodes gecombineerd moet worden met gedegen sociaalwetenschappelijk onderwijs waarin aandacht wordt besteed aan de beleidscontext waarin deze methodes worden gebruikt. Daarnaast kunnen inzichten uit de geesteswetenschappen worden ingebracht om tot een diepere reflectie te komen op wat kwantitatieve methodes ons wel en niet kunnen vertellen. Met dit soort opleidingen zouden studenten niet primair worden opgeleid tot experts in de technische aspecten van kwantitatieve methodes, maar vooral tot experts in het verantwoord gebruik van deze methodes in de context van beleidsontwikkeling. Hier bestaat een

grote maatschappelijke behoefte aan. Ik hoop dat we bij Universiteit Leiden initiatieven op dit gebied kunnen nemen. CWTS zou hier graag aan bijdragen.

De kroon

Ik begon mijn verhaal met de kroonindicator van CWTS. Over deze indicator is in de loop van de jaren veel gediscussieerd. De discussie ging over allerlei belangwekkende technische zaken. Ik denk echter dat er onvoldoende aandacht was voor de bredere context waarin de indicator wordt toegepast. Inmiddels is het belang van deze bredere context veel helderder. Het is volgens mij duidelijk geworden dat het zoeken naar een kroonindicator niet het meest productieve uitgangspunt is. De echte kroon waar we met z'n allen naar op zoek moeten is een gebalanceerd perspectief op hoe de scientometrie kan bijdragen aan constructieve onderzoeksevaluaties en verantwoord onderzoeksbeleid. Dit zie ik als de grote uitdaging voor de komende jaren. Ik kijk ernaar uit om hier samen met velen van u hard aan te werken.

Dankwoord

Dit brengt me bij de afsluiting van deze rede. Ik wil graag iedereen bedanken die op een of andere manier heeft bijgedragen aan mijn benoeming. Een aantal mensen wil ik persoonlijk noemen.

Dit zijn allereerst mijn leermeesters aan de Erasmus Universiteit Rotterdam, Uzay Kaymak, Rommert Dekker en Jan van den Berg: Ik heb veel van jullie geleerd. In het bijzonder waardeer ik de grote vrijheid die jullie me hebben gegeven in mijn eerste stappen als onderzoeker. Daar heb ik veel aan te danken.

Bij CWTS ben ik aangenomen door Ton van Raan, destijds directeur van het instituut. Ton, bedankt voor het vertrouwen dat je vanaf het begin in me hebt gesteld en voor de mogelijkheden die je me hebt gegeven om me met van alles en nog wat te bemoeien, zelfs met jouw geliefde kroonindicator.

Ik heb me verder kunnen ontwikkelen in de periode dat Paul Wouters directeur van CWTS was. Paul, bedankt voor alle inspiratie die je me hebt geboden, voor je sterke leiderschap en voor je bijdrage aan de totstandkoming van de leerstoel.

Op dit moment ben ik zelf onderdeel van de directie van CWTS. Graag bedank ik Sarah de Rijcke en Ed Noijons: Ik waardeer onze fijne samenwerking zeer.

Ik wil ook de leden van de Quantitative Science Studies onderzoeksgroep bedanken. Ik leer heel veel van en met jullie. In het bijzonder noem ik Rodrigo Costas, Vincent Traag en Nees Jan van Eck. Rodrigo, dankjewel voor het vele mooie en belangrijke werk dat je voor de onderzoeksgroep doet. Vincent, bedankt voor de stimulerende discussies en voor de vernieuwende ideeën die je in het onderzoek inbrengt. Nees, we zijn niet alleen collega's maar ook goede vrienden. Ik ben blij dat dat nog altijd prima samengaat. We werken veel samen. Jouw bijdrage aan de dingen die we samen realiseren is absoluut essentieel.

Ik eindig met m'n familie en schoonfamilie. Ik ben zeer bevoorrecht met de stabiele en gelukkige privésituatie die jullie me bieden. Dank jullie allemaal daarvoor. Eline, ik spreek het niet vaak genoeg uit, maar je bent een geweldige steun voor me. Wat ik heb bereikt, zowel binnen de wetenschap als daarbuiten, zou zonder jou absoluut niet mogelijk zijn geweest. Tim en Björn, sorry dat jullie zo lang naar jullie papa hebben moeten luisteren. Superknap dat jullie het hebben volgehouden!

Ik heb gezegd.