

Twijfel en ongeloof

METHODOLOGIE Veel medisch onderzoek dat wordt gepubliceerd klopt niet, volgens epidemioloog John Ioannidis. Hoe kan het wetenschappelijk bedrijf verbeterd worden?

Margriet van der Heijden



FOTO LEO VAN VELZEN

John Ioannidis: Blissfully doomed

John Ioannidis (1965) is geboren in New York en groeide op in Athene, waar hij ook medicijnen studeerde en promoveerde. Daarna werkte hij onder meer aan Harvard University en Tufts University in de Verenigde Staten. Tegenwoordig is Ioannidis hoogleraar aan de Stanford University in de VS, directeur van het Stanford Prevention Research Center en gashoogleraar aan het Imperial College in Londen. Dat hij de wetenschap inging was onvermijdelijk, zegt Ioannidis: “Het was onontkoombaar, ik kon er niet aan ontsnappen. En nog niet: de deur staat open, ik kan opstaan en weggaan en toch zit ik tot ver na middernacht achter mijn bureau. Uit passie voor een idee, of omdat ik nadenk over een nieuw onderzoeksprotocol – ik ben *blissfully doomed*”.

ONLANGS SCHREEF epidemioloog John Ioannidis (45) in het medisch vakblad *European Journal of Clinical Investigation* een recept voor succes. Tot een Nobelprijs leidt het niet, waarschuwde hij, maar het opent in de medische wetenschap vele deuren.

Dit waren de instructies: ‘Zorg dat je vaker co-auteur bent van wetenschappelijke artikelen – door resultaten stukje bij beetje op te dissen, door kwaliteitscontroles achterwege te laten en door onterecht co-auteurschap gewoon te accepteren. Negeer negatieve commentaren van reviewers. Blijf een artikel net zo lang insturen tot je een gat in het reviewsysteem treft. Citeer jezelf en verwelkom toevallige (onterechte) citaties.’

Ioannidis nam zo het publicatiesysteem van de medische wetenschapsbladen op de korrel. In dat systeem beoordelen onbezoldigde *peer reviewers*, collega-experts dus, de artikelen die onderzoekers insturen. Pas daarna worden de stukken in aangepaste vorm geplaatst – of afgewezen. Die kritische ‘zeef’ van de bladen moet borg staan voor de kwaliteit van de onderzoekers die erin publiceren. Telkens als hun artikel vervolgens geciteerd wordt, groeit de reputatie van die onderzoekers verder.

Solide, zou je denken.

Toch komt ‘*junk*’, rotzooi, in de bladen veelvuldig voor, schrijft Ioannidis. Hij wijst op de onderzoeker die in de top-300 van meest geciteerde wetenschappers terecht kwam, alleen door telkens zichzelf te citeren. En hij constateert dat sommige auteurs vanaf hun vijftigste last krijgen van ‘angstaanjagende schrijf-incontinentie’. Begin deze maand was Ioannidis de hoofdspreker op een symposium van de Radboud Universiteit Nijmegen, als ‘luis in de pels van de medische wetenschap’ maar ook als hoog gewaardeerd epidemioloog. Ioannidis is onder meer hoogleraar aan de prestigieuze Stanford University in de Verenigde Staten en gashoogleraar aan het befaamde Imperial College in Londen en hij heeft meer dan 400 publicaties op zijn naam.

Waaronder dus zeer kritische. Zo liet hij zien dat één op de zeven spraakmakende en toonaangevende resultaten uit de drie medische topbladen (*British Medical Journal*, *The Lancet* en *JAMA*) weersproken of afgezwakt wordt in vervolgonderzoek (*JAMA*, 13 juli 2005). In een famous essay somt hij vervolgens zes redenen op waarom zo veel onderzoek niet deugt (*PLoS*, augustus 2005). Titel: ‘*Why most published research findings are false*’.

Maar provoceren? “Neece”, zegt Ioannidis en er trekt een frons over zijn zachttaardige gezicht. “Ik geloof niet dat ik wil provoceren. Ik wil begrijpen hoe wetenschappers in verschillende deelgebieden van het medisch onderzoek te werk gaan.”

U deed uw onderzoek naar de topresultaten in 2005. Is het nu beter?
“Of onderzoek uit 2011 wel standhoudt, kunnen we pas rond 2020 nagaan. Maar als ik moest gokken zou

ik zeggen dat er weinig is veranderd. Het is nog steeds reuze aantrekkelijk om sterk afwijkende resultaten in de grote onderzoeksbladen te publiceren. Onvermijdelijk is een deel van die resultaten overdreven.”

Iedereen vergist zich wel eens, maar welk percentage onjuiste resultaten is nog acceptabel?
“Zolang een fout verder geen gevolgen heeft en er niemand gewond raakt door verkeerde ingrepen is het niet zo’n punt. In andere gevallen wil je geen enkele fout toestaan. Neem bijvoorbeeld een groot gerandomiseerd onderzoek naar het effect van een dure behandeling die op de grens balanceert tussen voordeel voor de patiënt en schade door bijwerkingen. Met zulke behandelingen zijn in de toekomst misschien miljardeneuro’s gemoeid. Patiënten kunnen nadeel ondervinden. Dan moet je de kosten en baten heel nauwkeurig afwegen.”

In 2008 dook u in het observationeel onderzoek naar verbanden tussen genen en ziektes (*American Journal of Epidemiology*): 95 procent van de resultaten viel door de mand. Is dat acceptabel?
“Nou, juist dit onderzoeksveld van de genetische epidemiologie is sinds 2005 [uit dat jaar waren de nageplonsde resultaten, red.] volledig veranderd. Destijds werkten kleine groepen in harde competitie, waren de resultaten gefragmenteerd en stelde men aan de statistiek soepele eisen. Tegenwoordig werken genonderzoekers in grote consortia samen en kloppen hun resultaten nagenoeg altijd – al weet ik niet of zo’n grootschalige samenwerking voor elk onderzoeksterrein ideaal is.

“Ik denk dat onderzoekers in elk veld zelf moeten vaststellen welke gevolgen fouten hebben en hoeveel fouten ze willen toestaan. Soms is het zo lastig om informatie te destilleren uit een complexe gegevensrij, dat je blij bent als één procent van de nieuwe vondsten niet meteen weerlegd wordt.” Hij knikt bijna verontschuldigend.

Laten we wat punten doornemen die volgens u het percentage onjuiste resultaten omlaag kunnen brengen. U verzet zich bijvoorbeeld tegen ‘salami slicing’, het stukje bij beetje publiceren van resultaten...
“Ja, dat hangt samen met die werkwijze die vooral een paar jaar geleden veel voorkwam, waarbij individuele onderzoekers elkaar zoveel mogelijk probeerden af te troeven met extravagante resultaten. Vaak namen ze niet eens de tijd om elkaars spectaculaire resultaten te repliceren. Ze joegen liever zelf op iets nieuws. Zo raakten de vakbladen vervuld met allerlei vondsten die geen stand hielden.

‘Onderzoekers bieden hun negatieve resultaten niet eens aan tijdschriften aan’

den of gewoon helemaal niet gerepliceerd waren.

“Volgens mij is het beter als een overkoepelende groep onderzoekers alle meetgegevens verzamelt en onderzoeksprotocollen vaststelt. De individuele onderzoekers mogen zich natuurlijk niet weerhouden voelen om ideeën te opperen en zelf met de data aan de slag te gaan. Maar het voordeel is dat hun resultaten binnen zo’n groep snel gerepliceerd kunnen worden – en al dan niet weerlegd.”

Gebeurt dat?
“Ja, neem die genetische epidemiologie: daarin kun je nu alleen nog resultaten publiceren als ze gerepliceerd zijn.”

Maar grote groepen zijn star.
“Daarvoor ben ik niet zo bang. Onderzoekers zijn per definitie eigenzinnig en vernuftig. Ze zullen binnen zo’n groep altijd manieren vinden om hun onafhankelijkheid te bewaren en eigen ideeën te opperen.”

Wat vaak niet lukt: negatieve resultaten publiceren. Resultaten die tonen dat er géén verband bestaat tussen een gen en een aandoening of een stof en een effect. En die belangrijk zijn omdat ze toevallige of overdreven vondsten die ten onrechte wel zo’n effect suggereren, kunnen temperen.
“Zeker, maar het is niet per se zo dat vakbladen zulke negatieve resultaten weigeren, zoals vaak wordt gezegd. Het grootste probleem is dat onderzoekers hun negatieve resultaten niet eens aanbieden. Misschien verliezen zij liever geen tijd met het schrijven van artikelen over verbanden die niet bestaan. Misschien voelen ze zich er ongemakkelijk bij...”

En dat is op te lossen door...
“...een andere manier van belonen. Nu belonen we het spectaculaire resultaat. Dat geeft veel citaties en leidt vaak tot een aanstelling. Het zou beter zijn om de onderzoekers te belonen die hun resultaten nauw omschrijven en van wie de resultaten repliceerbaar zijn.”

Nog een punt: onderzoekers met publicatiedrift nemen niet altijd de tijd om het commentaar van reviewers te verwerken.
“Sterker, het gebeurt dat een artikel na een paar maanden ongewijzigd weer bij hetzelfde tijdschrift binnenkomt. Kennelijk hopen de auteurs dan dat het door de mazen van het net glipt. Of ze bieden het ongewijzigd bij andere bladen aan. Kijk, er zijn ongeveer 100.000 medische vakbladen wereldwijd – al staat publiceren in 99.950 daarvan gelijk aan vergetelheid en zijn er zelfs bladen waarvan je niet durft te vertellen dat je artikel erin is verschenen.” Lacht.

Zo boet peer review aan belang in?
“We moeten het absoluut niet afschaffen. Het kan wel veel beter. Bijvoorbeeld: als een artikel nu in een vakblad verschijnt weten we impliciet dat het door reviewers is beoor-



ILLUSTRATIE RIK VAN SCHAGEN

deeld. Maar waren dat echte experts? Hadden ze conflicterende belangen? Hebben ze twee dagen aan het stuk besteed of twee minuten? We hebben geen idee. Het zou goed zijn als bekend was wat het reviewen behelsde. Of als de commentaren van de reviewers openbaar waren.

“Ideaal lijkt mij als bovendien de achterliggende informatie bij elk artikel tot in detail openbaar beschikbaar is: de ruwe data, de onderzoeksprotocollen. Dan kunnen collega’s een onderzoek zonodig helemaal nalen. Dat kost vaak maanden. Daar aan komen reviewers nooit toe.”

Sommige collega’s zeggen dat uw leidmotief voor striktheid vooral belangrijk is voor grote gerandomiseerde onderzoeken waarin patiënten tevens proefkonijnen zijn. Mogen onderzoekers die observeren en analyseren – zonder in te grijpen met nieuwe behandelingen of middelen – met meer vrijheid en speelruimte op vonden jagen?
“Het punt is: ook zulk observationeel onderzoek leidt geregeld tot allerlei aanbevelingen. Eet meer fruit. Verf je haar niet. Slaap korter. Slaap aan komen reviewers nooit toe.”

uit op medisch onderzoekers om advies te geven en te vertellen hoe het zit, maar vaak kleven nog grote onzekerheden aan hun resultaten. Als zij dan toch snel naar buiten treden en niet over die twijfels spreken, dan wordt een voorlopig en onzeker resultaat in de publieke perceptie ineens een voldongen feit. En wanneer zulke feiten te vaak later niet kloppen, lijken onderzoekers frauduleus.”

Dus: noem de onzekerheden?
“Zeker: wetenschap draait om twijfel en onzekerheid. Helaas worden die

meestal weggewuifd. In de Verenigde Staten helemaal – daar vertaalt men elk resultaat het liefst direct in concrete acties. Zelfs de directeur van de National Institutes of Health doet eraan mee en roept dat elke dollar die in medisch onderzoek geïnvesteerd wordt, twee dollar zal opleveren. Maar het is belachelijk om te denken dat medisch onderzoek garant kan staan voor dollarwinsten en klinkende resultaten. De waarheid is dat we miljoenen dollars zullen verspillen en maar een paar resultaten zullen behalen – en dat is het beste wat we kunnen doen.”

Is er een verband? Echt?

Medische wetenschappers werken met mensen. Dat maakt hun werk lastig. Terwijl chemici gemakkelijk over miljoenen moleculen beschikken, is de omvang van medische experimenten meestal beperkt tot tientallen of honderden proefpersonen. Zeker als je wilt onderzoeken of een middel geen bijwerkingen heeft, hou je de onderzoeksgroep juist beperkt. Nog iets: terwijl moleculen allemaal identiek zijn, is ieder mens anders. Iedereen heeft zijn eigen genenpakket; ook reageren vrouwen soms anders op medicijnen dan mannen, kinderen anders dan volwassenen, Aziaten anders dan blanke Europeanen en zo voorts.

Daarnaast staat elk mens aan unieke omgevingsinvloeden bloot. Het dieet van Japanners is anders dan dat van Amerikanen. Op het Franse plateau is de lucht schoner dan in Tokio. De een rookt, de ander drinkt. Om uit die ingewikkelde brij toch een verband te destilleren tussen een geneesmiddel en een effect nemen onderzoekers hun toevlucht tot statistiek. En dat doen zij ook in ‘observatieneel onderzoek’ waarin zij niet ingrijpen met geneesmiddelen of behandelingen, maar waarin zij groepen mensen vergelijken om het verband tussen een genvariant en een aandoening, of tussen een stof en een symptoom vast te stellen. In een veelgebruikte methode formuleren zij eerst een hypothese. In de illustratie hiernaast bijvoorbeeld: rode snoepjes geven ADHD. Dat willen ze dus onderzoeken. Maar dat doen ze door eerst het tegenovergestelde te formuleren: rode snoepjes geven géén ADHD. Dat heet de ‘nulhypothese’ – die zegt dus: er is géén verband. Daarna tellen de onderzoekers hoe vaak mensen die wel of geen rode snoepjes eten ADHD hebben. Een mogelijke uitkomst is bijvoorbeeld dat van 100 snoepjeseters er 30 ADHD hebben. De onderzoekers berekenen vervolgens de kans op deze uitkomst als er inderdaad géén verband zou zijn.

Is deze kans kleiner dan 1 op 20, dan verwerpen ze de nulhypothese. (In vaktermen: bij een p-waarde van 0,05, ofwel: 5 procent). Hun conclusie is in dit geval: hoogstwaarschijnlijk is er wél een verband en klopt de oorspronkelijke hypothese.

En ja, 1 op 20 *klinkt* als een kleine kans – en dus als een gerechtvaardigde grenswaarde. Maar als er jaarlijks honderden onderzoeken worden gedaan, dan zijn er flink wat die toch alleen maar toevallig een verband vinden.

Het probleem is dat onderzoekers vrijwel uitsluitend zulke ‘positieve’ resultaten publiceren. Dat geeft een vertekend beeld, omdat toevallige uitschieters dan ten onrechte niet gecorrigeerd worden door proeven die geen verband vinden. Met een beetje pech kan een foutief idee zo jaren rondzingen.

Moeten dus die statistische methoden op de schop? De befaamde epidemioloog John Ioannidis denkt van niet. Soms werken andere methoden beter, zoals de Bayesiaanse statistiek, zegt hij. Maar het zal mis blijven gaan, zolang voornamelijk positieve resultaten in de bladen verschijnen en zolang onderzoekers alleen opgewonden raken over positieve resultaten.