

WETENSCHAP EN WAARHEID ZIJN NIET ÉÉN

Griekse internist zet vraagtekens bij resultaten van onderzoek

De meeste gepubliceerde onderzoeksresultaten zijn niet waar, zo berekende de Griekse internist en hoogleraar epidemiologie John Ioannidis.

‘Praktiserend artsen moeten nadenken over de geloofwaardigheid van een behandeling. Een p-waarde alleen zegt mij niets.’

EVERT PRONK

Op veel websites over multiple sclerose staat dat vaders de neurologische aandoening twee keer vaker doorgeven dan moeders. Deze wetenschap is afkomstig van onderzoekers van de prestigieuze Mayo Clinic. Zij publiceerden hun bevindingen in juli van vorig jaar in *Neurology* (2006; 67: 305-10). De patiëntenverenigingen zullen hun websites moeten aanpassen. Er is niets van waar, blijkt uit een veel groter onderzoek dat Canadese wetenschappers komende september, opnieuw in *Neurology*, publiceren. Vaders en moeders geven MS in gelijke mate door. Het vakblad heeft de resultaten afgelopen woensdag 27 juni al op het internet gezet (www.neurology.org).

Wie het recente dispuut over het diabetesmiddel rosiglitazon en het al dan niet verhoogde risico op myocardinfarct heeft gevolgd, weet dat wetenschap en waarheid niet één zijn. Volgens de

vijver van de medische wetenschap. De plons en de golven zijn echter bijna alleen opgemerkt door artsen en wetenschappers die zich met methodologische vraagstukken bezighouden. Wellicht is de flinke dosis wiskunde die Ioannidis in het artikel (*PLoS medicine* augustus 2005, volume 2, Issue 8, e 124 blz. 0696-0701) stopte, daar debet aan. De boodschap zelf is immers schokkend genoeg.

KANS OP WAARHEID

In de berekeningen van Ioannidis spelen grofweg drie factoren een rol. Het gaat om de combinatie van het effect en de onderzoeksgrootte, met andere woorden de power van de studie, versturende factoren anders dan de kans, oftewel bias, en de voorafkans dat een gevonden relatie klopt. Hieruit berekent hij de positief voorspellende waarde, of met andere woorden, de kans dat een gevonden uitkomst overeenkomt met de waarheid.

In het artikel in *PLoS* laat Ioannidis zien dat voor de meeste studies geldt

spellende waarde van 0,85. Dat betekent dat in 85 procent van de gevallen de uitkomst later niet wordt weerlegd. Maar bij kleinere, minder transparant uitgevoerde studies, waarbij de mogelijkheid bestaat dat de eindpunten zijn aangepast aan de uitkomsten, daalt deze maat voor de waarheid tot zelfs 20 procent.

ZONDER WISKUNDE

Eerder deze maand was Ioannidis te gast in het AMC waar hij de Ruyschlezing hield. Een mooie gelegenheid om de hoogleraar van de universiteit van Ioannina in Griekenland en Tufts University School of Medicine in Boston te vragen of hij ook zonder wiskundige formules kan uitleggen waarom onderzoeksresultaten zo onbetrouwbaar zijn.

Allereerst wil Ioannidis kwijt dat hij niet de bedoeling had een provocatief stuk te schrijven. ‘Het is een uitvloeisel van mijn empirische werk waarin ik heb bestudeerd hoe vaak onderzoek wordt gerepliceerd, hoe resultaten van grote en kleine studies zich verhouden, en die van epidemiologisch onderzoek en RCT’s, et cetera. Het is een theoretisch concept waarin ik mijn bevindingen heb samengevoegd om dat als startpunt te nemen voor de vraag of onderzoeksuitkomsten geloofwaardig zijn en in welke mate ze geloofwaardig zijn.’

Vervolgens licht hij toe waarom de veelgebruikte p-waarde van studies niet zo veel zegt over de geloofwaardigheid van onderzoeksresultaten. ‘De waarde die aan de p-waarde wordt gehecht, is

Griekse internist en epidemioloog John P.A. Ioannidis zijn gepubliceerde onderzoeksresultaten zelfs vaker niet, dan wel waar. Met zijn essay *Why most published research findings are false* gooide hij anderhalf jaar geleden een steen in de

dat de kans dat de uitkomst overeenkomt met de waarheid kleiner is dan 0,5. Alleen een goed uitgevoerde, voldoende grote *randomised controlled trial* (RCT) en een meta-analyse van goed uitgevoerde RCT’s hebben een positieve voor-

Als je maar genoeg onderzoekt, krijg je vanzelf een keer een significant resultaat

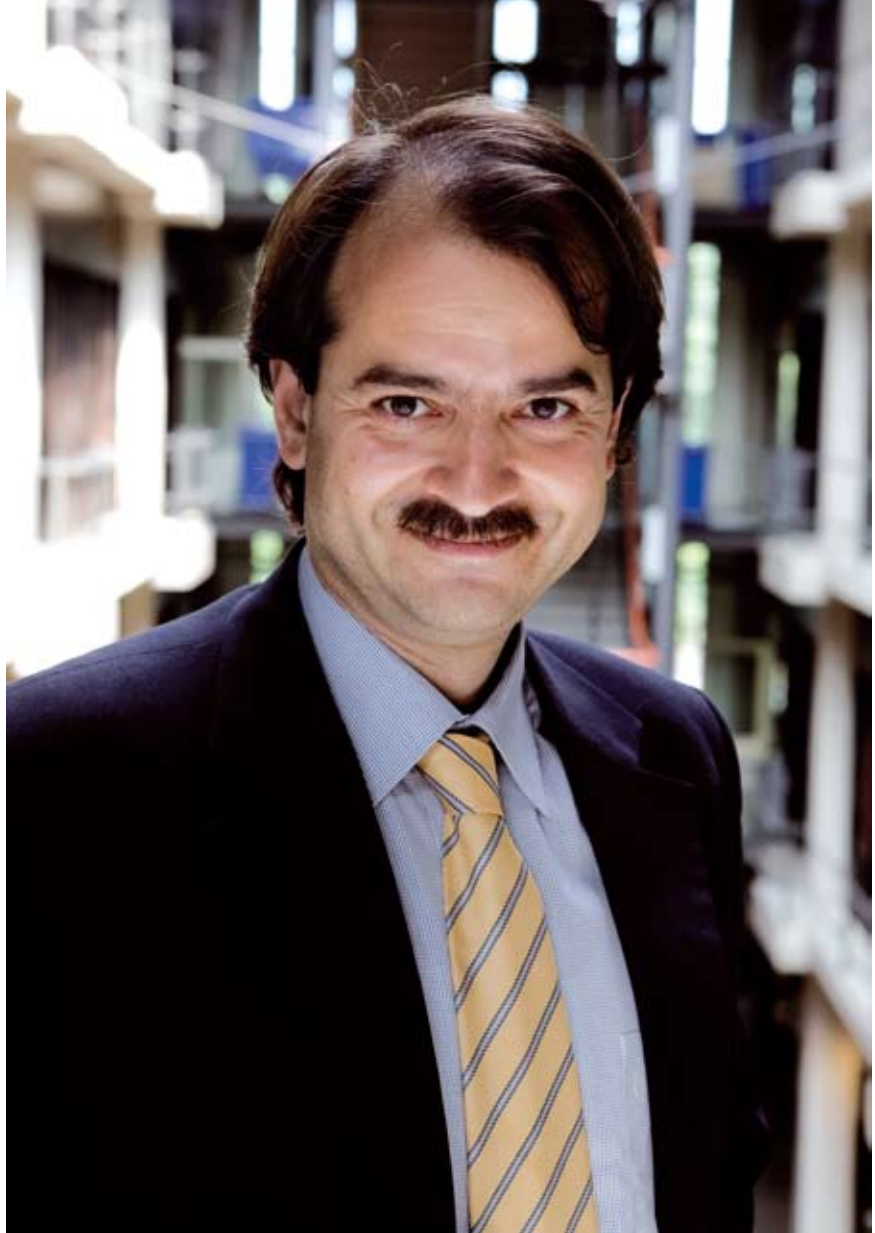
veel te groot. De p-waarde was prima toen er nog niet zo veel wetenschap werd bedreven als nu. Een p-waarde beneden de 0,05 zegt dat de kans dat de gevonden waarde op toeval berust kleiner is dan 5 procent. Maar er wordt tegenwoordig zo onnoemelijk veel onderzocht, dat er dus ook in absolute zin heel veel studies zijn waarvan de uitkomsten inderdaad toeval zijn.'

'Bovendien zijn de tijden van het leggen van grote verbanden, zoals tussen roken en longkanker voorbij', vervolgt Ioannidis. 'Er wordt bijna alleen nog maar op kleine effecten gejaagd, door veel te veel onderzoeksteams. Die moeten allemaal significante resultaten behalen, want die kun je publiceren. Dit geeft het gevaar van selectie op resultaten die toevallig significant zijn. Iemand publiceert dat het eten van wortels de bloeddruk verlaagt, maar wat we niet weten is hoeveel voedingsmiddelen er zijn getest en naar hoeveel eindpunten er is gekeken. Als je maar genoeg onderzoekt, krijg je vanzelf een keer een significant resultaat. Dit maakt het gevaarlijk.'

REPLICEREN

Veel onderzoeksgroepen is toch juist een voordeel? 'Dat hangt ervan af. Als ze samenwerken wel, maar dat gebeurt maar weinig. Er is veel competitie dus ze delen hun gegevens niet en maken deze ook niet openbaar, zodat onderzoekers de resultaten kunnen controleren. Het zou goed zijn als onderzoekers vaker studies proberen te repliceren. Nu gebeurt dat alleen als twee onderzoeksteams toevallig tegelijkertijd aan hetzelfde onderwerp werken. Er wordt geen prioriteit gegeven aan het repliceren van resultaten. Het is in het huidige waarderingssysteem niet aantrekkelijk om tweede te worden. Toch is het van het grootste belang, want zonder replicatie is het onduidelijk wat de waarde van een onderzoek is. Het leidt tot een vloed van veronderstelde resultaten.'

Als de berekeningen van Ioannidis correct zijn, is de kans om bij replicatie tot een andere uitkomst te komen groter dan de kans dat het onderzoek tot dezelfde conclusie leidt. Rationeel gezien is repliceren dus juist slim. 'Precies. Dit geldt met name voor onderzoeksgebieden waarvan te verwachten is dat de geloofwaardigheid laag is, zoals epidemiologisch onderzoek naar relaties tussen voedingsmiddelen en klassieke geneesmiddelstudies waarbij duizenden



John Ioannidis: 'De geloofwaardigheid is een compositie van de hoeveelheid evidence, de consistentie en bescherming tegen bias.'

middelen worden getest. Bij dergelijk onderzoek zijn de *pre-study odds*, oftewel de voorafkansen, laag.'

De kans dat een in wetenschappelijk onderzoek gevonden relatie echt bestaat, is groter als een studie is gebaseerd op kennis over de biologie van een aandoening en de relatie van een bepaalde stof daarin. Toch zegt ook dit volgens Ioannidis niet alles. 'Er zijn meer dan duizend publicaties met positieve resultaten in een dierproefmodel voor een beroerte. Maar hoeveel behandelingen zijn er nu? Misschien trombolysen, maar ook dat is nog twijfelachtig. De resultaten in de dierproeven bleken na analyse geen enkele voorspellende waarde te hebben voor resultaten bij humane studies. Met andere woorden, het dierexperimenteel onderzoek gaf geen richting aan het onderzoek. Voor we dierexperimenteel onderzoek afschaffen, moeten we goed uitzoeken wanneer het zinnig is.

Het moet niet gedaan worden, omdat het nu eenmaal gedaan moet worden.'

BELANG VAN BIAS

Behalve kans is volgens Ioannidis ook bias een belangrijke factor. 'Er is inmiddels heel wat bekend over de verschillende vormen van bias. Of bias een rol speelt, is achteraf vaak oncontroleerbaar omdat de publicatie maar een deel van het onderzoekstraject laat zien. Dat is belangrijk, want dan kun je proberen er rekening mee te houden. Maar er is ook bias waar je niet van tevoren aan kunt denken. Neem het onderzoek naar middelen tegen aids. Nu zijn er goed werkende middelen tegen hiv, maar aan het begin van de zoektocht naar een effectieve behandeling waren er veel positieve studies met immuunmodulatoren en immunoglobuline. Alle data waren destijds positief terwijl we nu weten dat die middelen niet werken. Daar moet >>

<< bias een rol hebben gespeeld. Men wilde graag een remedie vinden.'

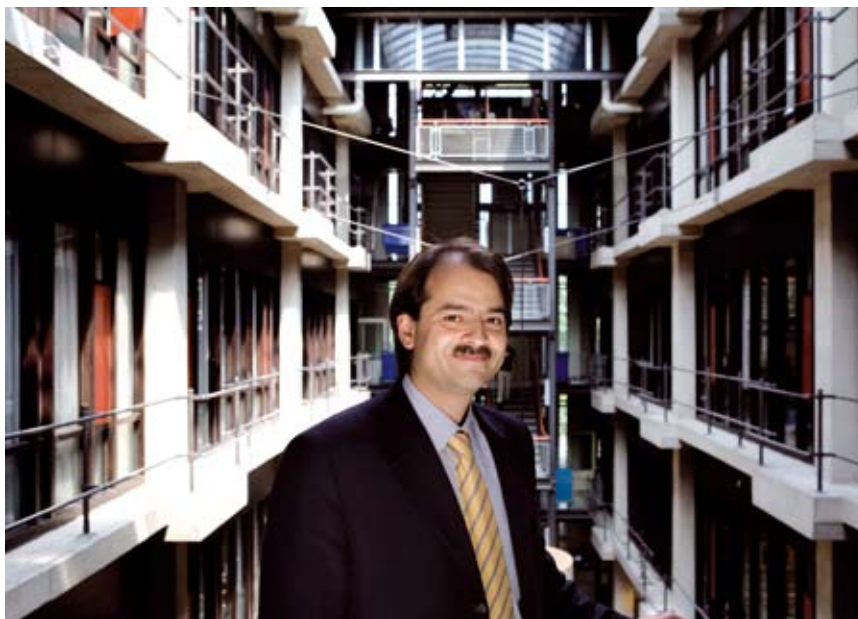
Het belang van bias baseert Ioannidis niet op louter gevoel. Hij heeft er een groot onderzoek naar gedaan door een vergelijking van de Europese en Amerikaanse met de Chinese literatuur over genetische associaties. 'Het Chinese equivalent van PubMed bevat meer dan 9000 wetenschappelijke tijdschriften. Publicaties in deze tijdschriften gaven altijd significante effecten die bovendien groter waren, terwijl de studies gemiddeld kleiner zijn. In de Europese en Amerikaanse literatuur zijn de oddsratio's gemiddeld 1,3 en in de Chinese literatuur 3. Van een deel van de studies is later in goed gecontroleerde, veel groter uitgevoerde onderzoeken aangetoond dat er geen relaties bestonden. Dus een oddsratio van 1. In feite komt het erop neer dat voor het onderzoek naar genetische associaties de maten voor bias in China en het westen 3 respectievelijk 1,3 zijn.'

GELOOFWAARDIGHEID

En wie zegt nu dat de berekeningen van Ioannidis over de geloofwaardigheid van wetenschappelijke publicaties correct zijn? Lachend: 'Er moet meer empirisch werk komen over publicaties. Over bias is al behoorlijk wat bekend, maar het is zeker nog niet compleet. Het is interessant om te onderzoeken of de bias omlaag gaat door stappen als *trial registration*.' De eis van wetenschappelijke tijdschriften om onderzoek vooraf te registreren is een goede ontwikkeling die volgens Ioannidis de geloofwaardigheid ten goede komt. 'Daarmee weet je tenminste van het bestaan van studies, ook als de resultaten uiteindelijk niet zijn gepubliceerd. The Lancet vraagt ook het protocol in te sturen, zodat je kunt zien of er selectieve analyse is gedaan, of selectieve rapportage is toegepast.'

'Maar vergis je niet. Slechts een klein deel van de studies wordt momenteel geregistreerd. Van het merendeel weten we niet van het bestaan. Het zijn toch ook vaak onderzoeken uit gespecialiseerde bladen, die geen trialregistratie eisen, die de praktijk in het ziekenhuis bepalen. Het zou bovendien goed zijn als ook andersoortige studies dan trials worden geregistreerd.'

Er zijn nog meer goede ontwikkelingen die volgens Ioannidis de geloofwaardigheid van gepubliceerde resultaten vergroten. 'Je ziet steeds vaker dat men studieresultaten direct verwerkt



BEELD: DE BEELDREDAKTIE, MARCO ORHUIZEN

'Een goed geïnformeerde patiënt is minder verrast als iets niet het gewenste effect heeft.'

in een meta-analyse, zodat zichtbaar is hoe de data zich verhouden tot de bestaande kennis. Ook meer data publiekelijk maken, helpt veel. Nature Genetics vraagt bij microarrays tegenwoordig de gehele dataset op, zodat anderen kunnen zien wat er precies is gedaan. Ook zijn er initiatieven om data te delen in netwerken van onderzoekers. Aan de publicatie in Nature waarin onlangs de ontdekking van drie nieuwe borstkankergenen werd beschreven, werkten 25 onderzoeksgroepen die de resultaten van elkaar hebben gerepliceerd.'

FORMULE

Een aanscherping van de p-waarde acht Ioannidis niet zinvol. 'Statistiek is belangrijk, maar je moet weten wat de statistiek betekent. Je kunt de grenzen opschuiven, maar het blijven grenzen en daar ben ik geen voorstander van. Het leidt tot automatiseren. Ze zetten mensen niet aan het denken over de waarde van de uitkomsten en wat er is gebeurd. Een p-waarde alleen zegt mij niets. In een meta-analyse over de relatie tussen vitamine-D-receptorpolymorfismen en het negatieve effect daarvan op de botdichtheid was de p-waarde 10-16. Uit een goed opgezette transparante studie groter dan de meta-analyse volgt een oddsratio van 1. Er is helemaal geen relatie tussen deze polymorfismen en de botdichtheid. Er moet sprake zijn geweest van selectieve analyse, selectieve rapportage, selectieve presentatie, publicatiebias of alles bij elkaar.'

'Het gaat om de geloofwaardigheid in plaats van de p-waarde', benadrukt

Ioannidis. De geloofwaardigheid is een compositie van de hoeveelheid evidence, de consistentie en bescherming tegen bias. Dit is in een formule te stoppen, want dat heb ik gedaan, maar dat helpt mensen niet over de geloofwaardigheid na te denken.'

Hoe moet een praktiserend arts dan omgaan met het gegeven dat er nogal wat twijfel is over de waarde van uitkomsten van klinische studies? 'Het vergt wat oefening, maar artsen moeten nadenken over de geloofwaardigheid van een behandeling. Wat zijn de onzekerheden en hoe breng ik dat over aan de patiënt? Dat voorkomt veel leed. Een goed geïnformeerde patiënt is minder verrast als iets niet het gewenste effect heeft.'

'En denk nu niet dat ik meen dat studies met lage geloofwaardigheid geen waarde hebben. Als er geen enkele andere behandeling is, en je hebt alleen een kleine studie die enig gunstig effect van een middel laat zien, dan pas je het toch toe. Ook als de kans dat het echt werkt maar 20 procent is. Maar dat moet je de patiënt wel uitleggen.' ■

Het artikel 'Why most published research findings are false' van John Ioannidis en meer MC-artikelen over de waarheid in de wetenschap vindt u via www.medischcontact.nl/dezeweek.

