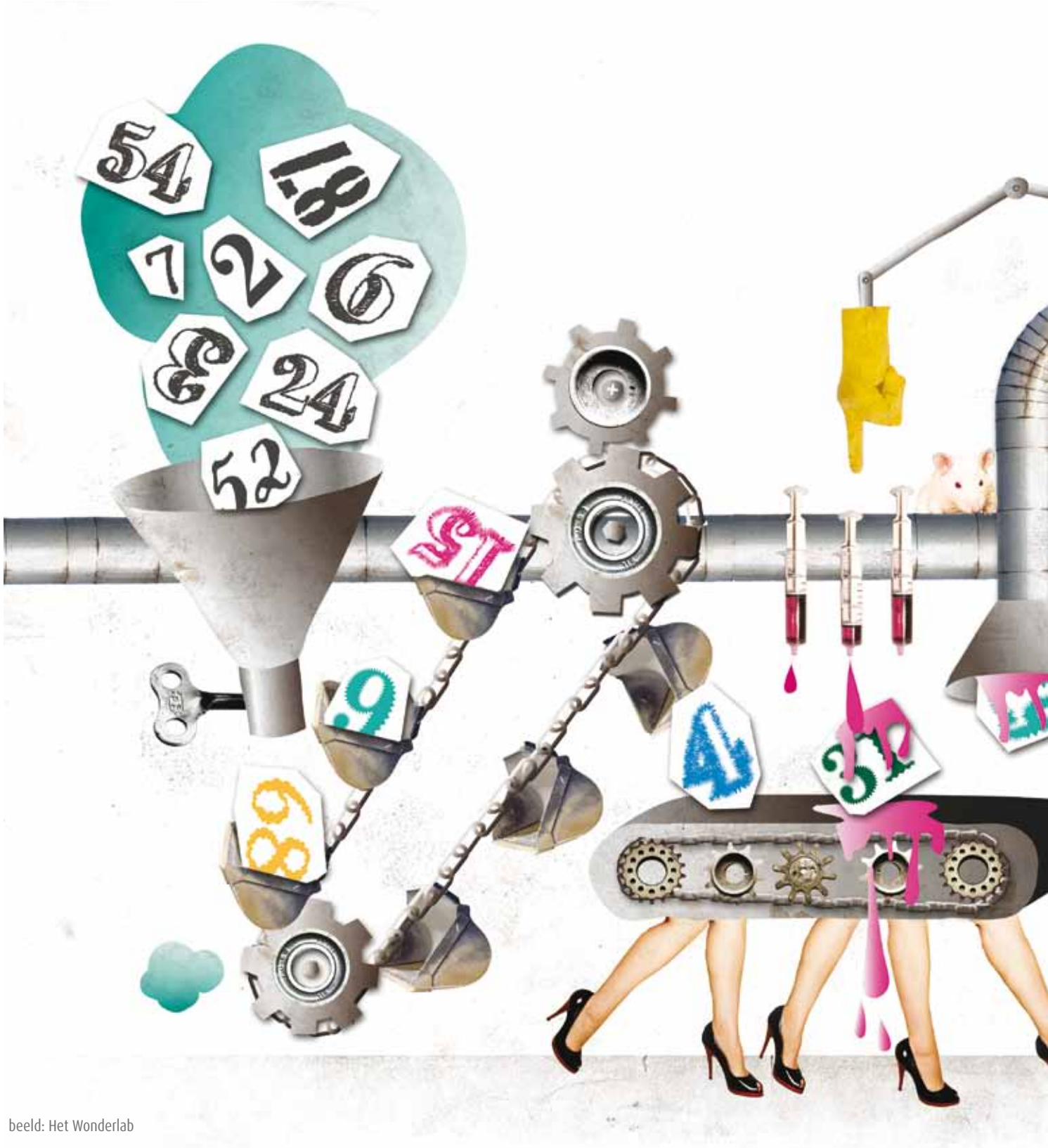


Een peer-reviewed



beeld: Het Wonderlab

mijnenveld

Navigatiehulp voor de gebruiker van wetenschappelijke literatuur

Dat alles wat in medisch-wetenschappelijke tijdschriften staat ook waar is, geloof je vast niet meer. Maar kun je de tekortkomingen en het bedrog ook meteen herkennen? De belangrijkste instinkers op een rij.

Hans van Maanen

Het blijft, ook voor een geharde wetenschapsjournalist, nog steeds verbazingwekkend hoe vaak auteurs van medisch-wetenschappelijke artikelen proberen hun lezers op het verkeerde been te zetten. Data in de *abstract* kloppen, zelfs in gerenommeerde peer-reviewed tijdschriften, vaak niet met wat er in het artikel staat, beweringen over significantie kunnen daar hoe dan ook niet geloofd worden, en zelfs uit de in het artikel gegeven cijfers is het moeilijk of zelfs onmogelijk het werkelijke risico of profijt van een behandeling op te maken.^{1,3} De medisch-wetenschappelijke literatuur is, kortom, een mijnenveld. Zonder goede navigatiehulp is er, zeker voor beginners, vrijwel geen doorkomen aan.

Kranten

Veel onderzoek is er niet naar gedaan, maar het is wel vrij duidelijk dat persberichten van academisch-medische instellingen veel te wensen overlaten.⁴ Journalisten maken veel gebruik van deze persberichten; als krantenstukken niet letterlijk daarvan zijn overgeschreven, zijn ze meestal slordig overgeschreven. Cijfers worden weggelaten, nu eens met een beroep op het onverstand van de journalist, dan weer met een beroep op het onverstand van de lezer. De goeden niet te na gesproken, maar het is werkelijk slechts zelden dat uit een krantenbericht de exacte toedracht en reikwijdte van een medische vinding gedistilleerd kan worden.

Krantenberichten hebben vaak de eenvoudige vorm 'x verhoogt de kans op y', met soms nog als toevoeging 'met z procent'. Soms, niet altijd. NRC Handelsblad liet het bijvoorbeeld op 17 november 2009 bij: 'Meer ongelukken bij mooi weer', zonder verdere percentages. Of intensivisten maatregelen moesten nemen bij zomerse dagen bleef



daardoor onduidelijk. De Volkskrant meldde op 17 april 2010 weliswaar dat ouderen die 'goed eten' 38 procent minder kans op alzheimer hebben, maar verzuimde erbij te zetten hoeveel kans ouderen die niet goed eten dan hebben.

Ook in de wetenschappelijke literatuur zijn die cijfers soms goed verstoppt. In het artikel waarop NRC Handelsblad zich baseerde, staat in de abstract wel dat mooi weer leidde tot 10,1 procent (95%-betrouwbaarheidsinterval 9,3 tot 11,4 procent) meer trauma's dan normaal, maar pas diep in het stuk wordt duidelijk dat het daggemiddelde op 26,9 zat met een standaardafwijking van 8,2 ongevallen. Het komt dus neer op amper 3 ongevallen extra bij mooi weer.⁵

Ook de auteurs van het oorspronkelijke artikel over het anti-alzheimerdieet zijn niet scheutig met cijfers. Ze verzuimen voor te rekenen dat het risico in vier jaar afneemt van 16 procent in het ongezondste tertiel naar 10 procent in het gezondste tertiel.⁶ Als 100 bejaarden hun voedingsgewoonten volledig omgooien, blijft 6 van hen alzheimer bespaard (al zullen 10 van de 100 desondanks in die vier jaar alzheimer krijgen). Het is kortom geen wonder dat uit een krantenbericht zelden de exacte toedracht en reikwijdte van een medische vinding kan worden gedestilleerd.

Aspirine

Onder de kop: 'Minder kanker door dagelijks aspirientje', begon de Volkskrant op 8 december 2010: 'Een aspirientje per dag verlaagt de kans op een aantal vormen van kanker met ruim 20 procent.' We herkennen inmiddels de x-y-z-vorm, en in de rest van het artikel wordt dat niet goedgemaakt (en het gaat om de kans op overlijden aan kanker). De auteurs van het wetenschappelijke artikel komen inderdaad uit op een relatief risico van 0,77.⁷ Want nadat 25.570 mensen waren geteld, overleden in de groep die

minstens vier jaar dagelijks aspirine hadden geslikt 327 mensen, in de andere groep 20 meer. Anders gezegd, de kans aan kanker te overlijden daalde met 0,7 procent: als 147 mensen minstens vier jaar trouw slikken, wordt na twintig jaar 1 sterfgeval voorkomen; 146 mensen slikken voor niets en 1 van hen zal in die tijd door de aspirine een ernstige bloeding krijgen. Voor de totale sterftekans na twintig jaar blijkt het allemaal weinig uit te maken ($HR = 0,96$; 95%-betrouwbaarheidsinterval 0,90 tot 1,02).

Ook de wetenschapsredactie van NRC Handelsblad was vol van het onderzoek, en publiceerde er zowel op 7 december over ('Aspirientje per dag helpt tegen kanker') als op 9 december ('Aspirine komt op tegen kanker'). In het tweede artikel komen deskundigen aan het woord die wat minder enthousiast zijn over aspirientjes, maar heldere berekeningen waarmee lezers zelf hun afwegingen kunnen maken, ontbreken nog steeds.

Genetica

Bijzondere voorzichtigheid is altijd geboden bij allerlei genen die voor allerlei aandoeningen en eigenschappen verantwoordelijk worden gehouden. Door het publiek wordt genetica veelal als ja-nee-kwesties beschouwd – met het gen: impulsief, overspelig enzovoort; zonder het gen: bedachtzaam, monogaam enzovoort – maar het zijn uiteraard slechts waarschijnlijkheden. Ook al staat er 'Een gen voor impulsiviteit' boven een beschouwing in Nature, het gaat om 17 van 228 ernstig impulsieve, gewelddadige Finnen die genvariant Q20* dragen tegen 7 van 295 controles.⁸ Voor gedragsgenetici een enorm verschil, maar toch: 92,5 procent van de impulsieve schurken zijn geen drager van het vermeende impulsiviteitsgen, 2,4 procent van de brave burgers wel.

Het is maar hoe je het inkleedt – soms klinken procenten beter, soms absolute aantallen. Voorstanders van het bevolkingsonderzoek naar borstkanker zeggen bijvoorbeeld: de sterfte door screening daalt met 26 tot 30 procent, terwijl tegenstanders zeggen: er moeten 2000 vrouwen 10 jaar meedoen om het leven van 1 vrouw te verlengen.^{9,10} En een sterftekans van 2 procent klinkt nu eenmaal altijd anders dan een overlevingskans van 98 procent. Ook artsen zijn hiervoor niet ongevoelig.¹¹

Significantie

'Significantie' is het toverwoord in de wetenschap. Roep 'p<0,05' en het lab staat op zijn kop. Een publicabel resultaat!

Maar om te beginnen is het helemaal niet moeilijk om significante resultaten te bereiken: je neemt gewoon een heel grote steekproef. Daarmee worden zelfs de kleinste verschillen significant. Het onderzoek naar de traumagevallen in Groningen hierboven omvatte 354.150 gevallen in 36 jaar – voldoende om verschillen van 1 procent significant te maken.⁵ Nieuwe maan bracht 2 procent (95%-betrouwbaarheidsinterval 1,01 tot 1,03) meer ongevallen dan volle maan, en ook de windrichting had een 'kleine doch significante invloed'. Dat stond dan weer niet in de



samenvatting. Andere factoren – luchtdruk, wolkendek, zicht, neerslag en nog zo wat – hadden trouwens geen invloed.

Hier herkennen we de volgende zonde: de ‘visexpeditie’. Wie twintig onzinvariabelen gaat meten, vindt er puur door toeval altijd wel één die significant is op 5%-niveau. Zo bleek ooit dat de kleur noch de maat van de beha het risico op borstkanker verhoogt, maar dat het aantal uren dat de beha wordt gedragen wél invloed heeft.¹²

De term ‘significant’ heeft alleen betekenis bij echte experimenten waarin twee hypothesen met elkaar worden vergeleken. Bijvoorbeeld de hypothese dat er geen verschil is in het aantal ongevallen bij mist tegen de hypothese dat er wel verschil is. Als de onderzoekers dan gaan tellen en er blijkt meer verschil dan door het toeval kan worden verklaard, is die hypothese onhoudbaar – waarmee de andere niet automatisch de juiste is. Andersom is het goed om van tevoren aan te geven wat nog een ‘interessant’ of ‘klinisch relevant’ verschil is, zodat de steekproef niet te groot wordt en onbeduidende verschillen van bijvoorbeeld 1 mmHg niet vermeldenswaard worden gevonden. Hoewel hoofdredacties van medische tijdschriften al in 1988 besloten kale significantietesten in de ban te doen ten gunste van betrouwbaarheidsintervallen, blijkt zo’n twintig jaar later dat nog meer dan 20 procent van de artikelen niet aan deze voorwaarde voldoet. Bovendien wordt in 80 procent van de gevallen statistische significantie gelijkgesteld aan klinische relevantie.¹³

Publicatiebias

Een van de grootste problemen die op het ogenblik vooral het medisch onderzoek bedreigen, is de ‘publicatiebias’. En dan betreft het niet eens zozeer het verschijnsel dat significante resultaten eerder aangeboden en sneller gepubliceerd worden dan niet-significante.¹⁴ Met name de farmaceutische industrie heeft er een handje van om onderzoek dat niet goed uitkomt, niet te publiceren.¹⁵ Of onderzoek dat niet helemaal goed uitkomt, zo verwarrend te presenteren dat het toch goed uitkomt.¹⁶ Of

Farmaceutische industrie jaagt mensen de dood in

onderzoek dat wel goed uitkomt, drie, vier, desnoods vijf keer te publiceren.¹⁷

De farmaceutische industrie is op grote schaal de wetenschappelijk literatuur aan het vervuilen, en daarmee de geloofwaardigheid van wetenschap aan het ondergraven. En, nog veel belangrijker, mensen de dood in aan het jagen. Maatregelen van medische tijdschriften om hier wat tegen te doen, worden omzeild.¹⁸ Nu klinische trials moeten worden geregistreerd, wordt bijvoorbeeld dui-



delijk dat onafhankelijk onderzoek veel prompter wordt gepubliceerd – ook bij ongunstige uitkomsten – dan onderzoek van de industrie. Verder wijken uitkomstmaten soms af van die in het protocol, en er zijn gevallen bekend waarin het protocol nog na afloop van de dataverzameling werd gewijzigd.

Bij elk wetenschappelijk artikel is het daarom goed om niet alleen te bedenken wie het waarom en waarom nu heeft geschreven, maar ook hoe het past in de rest van de wetenschappelijke literatuur. Het staat niet altijd in het artikel zelf.

Base rate

We eindigen maar met een wat vrolijker noot: het vergeten van de *base rate*, de ‘uitgangsfrequenties’. Het grappigste voorbeeld daarvan was natuurlijk een artikel in *Nature* waarin werd voorgesteld terroristen op vliegvelden op te sporen met infraroodcamera’s (liggende mensen krijgen warme ogen).¹⁹ De test pikte 6 van 8 leugenaars eruit, en vergiste zich slechts bij 1 op 12 niet-leugenaars. Dat lijkt een mooie score, tot we bedenken hoe de verhouding terroristen versus passagiers op een vliegveld is – er vertrekken per jaar ongeveer 40 miljoen reizigers van Schiphol.²⁰ In de artspraktijk liggen de verhoudingen minder extreem, maar ook bij kwesties als het testen op hiv-besmetting, hieprikken, kankerscreening en positieve labuitslagen bij zeldzame aandoeningen is het van belang hierop bedacht te zijn. Temeer omdat keer op keer blijkt dat gezondheidswerkers hier akelig weinig kaas van hebben gegeten. Weinig gynaecologen (34%) en verloskundigen (0%) kunnen een zwangere vrouw vertellen wat een positieve uitslag bij een test op downsyndroom betekent voor de kans op een aangedaan kind.²¹ Maar zelfs als zij het zelf wel begrijpen, is het twijfelachtig of hun boodschap goed overkomt. In een representatieve steekproef werd de Duitse bevolking eens gevraagd wat ‘40 procent’ betekent: is dat een vierde, één op veertig, of vier op tien? Iets meer dan de helft (52%) van de ondervraagden gaf het juiste antwoord, 32 procent gaf een fout antwoord en 14 procent had geen idee.²² *Ais*

De noten staan onder het artikel op www.artsinspe.nl