

# Statistieken, wetgeving en beleid\*

*At long last— now, isn't it terrific:  
Swords into spreadsheets, and digital tears—  
How rulers rule, is truly scientific.*

– Anonieme straatzanger (Covent Garden Market, november 2002).

*It starts from a number; then there is a missing link;  
and it finishes up with draconian legislation.*

–John Brignell

Regelmatig duiken in de media sensationele verhalen op van het type ‘Onderzoek wijst uit dat V een oorzaak is van W’ of ‘Gebruikers van S hebben x keer meer risico op T’. Vaak zijn dat althans op het eerste gezicht erg verontrustende berichten.<sup>1</sup> Scholen, verkiezingscampagnes en brochures en campagnes voor fondsenwerving van allerlei pressie- en actiegroepen dragen ruimschoots hun steentje bij tot het creëren en instandhouden van wat men wel ‘de cultuur van de angst’ noemt.<sup>2</sup>

Bijna al die waarschuwingen berusten op statistisch onderzoek; slechts uitzonderlijk blijkt de berichtgeving te verwijzen naar klinisch of gecontroleerd experimenteel onderzoek. Zij heeft het bijna uitsluitend over verbanden die ‘mogelijk’, ‘naar men vermoedt’, ‘waarschijnlijk’ of ‘het valt niet uit te sluiten’ interessante onderwerpen voor wetenschappelijk onderzoek zijn. Het is evenwel pas nadat onderzoek is gedaan en nadat het de test van de wetenschappelijke kritiek heeft doorstaan dat men kan zeggen of er een causaal mechanisme is dat de statistisch gevonden ‘verbanden’ verklaart. De ‘cultuur van de angst’ drijft niet op wetenschappelijk onderzoek maar op de ruchtbaarheid die gegeven wordt aan met louter statistische methoden gevonden verbanden.

De opmars van de statistiek is niet beperkt tot de media. De uitbreiding van het staatsapparaat in de twintigste eeuw ging gepaard met bureaucrativering en professionalisering. Allerlei ‘deskundigen’ werden aangetrokken om gegevens te verzamelen, onderzoek te doen en beleid te formuleren en te evalueren. Vandaag vin-

---

\* Dit hoofdstuk gaat over het gebruik van statistieken in wetgeving en beleid. Referenties naar publicaties op het vlak van de toegepaste statistiek illustreren de discussie daaromtrent. Zij impliceren geen oordeel over de in die artikelen gevonden correlaties. Dat geldt ook voor de correlaties en relatieve risico's waarvoor geen referentie wordt gegeven.

<sup>1</sup> In de eerste zes maanden van 2001 kregen de Britten via de nationale dagbladders ten minste honderdvijftig waarschuwingen over de gevaren van allerlei etenswaren, activiteiten, gemoedstoestanden, gewoonten, levensstijlen, woonomgevingen, toestellen, geneesmiddelen, operaties, zwangerschappen, weersomstandigheden, en wat al niet. Zie *Health Wars: The Phantom Menace, Volume 1: An Audit of Health Scares January – June 2001*, Compiled by Josephine Gaffikin (FOREST Occasional Paper 3, Audley House, 13 Palace Street, London SW1, December 2001). Het verschijnsel blijft uiteraard niet beperkt tot Groot-Brittannië of de geschreven pers.

<sup>2</sup> Frank Furedi, *The Culture of Fear*, Cassell, London 1997.

den heel wat afgestudeerden van de universiteiten en hogescholen een broodwinning als beleidsmedewerker, onderzoeker of consultant van de overheid.

Ministeriële departementen, overheidsdiensten en ook internationale bureaucratieën staan onder druk om hun werking en budgetten rechtvaardigen. Zij moeten dus aantonen dat zij zich met ‘belangrijke problemen’ bezighouden en ook dat zij bereid en bekwaam zijn daar wat aan te doen. Datzelfde geldt voor politici en politieke partijen. Het uitgangspunt van de verzorgingsstaat is immers dat ‘de mensen zelf’ niet bekwaam of in staat zijn hun leven — dat ‘toch zo ingewikkeld geworden is’ — op een behoorlijke manier te leiden. De implicatie is dat zij hun leven beter door deskundigen kunnen laten leiden. Aan de politici komt dan de taak toe de relevante deskundigen te selecteren. Zij bepalen ook welk onderzoek een ‘politiek gevolg’ krijgt of genegeerd wordt, en hoeveel belastinggeld voor een en ander nodig of beschikbaar is.

Uiteraard laten de politici zich voor die taken graag bijstaan door deskundigen — vaak dezelfde lieden die in aanmerking komen om het te subsidiëren onderzoek uit te voeren. Zo ontstaat er collusie van politieke en academische autoriteiten en elites. Als gevolg daarvan gaan in een aantal ‘gevoelige kwesties’ (zure regen, klimaatopwarming, BSE, aids en dergelijke<sup>3</sup>) honderden miljoenen euro naar het onderzoek van een enkele hypothese. Alternatieve hypothesen blijven in de kou staan. Hoogstens enkele onderzoekers buiten de grote (dat wil zeggen massaal gesubsidieerde) instituten houden zich ermee bezig. Als zij \$10.000 subsidies ontvangen van bepaalde industriële sectoren dan wordt hun dat zwaar aangerekend. Zij zijn dan per populaire definitie ‘niet onafhankelijk’. Tien of honderd keer hogere subsidies—of aanstellingen in een onderzoeksinstituut of programma—van de EU, de Amerikaanse overheid, het EPA, de WGO, *Greenpeace* en andere ‘foundations’<sup>4</sup> met een min of meer uitgesproken politieke agenda hebben daar-

---

<sup>3</sup> Gemeenschappelijk aan deze gevallen is dat zij beginnen met de publicatie van echte rampenscenario’s, die echter jaar na jaar worden afgezwakt. Van de in 1990 voorspelde opwarming van de aarde met 3-5°C tegen 2100 bleef in 1995 nog slechts 1-2°C over. Als de voorspellingen nog verder zakken dan zal Groenland nog lang moeten wachten vooraleer het opnieuw eer kan doen aan zijn naam. Die kreeg het omstreeks 1000 n.C. van de Vikingkolonisten onder leiding van Erik de Rode. De vikings verlieten Groenland in de veertiende eeuw. Dat was het begin van de ‘kleine ijstijd’, die in de zeventiende eeuw zijn laagste temperaturen bereikte en nu langzaam naar zijn einde lijkt te gaan.

De voorspelde aids-epidemie is uitgebleven (behalve in Afrika waar men echter andere criteria en goedkopere en onbetrouwbare methoden gebruikt om aidsgevallen te identificeren—wat de vraag oproept of ‘aids in Afrika’ wel dezelfde ziekte is als aids in Europa of Amerika). Ook de voorspelde BSE-epidemie laat op zich wachten evenals de eveneens voorspelde overdracht ervan op mensen. ‘Zure regen’ is zo goed als volledig uit de belangstelling verdwenen.

<sup>4</sup> Het *Environmental Protection Agency* (EPA) had in 1999 een budget van 7,6 miljard dollar. De WGO (Engels WHO) had in 1997-1998 een budget van 842 miljoen dollar. Volgens V. Ekamov (‘Reform of the WHO’, *The Lancet*, 347, juni 1996, p.1536-1537) besteedt de WHO 75% van zijn budget voor administratieve overhead. De WGO ontvangt bovendien vele honderden miljoenen dollar van ‘private donateurs’ (vrijwillige bijdragen van lidstaten, vooral de VS, organisaties en bedrijven, vooral uit de farmaceutische sector). Zie J.P. Vaughan, S. Mogedal, S.E. Kruse, K. Lee, G. Walt, K. de Wilde, ‘Financing

entegen blijkbaar geen effect op de onafhankelijkheid van onderzoekers. Integendeel, althans voor de media lijken zij veeleer certificaten van ‘onafhankelijkheid’ te zijn.<sup>5</sup>

Statistieken, tellingen, steekproeven en bewerkingen uitgevoerd op de resultaten daarvan zijn nu belangrijke instrumenten van politieke propaganda en beleid. Statistisch berekende risicowaarden worden gewikt en gewogen, aanvaardbaar of te hoog te bevonden. Wij vinden ze terug in allerlei wettelijke en administratieve normen en maatstaven: veiligheidsnormen, concentratienormen, doseringsnormen, snelheidsnormen, en dergelijke. Net boven of net onder de norm — het kan verstrekkende financiële en andere gevolgen hebben. Producten, diensten, beroepen en activiteiten worden gereguleerd, aan belastingen onderworpen of zelfs verboden op grond van ‘statistische bewijzen’ van hun schadelijke effecten op veiligheid, gezondheid, milieu en dergelijke.

De leek kan gemakkelijk de indruk krijgen dat het maatschappijmanagement van de moderne beleidsstaat met alle mogelijke waarborgen van de wetenschappelijke methode omkleed is. Er zijn echter kritische stemmen. Veel van de in het publieke leven gebruikte of geciteerde statistieken zijn controversieel of worden door specialisten zelfs als ondeugdelijk beschouwd. Beschuldigingen van incompetent of bedrieglijk gebruik en misbruik van statistieken zijn niet van de lucht.<sup>6</sup> De term ‘junk science’ (rommelwetenschap) is nu al goed ingeburgerd in discussies over voedselhygiëne, gezondheids- en milieubeleid.<sup>7</sup> In grote rechtszaken, waar reusachtige schadevergoedingen op het spel staan, wordt gegoocheld met statistieken van verdacht allooi. Velen — vooral de zogeheten ‘deep pockets’ (bedrijven, ziekenhuizen, gemeentelijke overheden, hun verzekeraars) — geven er de

---

the World Health Organization: global importance of extrabudgetary funds’ *Health Policy*, 35, 1996, p.229-245. Voor 1998 rapporteerde *Greenpeace International* een netto-inkomen van 101 miljoen dollars. (Zie ook de officiële websites van die organisaties.)

<sup>5</sup> Sommige industrieën (o.a. farmaceutica, medische en milieutechnologie) hebben ‘partnerships’ met—en laten door hen gefinancierd onderzoek vaak verschijnen onder de auspiciën van—die organisaties. Commerciële redenen—verkoop van medische producten en technische installaties—liggen daarbij voor de hand. Andere redenen worden eveneens geciteerd: medische producten en behandelingen (o.a. radiologie, ontsmettingsmiddelen) zijn grote risicofactoren voor allerlei aandoeningen. (John W. Gofman, *Radiation from Medical Procedures in the Pathogenesis of Cancer and Ischemic Heart Disease*, Committee for Nuclear Responsibility Book Division, San Francisco, 1999). Hun aansprakelijkheidsrisico is dermate groot (vooral in de Amerikaanse ‘culture of litigation’) dat vele bedrijven bereid zijn te investeren in goede relaties met het EPA, de WGO en andere ‘public health authorities’. Zij hebben er belang bij het onderzoek te concentreren op ‘andere hypothesen’. De eis dat universiteiten meer aan ‘derde geldstroomactiviteiten’ gaan doen—en de status van wetenschappelijke medewerkers mede laten afhangen van de mate waarin zij de kassa doen rinkelen (in Nederland anno 2002, het UFO-project van de VSNU)—zal nog tot creatievere invullingen van ‘onafhankelijk onderzoek’ leiden.

<sup>6</sup> Bijv. Hans-Joachim Maes: ‘World Health Organization ... mit Entsetzen Scherz.’, *Deutsches Ärzteblatt*, 98/25, 22 juni 2001, p.1664-1666.

<sup>7</sup> Verschillende organisaties brengen regelmatig verslag uit over nieuwe gevallen van ‘junk science’. Zie o.a. The American Council on Science and Health ([www.acsh.org](http://www.acsh.org)), Steven J. Milloy’s [JunkScience.com](http://JunkScience.com), en [www.Quackwatch.com](http://www.Quackwatch.com).

voorkeur aan zich naar grondloze normen te schikken of geschillen in der minne te regelen. Dat is goedkoper dan een aanslepend geding met onzekere afloop te beginnen en ondertussen in de pers te schande gemaakt te worden.

Alleen al dreigen met een geding wordt zo voor advocaten, forensische deskundigen en adviseurs een lucratieve bezigheid. Op dezelfde wijze kunnen politici en ambtenaren druk uitoefenen op bedrijven, sectoren of bevolkingsgroepen door te dreigen met reguleringen en normeringen waarvan zijzelf weten dat zij geen serieuze basis hebben. Dergelijke gevallen van 'legale afpersing' zijn verre van zeldzaam, zowel op het nationale als het internationale vlak.<sup>8</sup>

Toch laten leken—en veel mensen zijn leken als het op statistieken aankomt—zich blijkbaar vlot imponeren door cijfers, percentages, risicometingen en dergelijke. Politici en ook juristen zijn geen uitzondering op die regel. Dat is geen goede zaak. Zij hebben geregeld te maken met wetten, beleid en argumentaties waarin statistische gegevens, theorieën en conclusies prominent aanwezig zijn. Als zij niet bekwaam zijn goed en slecht statistisch onderzoek, gepast gebruik en misbruik van statistiek te onderscheiden dan lopen zij het risico onzin te verkondigen en veel kwaad te doen. Dat er gevallen zijn waarin er met de toepassing van statistische methoden prachtige en nuttige resultaten worden geboekt, is onbetwistbaar. Er zijn echter ook andere gevallen. Van de vooroordelen en de statistische incompetentie van nationale, supranationale of internationale overheids- en beleidadviserende instanties kunnen duizenden tot miljoenen mensen het fysieke of financiële slachtoffer zijn.

Dat risico is reëel. In 1991 werd onder druk en invloed van *Greenpeace* en het *Environmental Protection Agency* (EPA, een Amerikaanse overheidsdienst) het chloreren van drinkwater in Peru stopgezet.<sup>9</sup> De maatregel leidde vrijwel onmiddellijk tot een cholera-epidemie. Ten minste achthonderdduizend mensen werden ziek, waarvan ten minste een zesduizendtal met dodelijke afloop. De maatregel was gebaseerd op de toen door *Greenpeace* gepropageerde *idée fixe* dat chloor in alle concentraties en doseringen giftig is en op de classificatie door het EPA van chloor als een risicofactor voor kanker. Chloor is inderdaad een bijtende stof die bijna alle metalen kan aantasten. Het is ook een component in honderden voor plant en dier vitale stoffen, onder meer zout en de menselijke verteringsapparaat. Voor het ontsmetten van drinkwater is het buitengewoon goed geschikt.

Onoordeelkundig gebruik van statistieken heeft lange tijd officiële en andere instanties ertoe aangezet jonge ouders aan te bevelen baby's op hun buik te laten

---

<sup>8</sup> Een econoom van de Wereldbank: "Policy based lending is where the bank really has power--I mean brute force. When countries really have their backs against the wall, they can be pushed into reforming things at a broad policy level that normally, in the context of projects, they can't." (Geciteerd in Kamran Abbasi, 'The World Bank and world health: healthcare strategy', *British Medical Journal*, 318, April 3 1999, p.933-936.)

<sup>9</sup> *Nature* (November 28, 1991), U.B. Panisset, *International Health Statecraft: Foreign Policy and Public Health in Peru's Cholera Epidemic* (Lanham MD: University Press of America, 2000). Volgens *Greenpeace* ([Archive.greenpeace.org/~toxics/reports/cholera-chlorine.pdf](http://Archive.greenpeace.org/~toxics/reports/cholera-chlorine.pdf)), dat campagne voerde voor een algemeen chloorverbod, had de epidemie kunnen vermeden worden door het water te koken en betere persoonlijke hygiëne.

slapen. Toen de statistieken een opmerkelijke toename van wiegendoed suggereerden werd dat advies stopgezet. Economisch en monetair beleid, vaak gebaseerd op een combinatie van achterhaalde statistieken uit diverse bronnen, heeft in de twintigste eeuw meer dan eens miljoenen mensen in zware financiële problemen gebracht.

Ofschoon dergelijke episodes vragen rond verantwoordelijkheid en aansprakelijkheid oproepen, blijken maatschappelijke instellingen (ook de rechtbanken) daar weinig oor voor te hebben. Beleid en wetgeving zijn officieel erkende methoden om straffeloos ‘effecten te externaliseren’. De slachtoffers die zij maken hebben gewoon pech.

Het is voor academisch geschoolde juristen — die toch aanspraak maken op een ‘wetenschappelijke vorming’ — van belang enig inzicht te hebben in het gebruik van statistische methoden. In deze tekst kunnen wij uiteraard geen volledige instructie in de wetenschap van de statistiek en haar toepassingen geven. Dat zou ons al snel tot de studie van de wiskundige waarschijnlijkheidsleer brengen en die ligt ver buiten het bestek van de juridische opleiding. Wij beperken ons tot een discussie van enkele elementaire begrippen en problemen.

## Dingen en cijfers

Statistiek is een tak van de wiskunde en als zodanig een volwaardige wetenschappelijke discipline. Zij houdt zich bezig met de studie van de formele relaties tussen verzamelingen van abstracte objecten. De toepassing van de statistische methoden op de werkelijkheid is evenwel geen sinecure. Reële objecten zijn geen abstracte objecten. Neem een zo alledaags object als een huis. Hoeveel huizen zijn er in een bepaald dorp? Verschillende waarnemers zullen een verschillend aantal huizen tellen, bijvoorbeeld omdat zij ‘randgevallen’ (hutten, stacaravans, bungalows, appartementen gebouwen, als woning gebruikte opslagplaatsen, vervallen gebouwen, en dergelijke) verschillend beoordelen.

Als alleen al het tellen van dingen problemen oplevert, dan is ook het verwerken van die gegevens problematisch. Als men gegevens heeft over het aantal huizen in elke Nederlandse provincie, dan lijkt het voor de hand te liggen dat men door eenvoudige optelling kan berekenen hoeveel huizen er in Nederland zijn. Dat is echter niet het geval. De provinciale huizentellers hebben immers niet noodzakelijk allen dezelfde criteria gehanteerd. Misschien zijn er gebouwen, die al dan niet als huis zouden geteld zijn, helemaal buiten het gezichtsveld van de tellers gebleven. Niet alle tellers zijn even zorgvuldig en gewetensvol. De gegevens zijn dus hoogstens min of meer volledig en betrouwbaar.

De problemen nemen nog toe als men in de plaats van huizen te tellen indirecte bronnen (belastingaangiften bijvoorbeeld of orderboekjes van loodgieters en dakwerkers) gebruikt om het aantal huizen vast te stellen. Nog vele malen onbetrouwbaarder zijn niet of moeilijk controleerbare bronnen, zoals herinneringen en rapportages achteraf.

Wat voor huizen geldt, geldt nog veel meer voor ‘dingen’ als ziektes, aandoeningen, gemoedstoestanden, koffiedrinkers, rokers, kansarmen, wiskundig be- gaafden en andere categorieën van zaken die moeilijk te identificeren zijn of de relevante eigenschap in meerdere of mindere mate hebben. Toch zijn dergelijke ‘dingen’ vaak voorwerp van statistisch onderzoek.

Waar het om gaat is dit: wij kunnen geen statistische methoden toepassen op re- ele objecten maar alleen op abstracte objecten zoals resultaten van tellingen, me- tingen of schattingen. Kortom, wij hebben getallen nodig. Voor het gebruik van de statistische methoden maakt het niets uit of wij met zestig huizen of zestig aal- wormen te maken hebben; alleen het getal 60 telt. Hoe wij aan dat getal komen is geen probleem van de wiskundige statistiek. Het is uiteraard wel een probleem voor wie statistische methoden wil gebruiken om iets over huizen of over aalwor- men te weten te komen. Toegepast statistisch onderzoek kan niet beter zijn dan de kwaliteit van de gegevens waar het op gebaseerd is. Hier geldt het GIGO-principe: ‘Garbage in, Garbage out’.

Zodra wij over de getallen beschikken, kunnen wij daar allerlei wiskundige be- werkingen op uitvoeren: optellen, aftrekken, vermenigvuldigen, delen en ook meer ingewikkelde operaties. Dat geeft al snel eigenaardige resultaten. Zo kunnen wij bijvoorbeeld tot de vaststelling komen dat er vorige maand 3,3 verkeersdoden per dag vielen of dat er per Nederlands gezin 1,92 kinderen zijn. Wij kunnen ons wel gemakkelijk een voorstelling maken van 3 of 4 verkeersdoden en van 1 of 2 kinderen—maar van 3,3 doden of 1,92 kinderen? Bewerkingen uitgevoerd op tel- lingen van bestaande dingen leveren niet noodzakelijk een uitkomst op die zelf beantwoordt aan een telling van bestaande dingen.

Het is alleen zinvol over gemiddelden te spreken als er dingen zijn die boven het gemiddelde liggen en andere die daar onder liggen. Er is geen relatieve ar- moede zonder relatieve rijkdom. Er zijn kansrijken omdat er kansarmen zijn. Het is zinloos te willen ‘dat iedereen ten minste het gemiddelde haalt’ (tenzij men ook wil dat niemand meer dan het gemiddelde haalt).

Mensen zijn niet onsterfelijk. Sommige doodsoorzaken komen meer voor dan andere. Vooruitgang in de strijd tegen de ene of de andere ziekte of een betere be- schermt tegen een bepaald soort ongevallen veranderen daar niets aan. Er is al- tijd wel ‘een belangrijkste doodsoorzaak’. Wij kunnen niet de kans verminderen dat iemand aan ziekte Z overlijdt zonder de kans te vergroten dat hij door een an- dere oorzaak sterft. Dat wil zeggen: elke succesrijke behandeling voor Z is een positieve risicofactor (zie hieronder) voor andere overlijdensoorzaken. Stel dat er nu in vergelijking met dertig jaar geleden jaarlijks relatief meer mensen door ziekten of aandoeningen sterven. Dat is niet per se een reden tot ongerustheid. Misschien zijn er nu relatief minder dodelijke ongevallen.

## Correlaties

Wanneer wij over twee verzamelingen van cijfers beschikken, dan kunnen wij die vergelijken en de maat van hun overeenkomst of verschil aangeven. Dat doen wij

met behulp van een correlatiecoëfficiënt (CC), waarvan de waarde maximaal +1 en minimaal -1 is. Hoe wij die berekenen is hier niet van belang. De volgende allereenvoudigste voorbeelden maken duidelijk wat zo'n coëfficiënt eigenlijk meet. We vergelijken twee verzamelingen van getallen X en Y, die elk vier getallen tellen.

<b>X1</b>	1	1	0	0	<b>CC</b>
<b>Y1</b>	1	1	0	0	<b>1</b>

Y is perfect gelijk aan X. Het is dus niet verwonderlijk dat de CC maximaal is.

<b>X1</b>	1	1	0	0	<b>CC</b>
<b>Y2</b>	0	0	1	1	<b>-1</b>

Y is zeker niet gelijk aan X; integendeel, Y is als het ware een omgekeerd beeld van X. Met een '1' in X komt een '0' in Y overeen en met een '0' een '1'. Kortom, er zit een systeem in de verschillen. De CC-waarde is in dit geval minimaal, namelijk -1.

<b>X1</b>	1	1	0	0	<b>CC</b>
<b>Y3</b>	1	0	1	0	<b>0</b>

In dit geval is er geen overeenkomst, zelfs niet als spiegelbeeld. Een '1' in X correspondeert nu eens met een '1' in Y dan weer met een '0'; en hetzelfde geldt voor een '0' in X. De CC bedraagt 0, dat wil zeggen dat er geen overeenkomst of correlatie tussen de twee verzamelingen is.

Nu moeten wij niet overhaast concluderen dat  $CC=1$  betekent dat de twee verzamelingen identiek zijn. Beschouw dit voorbeeld:

<b>X2</b>	1	1	1	1	1	0	0	<b>CC</b>
<b>Y4</b>	12	12	12	12	12	5	5	<b>1</b>

Hoewel X en Y niet identiek zijn is er toch een maximale correlatie tussen de twee ( $CC=1$ ). Met een '1' in X correspondeert onveranderlijk een '12' in Y; en met een '0' in X komt onveranderlijk een '5' in Y overeen. Correlatie is dus wat anders dan identiteit.

<b>X2</b>	1	1	1	1	0	0	<b>CC</b>
<b>Y5</b>	12	12	12	12	25	25	<b>-1</b>

Ook hier is er een 'perfecte' correlatie tussen de twee reeksen maar met dit verschil dat met een *hogere* waarde in de eerste reeks (1 is groter dan 0) een *lagere* waarde (12 is kleiner dan 25) in de tweede reeks correspondeert.

<b>X2</b>	1	1	1	1	1	0	0	<b>CC</b>
<b>Y6</b>	12	18	22	15	29	6	3	<b>0,8</b>

(Exacter  $CC = 0,7953502167$ ) Hier is de correlatie van de twee verzamelingen niet perfect, maar toch vrij hoog. Met een *hogere* waarde (1) in X correspondeert een min of meer hoge waarde (12 tot 29) in Y. Met een *lagere* waarde (0) in de eerste reeks correspondeert een min of meer lage waarde (3 of 6) in de tweede reeks.

<b>X2</b>	1	1	1	1	1	0	0	<b>CC</b>
<b>Y7</b>	12	0	2	5	12	25	-25	<b>0,2</b>

(Exacter  $CC = 0,1959828375$ ) Op het eerste gezicht is er hier geen systeem van overeenkomst te herkennen. Met de  $CC=0,2$  zitten wij inderdaad dicht in de buurt van  $CC=0$ , wat staat voor 'er is helemaal geen correlatie'.

Bekijken wij ten slotte nog een setje van twee getallenverzamelingen

<b>X3</b>	-7	2,5	800	12	45,04	151	-8	<b>CC</b>
<b>Y8</b>	0,25	0,56	5	1	3,27	4,56	0,12	<b>0,7</b>

(Exacter  $CC = 0,7341022013$ ) Ook hier is er op het eerste gezicht weinig overeenkomst. De relatief hoge correlatiecoëfficiënt 0,7 wijst echter op een eerder systematische samenhang tussen de getallen in X en de getallen in Y. Die is er ook: met een hogere waarde in X correspondeert onveranderlijk een hogere waarde in Y. Met het hoogste getal (800) in X correspondeert het hoogste getal (5) in Y. Met het tweede hoogste getal (151) in X correspondeert het tweede hoogste getal (4,56) in Y. Enzovoort. Met het laagste getal (-8) in X correspondeert het laagste getal (0,12) in Y.

Nog even herhalen:

- Een hoge *absolute* waarde van de correlatiecoëfficiënt (positief of negatief, in de omgeving van 1 of -1) betekent een sterke graad van overeenkomst tussen twee getallenreeksen. Een lage *absolute* waarde (positief of negatief, in de omgeving van 0) betekent een zwakke correlatie.
- Statistici werken met getallen en met verzamelingen van getallen. Het enige wat wij met statistieken kunnen bewijzen is dat er tussen twee of meer getallenreeksen bepaalde overeenkomsten of verschillen bestaan. Waar die getallen voor staan, of ze überhaupt ergens voor staan, dat maakt voor de statistiek niet uit.

## Toegepaste statistiek

Bij toegepaste statistiek komt het uiteraard aan op 'waar de getallen voor staan'. Nemen wij als voorbeeld de eerder getoonde verzamelingen X2 en Y6.

<b>X2</b>	1	1	1	1	1	0	0	<b>CC</b>
<b>Y6</b>	12	18	22	15	29	6	3	<b>0,8</b>

Wij gaan die getallen nu interpreteren, dat wil zeggen wij gaan ze beschouwen als symbolen die naar iets verwijzen. Wij kunnen bijvoorbeeld veronderstellen dat er een onderzoek gedaan is in de zeven gemeenten van een bepaalde regio van het land. Daar werd voor elke gemeente geteld hoeveel keer een medisch urgentieteam werd opgeroepen bij verkeersongevallen in de eerste zes maanden van het jaar 2000. Die oproepen staan per gemeente vermeld in Y. In X staat een '1' als de gemeente een mannelijke burgemeester had en een '0' als er een vrouwelijke burgemeester was.<sup>10</sup>

Er blijkt dus, althans in de onderzochte regio, een vrij sterk verband te bestaan tussen het optreden van medische urgentieteams bij verkeersongevallen en het geslacht van de burgemeester. Het is evenwel slechts een statistisch verband—een verband tussen twee getallenreeksen.

Het is niet waarschijnlijk dat er iets mis is met de gegevens van ons onderzoek. Het geslacht van een burgemeester staat meestal onomstotelijk vast en oproepen voor een medisch urgentieteam bij verkeersongevallen worden zorgvuldig geregistreerd en zijn niet zo moeilijk te onderscheiden van andere oproepen. Aangezien alle gemeenten in de regio werden onderzocht is er ook geen twijfel aan de 'representativiteit' van de onderzochte gemeenten voor de situatie in die regio. Daarmee is echter niets gezegd over de vergelijkbaarheid van de verschillende gemeenten, bijvoorbeeld qua grootte, industriële activiteit, verkeersdichtheid, verkeersinfrastructuur en dergelijke. Ook is niets gezegd over de representativiteit van de periode waarin het onderzoek liep. Als zou blijken dat er in de gemeenten met een mannelijke burgemeester in de eerste helft van 2000 toevallig veel uitzonderlijke verkeerssituaties waren (wegwerkzaamheden, sportmanifestaties, wegblokkades) dan zou dat het verschil in de cijfers kunnen verklaren.

De statistiek bewijst of verklaart niets. Zij geeft in het beste geval een beschrijving aan de hand van getallen van een situatie die misschien wel, maar misschien ook niet, om een verklaring vraagt. Een hoge correlatie kan louter toevallig zijn of het resultaat van een buiten het onderzoek gebleven factor. Uit Duits onderzoek bleek bijvoorbeeld dat huizen bewoond door grote gezinnen vaker door ooievaars als nestplaats werden uitgekozen. Dat betekent niet dat er een grond van waarheid is in het bakerpraatje dat ooievaars baby's brengen. Grote gezinnen wonen meestal in grotere huizen — en die hadden (in de tijd van verwarming met haarden en kolenkachels) meer schouwen en schoorstenen, dus meer potentiële nestplaatsen voor ooievaars!

Terug naar ons voorbeeld. Het is gewoon een feit dat er in de gemeenten met een vrouwelijke burgemeester in de regio minder oproepen voor medische interventie waren dan in gemeenten met een mannelijke burgemeester. Daarmee is niet gezegd dat het geslacht van de burgemeester iets te maken heeft met het aantal verkeersongevallen waarbij dringende medische interventie nodig is. Vermoede-

---

<sup>10</sup> Wij gebruiken hier de cijfers '1' en '0' niet om resultaten van tellingen of metingen aan te duiden maar als symbolen voor verschillende categorieën van dingen. Wij hadden ook de letters 'M' en 'V' kunnen gebruiken, maar die worden niet herkend door de rekenmachine waarmee wij de correlatiecoëfficiënt berekenen!

lijk zullen er nog wel andere kenmerken zijn die de gemeenten met hogere aantallen medische interventies onderscheiden van de andere. Stel bijvoorbeeld dat de gemeenten met een vrouwelijke burgemeester de letter ‘b’ in hun naam zouden hebben en die met een mannelijke burgemeester niet. Wie zal de correlatie tussen het al of niet voorkomen van een letter in de naam van een gemeente en het aantal zware verkeersongevallen als interessant beschouwen?

Het is niet waarschijnlijk dat iemand enige *wetenschappelijke* belangstelling zou hebben voor onze statistiek alleen maar omdat wij de gemeenten volgens het geslacht van hun burgemeester hebben geclassificeerd. Het blijft dezelfde statistiek over dezelfde gemeenten en dezelfde feiten. Toch kan er heel wat buitenwetenschappelijke belangstelling voor onze statistiek bestaan. Een lokale krant zou er een artikel aan kunnen wijden onder de kop

### **Vrouwelijke burgemeesters goed voor verkeersveiligheid**

Recent onderzoek aan de universiteit van X heeft uitgewezen dat in onze regio zware verkeersongevallen, waarbij mensen ernstig gewond raken, vooral gebeuren in gemeenten met een mannelijke burgemeester. Volgens de onderzoekers werd in een periode van zes maanden per gemeente met een man als burgemeester gemiddeld 19,2 keer een beroep gedaan op een medisch urgentieteam voor verkeersslachtoffers. In de gemeenten met een vrouwelijke burgemeester was dat gemiddeld slechts 4,5 keer. De onderzoekers wijzen erop dat men hieruit geen voorbarige conclusies mag trekken en dat verder onderzoek nodig is om dit opvallende verschil te verklaren. Het verschil is groot genoeg om er niet onverschillig voor te blijven.

Die krantenkop is zonder meer misleidend. Hij suggereert dat er een causaal verband bestaat tussen het geslacht van de burgemeester en het aantal ernstige ongelukken in een gemeente. Het onderzoek heeft daaromtrent echter niets uitgewezen. De tekst zelf van het artikel doet de waarheid weinig geweld aan, maar is wel vaag. Het is bijvoorbeeld niet duidelijk of de laatste zin een redactionele commentaar is of een uitspraak die aan de onderzoekers wordt toegeschreven.

Wij kunnen ons voorstellen dat bepaalde groepen proberen politieke munt uit het artikel te slaan zonder daarbij uitdrukkingen als ‘het is wetenschappelijk / statistisch bewezen’, ‘onderzoek heeft aangetoond’ te schuwen. Het is moeilijk te voorspellen wat er met statistieken kan gebeuren zodra men ze uit de geborgenheid van het wetenschappelijk onderzoek haalt en in de openbaarheid gooit.

### **Het relatieve risico**

Wij kunnen onze statistiek ook op een ander manier voorstellen—een manier die het tendentieuze gebruik ervan nog aantrekkelijker maakt. Veronderstellen wij dat er ook cijfers bestaan voor het aantal interventies van de politie bij verkeersongevallen in de zeven gemeenten.

*Eerste hypothese:*

<b>Burgemeester</b>	<b>Interventie</b>	
	politie + urgentieteam	alleen politie
Man	96	853
Vrouw	9	341

De verhouding van de verkeersongevallen met medische interventie in de twee categorieën van gemeenten is  $96 / 9 = 10,67$ . Voor de verkeersongevallen zonder medische interventie is de verhouding  $853 / 341 = 2,50$ . De verhouding van die twee verhoudingen is

$$10,67 / 2,50 = 4,268.$$

Dat is wat statistici het ‘relatieve risico’ (RR) noemen. In ons geval is dat het relatieve risico van een mannelijke versus een vrouwelijke burgemeester voor ‘ernstige verkeersongevallen’. Het berekende relatieve risico zegt ons dat de kans dat iemand bij een verkeersongeval zwaargewond raakt meer dan vier keer groter is in een gemeente met een mannelijke dan in een gemeente met een vrouwelijke burgemeester!

Het relatieve risico 4,268 komt aardig overeen met de verhouding tussen de gemiddelden van de zware ongevallen in ‘mannelijke’ en ‘vrouwelijke’ gemeenten:

$$(96/5) / (9/2) = 19,2 / 4,5 = 4,267.$$

Dat is gemakkelijk te verklaren. Bij onze eerste hypothese zijn wij er van uitgegaan dat er in elke gemeente — ongeacht het geslacht van de burgemeester — ongeveer evenveel ongevallen zonder medische interventie gebeuren, namelijk circa 170. Statistisch gesproken maken die ongevallen onder de eerste hypothese dus geen verschil tussen de twee soorten gemeenten. Laten wij evenwel wat ‘spelen’ met het aantal verkeersongevallen zonder medische interventie.

*Tweede hypothese:*

<b>Burgemeester</b>	<b>Interventie</b>	
	politie + urgentieteam	Alleen politie
Man	96	1853
Vrouw	9	341

Er zijn nu duizend verkeersongevallen meer in de gemeenten met een mannelijke burgemeester. Die gemeenten zijn dus nog beduidend verkeersonveiliger dan wij eerst hadden aangenomen. Wat gebeurt er echter met het relatieve risico? Dat bedraagt nu

$$(96/9) / (1853/341) = 1,963$$

afgerond 2. Hoewel de verkeerssituatie in de ‘mannelijke’ gemeenten duidelijk slechter is dan onder de vorige hypothese, is het ter discussie staande relatieve risico meer dan gehalveerd. De kans dat iemand bij een verkeersongeval zwaargewond raakt is nog slechts twee keer groter in een gemeente met een mannelijke dan in een gemeente met een vrouwelijke burgemeester.

*Derde hypothese:*

<b>Burgemeester</b>	<b>Interventie</b>	
	politie + urgentieteam	Alleen politie
Man	96	1853

Vrouw	9	34
-------	---	----

Gemeten aan het aantal verkeersongevallen is de verkeerssituatie in de ‘vrouwelijke’ gemeenten nu heel wat veiliger: een totaal van 43 ongevallen tegenover 350 onder de twee vorige hypothesen. Het relatieve risico bedraagt nu evenwel

$$(96/9) / (1853/34) = 0.196$$

afgerond 0,2. De kans dat iemand bij een verkeersongeval zwaargewond raakt in een gemeente met een mannelijke is nog slechts *een vijfde* van die kans in een gemeente met een vrouwelijke burgemeester.

Noteer dat er aan de cijfers van de reeksen X2 en Y6 niets gewijzigd is. Het blijft mogelijk het eerder geciteerde krantenartikel onder dezelfde misleidende titel te publiceren zonder de statistische gegevens geweld aan te doen. Evenwel is het ook mogelijk, al naargelang de cijfers over het totale aantal verkeersongevallen in de betrokken gemeenten in dezelfde periode, andere en evenzeer misleidende titels te maken.

Relatieve risico’s worden vaak geciteerd in de pers. ‘Zoveel meer kans op kanker’, ‘Zoveel meer kans om in de gevangenis te belanden / werkloos / depressief / zwaarlijvig te worden’ en dergelijke meer zijn krantenkoppen die altijd de aandacht trekken. De geciteerde getallen doen er vaak niet toe. ‘Vijf keer meer kans’ klinkt indrukwekkend, maar ‘Vijftig procent meer kans’ evenzeer — hoewel met het eerste een relatief risico van 5 correspondeert en met het tweede een relatief risico van 1,5.

Bekijken wij nog even de volgende statistiek. Zij geeft voor een bepaalde luchtvaartmaatschappij aan hoeveel dodelijke slachtoffers er bij ongevallen met haar vliegtuigen gevallen zijn.

	Verongelukt	Niet verongelukt
Personeel	14	520
Passagiers	609	4123584
Pers/Pas	0,023	0,000126

Het relatieve risico bedraagt 182,5. Voor een piloot of steward (van die maatschappij) is de kans dat hij door een vliegtuigongeluk om het leven komt dus 182,5 keer groter dan voor een passagier. Dat is een enorm verschil. Het betekent echter niet dat men wel gek moet zijn om piloot of steward te worden. Het betekent zeker niet dat piloten en stewards een grote kans hebben om bij een crash om het leven te komen. Uit het relatieve risico kunnen wij niets afleiden over het *absolute risico*.

Relatieve risico’s duiken geregeld op in discussies rond wetsvoorstellen en beleidsvoorstellen, bijvoorbeeld om het roken te ontmoedigen of te verbieden. Anti-tabak lobbies komen graag met statistieken waaruit blijkt dat rokers jaarlijks 15 keer meer kans hebben aan een hartziekte te sterven dan niet-rokers. Uit hun cijfers blijkt echter ook dat het gaat om de verhouding tussen 7 per honderdduizend niet-rokers en 104 per honderdduizend rokers. Nu is  $104 / 7$  inderdaad wel ongeveer gelijk aan 15. Het geciteerde relatieve risico is dus wel correct. Uit de cijfers kunnen wij echter ook afleiden dat de jaarlijkse kans dat rokers *niet* aan een der-

gelijke ziekte overlijden 99,9% bedraagt van die kans voor niet-rokers. Voor longkanker, waar het gaat om 7 per honderdduizend niet-rokers en 166 per honderdduizend rokers, bedraagt het relatieve risico bijna 24. Bekijken wij echter de jaarlijkse kans om *niet* aan longkanker te sterven dan bedraagt die voor rokers 99,8% van die kans voor niet-rokers. Of: hoe ‘vierentwintig keer meer kans’ zich laat vertalen als ‘praktisch evenveel kans’!

Omzetting van het genoemde jaarlijkse risico in een risico over het gehele leven is gebaseerd op allerlei vermoedens en hypothesen en is dus controversieel. Het absolute risico van overlijden aan longkanker voor zware rokers zou tussen 4% en 6% liggen. Een groot aantal daarvan bereikt evenwel een hoge leeftijd, boven de gemiddelde levensverwachting.

Wij moeten er uiteraard rekening mee houden dat er grote statistische verschillen zijn tussen bevolkingsgroepen en nationaliteiten<sup>11</sup> en dat de nationale statistieken van verschillende landen niet altijd (meestal niet) vergelijkbaar zijn. In 1976 bedroeg het sigarettenverbruik per volwassen persoon in België en Mexico gemiddeld 1500. In België waren er per honderdduizend inwoners ongeveer 55 mensen die aan longkanker overleden zouden zijn, in Mexico minder dan vijf.<sup>12</sup> Longkanker is, zoals de meeste kankers, een ziekte van oude mensen, en de gemiddelde leeftijd bij overlijden was in België in 1976 beduidend hoger dan in Mexico. Ook tussen Noord- en Zuid-Europa bestaan er opmerkelijke verschillen in het relatieve risico van longkanker.<sup>13</sup>

## Epidemiologisch onderzoek

De epidemiologie is de statistische studie van ziekteverschijnselen. Zij is vooral belangrijk wanneer het gaat om zeldzame maar ernstige aandoeningen waarvan men mag aannemen dat zij een heel specifieke oorzaak hebben. Zij dankt haar aanzien in het bijzonder aan de succesrijke toepassing van statistische methoden bij het identificeren van de oorzaak van cholera. Door registratie van de woonplaats van cholera-patiënten konden onderzoekers achterhalen dat cholera vooral voorkwam bij mensen die in de buurt van bepaalde publieke waterputten woonden. Er was een sterke correlatie tussen het aantal patiënten en die waterputten. Het medisch en wetenschappelijk onderzoek kon zich dank zij dat gegeven toespitsen op de kwaliteit van het drinkwater. Dat was een spectaculaire doorbraak in de strijd tegen die oude plaag. Let wel, de statistische benadering bewees niet dat

---

<sup>11</sup> Zie de *Country Reports* van de WHO (website van de Wereldgezondheidsorganisatie). Voor Japan, met 59.0% rokers (1994, 15 jaar en ouder) waren er tien jaar geleden 81.2 longkankersterfgevallen per 100.000 rokers; voor de USA waren de overeenkomstige cijfers 28.1% rokers (1991) en 305.7 sterfgevallen per 100.000 rokers. Die grote verschillen blijven vooralsnog onverklaard.

<sup>12</sup> James Le Fanu, ‘A Healthy Diet—Fact or Fiction?’ in *Health, Lifestyle & Environment* (Social Affairs Unit / Manhattan Institute, London, 1991)

<sup>13</sup> Althans volgens het cultuurvergelijkend onderzoek van Ancel Keyes naar de ‘mediterrane levensstijl’, zie James Le Fanu, eerder geciteerd.

het drinkwater besmet was met de bacil die cholera veroorzaakt. Zij bracht alleen aan het licht waar men naar de oorzaak moest gaan zoeken.

Ook in andere epidemieën werd die methode met succes toegepast. Naarmate epidemieën echter zeldzamer werden, gingen meer en meer epidemiologen hun deskundigheid toepassen op situaties die daarvoor nauwelijks of niet geschikt waren. Vrij algemeen verspreide verschijnselen (niet alleen ziekten maar ook bijvoorbeeld criminaliteit, inkomenspositie, kinderrijkdom) werden zo onderzocht ofschoon er geen reden was om te geloven dat zij een heel specifieke maar onbekende oorzaak hebben. De epidemiologie hield op een hulpmiddel te zijn dat het mogelijk maakt te bepalen waar men het best naar de oorzaak van een verschijnsel kan zoeken. Zij ging zich bezighouden met de 'imaginative observation'<sup>14</sup> van allerlei gegevens in de hoop daarin positieve of negatieve correlaties te vinden en zo lijstjes met 'risicofactoren' op te stellen. Voor hartziekten zijn op die manier nu al meer dan driehonderd risicofactoren genoemd: alcoholisme, geheelonthouding, roken, te weinig melk, te veel melk, lawaai, snurken, leeftijd, zwaarlijvigheid, buitenechtelijke seks, Engels als moedertaal, trage baardgroei, gebruik van de pil, enig kind zijn, het vijfde of nog jongere kind zijn in een groot gezin, koffie, stadsbewoner, vroege menopauze, lage maatschappelijke status, enzovoort, enzovoort.<sup>15</sup> Die correlaties zijn bijna allemaal heel zwak. Vaak werden zij in ander onderzoek (in andere landen, met andere steekproeven) niet teruggevonden. Zij hebben bijna nooit geleid tot een wetenschappelijke of medische doorbraak.

Dat wil niet zeggen dat zij geen aanleiding zijn geweest voor allerlei speculaties. Integendeel, dergelijke correlaties worden door velen aangegrepen om de ene of de andere theorie te propageren en op grond daarvan producten, remedies, behandelingen of beleid aan de man te brengen. Theorieën verzinnen kan iedereen. Dat iemand een theorie publiceert en daar in de media belangstelling voor krijgt (of weet te bewerkstelligen), betekent uiteraard niet dat die theorie wetenschappelijk onderzocht, laat staan bevestigd is. Toch is een publicatie (een tijdschriftartikel of een presentatie op een congres) voor de media vaak genoeg om risicofactoren als 'oorzaken' en de daarbij gesuggereerde hypothesen als 'wetenschappelijke verklaringen' voor te stellen.

Er is evenwel reden genoeg om zelfs met heel zorgvuldig en competent uitgevoerd epidemiologisch onderzoek en de gepubliceerde resultaten ervan voorzichtig om te springen. Wij zullen dat aantonen met de volgende fictieve casus.

Laat ons aannemen dat zowel de factor F (een chemische stof bijvoorbeeld) als

---

<sup>14</sup> R. Doll & R. Peto, 'The causes of cancer', *Journal National Cancer Institute*, 66, 1981, p.1192-1312. De auteurs hebben er echter geen bezwaar tegen dat hun 'verbeelding' gebruikt wordt om dure onderzoekprogramma's en ingrijpende beleidsvoorstellen te lanceren. Noteer dat het artikel onder de titel 'causes of cancer' verscheen, terwijl het inhoudelijk alleen over risicofactoren ging! 'Oorzaken' veronderstellen een bekend etiologisch mechanisme of proces; 'risicofactoren' niet meer dan een statistisch verband.

<sup>15</sup> Twintig jaar geleden stond de teller op 246: P.N. Hopkins & R.R. Williams, 'A survey of 246 suggested coronary risk factors', *Arteriosclerosis*, 40, 1981, p.1-52. Ondertussen is de lijst elk jaar met enkele eenheden aangegroeid.

het verschijnsel V (bijvoorbeeld een ziekte) waargenomen worden op een eiland. De stof F is enkele jaren geleden op het eiland geïntroduceerd en heeft er op korte tijd een algemene verspreiding gevonden. Er zijn lieden die beweren dat F het risico van V verhoogt, maar zij hebben daar geen klinisch bewijs voor. De geneeskunde kent geen mechanisme (een fysiologisch proces bijvoorbeeld) dat de beweerde oorzaak F rechtstreeks of onrechtstreeks omzet in V, het beweerde gevolg. Dat verhindert niet dat er bezorgdheid ontstaat en dat er zich een debat ontspint.

Het eiland, dat een miljoen bewoners heeft, is verdeeld in 10 provincies. In elke provincie wonen honderdduizend mensen. Er zijn tien epidemiologen aan het werk, een in elke provincie. Onafhankelijk van elkaar nemen zij zich voor het verband tussen F en V te onderzoeken. Ieder van hen bestudeert de bevolking van zijn provincie. Wij nemen aan dat zij uitermate gewetensvol en competent zijn.

*Eerste hypothese:* Wij veronderstellen dat wij alles weten wat er over F en V op het eiland te weten is. Wij weten bijvoorbeeld dat een op tienduizend mensen daar aan V lijdt en ook dat de factor F het risico van V met ongeveer 50% verhoogt. Wij weten dat wel, maar op het eiland weet men het niet!

Eerst verzamelen onze epidemiologen in de geneeskundige archieven van hun provincie de gegevens over de V-lijdende in het laatste jaar voor F werd ingevoerd. Zo kunnen zij weten hoeveel mensen aan V leden in een F-vrije omgeving. De resultaten staan in de volgende tabel. Per provincie zijn er gemiddeld 10,2 patiënten. Aangezien er in elke provincie 100.000 mensen wonen, zijn die resultaten goed te verenigen met het [aan ons bekende] feit dat ongeveer een op tienduizend mensen aan V lijdt.  $1/10.000$  is van dezelfde orde als  $10,2/100.000$ .

1	2	3	4	5	6	7	8	9	10	Gem.	F
8	10	6	10	11	10	16	11	7	13	10,2	Neen

Vervolgens onderzoeken onze epidemiologen de toestand vandaag, na de introductie van F op het eiland. De resultaten staan in de volgende tabel.

1	2	3	4	5	6	7	8	9	10	Gem.	F
18	17	20	6	15	12	23	14	21	14	16	Ja

Vergelijking van de resultaten geeft het volgende

1	2	3	4	5	6	7	8	9	10	Gem.	F
8	10	6	10	11	10	16	11	7	13	10,2	Neen
18	17	20	6	15	12	23	14	21	14	16	Ja
Verandering (in procenten)											
+125	+70	+233	-40	+36	+20	+44	+27	+200	+8	72,3	
Relatief risico van F voor V											
2,25	1,7	3,33	0,6	1,36	1,2	1,44	1,27	3,1	1,08	1,733	[1,5]

Het gemiddelde relatieve risico (1,733) komt dicht in de buurt van de [alleen aan ons bekende] werkelijke waarde van 1,5. Merk evenwel op dat per provincie de individuele cijfers sterk variëren, van 0,6 tot 3,33. Dat is niet verwonderlijk gezien de waargenomen veranderingen in de aantallen V-lijdens. Die veranderingen gaan van een daling met 40% tot een stijging met 233%.

Wij herinneren eraan dat de onderzoekers onafhankelijk van elkaar werken. Zij hebben ook de bedoeling hun eigen resultaten te publiceren in het kader van het debat over de relatie tussen F en V. Het is echter weinig waarschijnlijk dat onderzoeker 4 een wetenschappelijk tijdschrift vindt dat zijn onderzoek wil publiceren. Hij heeft namelijk een negatief resultaat gevonden: Met F is er volgens hem niets aan de hand, integendeel het onderzoek doet vermoeden dat F het relatieve risico van V met meer dan een derde verkleint.

Onderzoekers 2, 5, 6, 7, 8 en 10 zullen ook niet op veel belangstelling moeten rekenen. In vergelijking met wat zij te vertellen hebben, zijn de resultaten van onderzoekers 3 en 9, respectievelijk  $RR=3,33$  en  $RR=3,1$  opmerkelijk. De kans is groot dat alleen zij hun onderzoek in de gezaghebbende (dat wil zeggen veelgelezen, ook bij een ruimer publiek dan de specialisten bekende) nationale tijdschriften zullen kunnen publiceren. Welke resultaten gepubliceerd worden en welke niet is van groot belang voor ons gedachte-experiment. Het zijn immers alleen de gepubliceerde onderzoeken die deel uitmaken van 'de wetenschap'.

Wij moeten hier rekening houden met ten minste twee factoren:

- Publicatievoorselectie (*publication bias*): ook voor wetenschappelijke tijdschriften geldt dat zij moeten verkopen en dus een voorkeur hebben voor artikels die belangstelling wekken.

Zolang die belangstelling vooral een wetenschappelijke is, is er geen probleem. Wanneer echter criteria als 'maatschappelijke relevantie', 'politieke correctheid', 'eerbied voor de publieke opinie' in het spel komen dan is het einde zoek. Voor tijdschriften die van advertentie-inkomsten moeten leven of van door politieke autoriteiten verstrekte subsidies zijn dergelijke criteria evenwel vaak van groot gewicht. Meestal zijn mensen meer geïnteresseerd in een duidelijke positieve of negatieve correlatie dan in geen of bijna geen correlatie. Als F invloed heeft op V, dan wil men dat weten. Als dat niet het geval is, dan is er niets aan de hand. Waarom dan een tijdrovend en duur publicatieproces beginnen?

- Onder statistici geldt de regel dat relatieve risico's tussen 0,5 en 2 niet overtuigend zijn om er wetenschappelijke conclusies of beleidsimplicaties aan te verbinden.

Dit tweede punt verdient enige uitleg.  $RR=1$  betekent in ons voorbeeld dat er geen verschil is tussen het percentage V-lijdens na de introductie van de F-factor en het percentage vóór de introductie ervan. Een relatief risico van 2 betekent een verdubbeling van het percentage V-lijdens. Dat lijkt veel, maar komt in ons voorbeeld toch maar neer op een stijging van 100 V-lijdens naar 200 (op een bevolkingstotaal van een miljoen, waarvan 1 op tienduizend aan V lijdt). Het beschreven onderzoek heeft niet bewezen dat de 100 patiënten die er nu meer zijn 'slachtoffer' zijn van de F-factor. Voor zover de epidemiologen op het eiland weten kan de stijging geheel of gedeeltelijk of in het geheel niets met F te maken hebben. (Wij weten dat het relatieve risico van F 1,5 bedraagt, maar wij houden

die kennis voor ons.) Voor zover zij weten kunnen er andere factoren in het spel zijn. Of wellicht waren de medische archieven toch niet zo betrouwbaar zijn als zij voor hun onderzoek hebben aangenomen.

Een relatief risico van 0,5 betekent in ons voorbeeld een halvering van het percentage V-lijders in de aan F blootgestelde bevolking in vergelijking met vroeger. Dat is echter hetzelfde als zeggen dat vroeger het percentage dubbel was van wat het nu is.

Van 100 naar 200, of van 100 naar 50—de verandering is te klein om er veel gewicht aan te hechten. Dat is zeker het geval als er geen kennis uit andere bronnen aanwezig is, als het om onderzoeken op kleine steekproeven gaat, of om extrapolaties van bijvoorbeeld een klein aantal proefdieren dat in laboratoria aan hoge doses werd blootgesteld naar menselijke populaties die nooit met dergelijke doses geconfronteerd worden.<sup>16</sup>

Om die reden negeren wetenschappelijke epidemiologen meestal elk relatief risico tussen 0,5 en 2. Sommigen vinden zelfs die grenzen te laks en hanteren de norm 'Geen belang hechten aan relatieve risico's tussen 0,3 en 3'.

Hanteren wij een  $RR > 2$  norm, dan kunnen de onderzoekers 3 en 9 (althans op basis van dit criterium) hun onderzoek publiceren. Voor de buitenwereld (waaronder de algemene media als kranten, radio en televisie, en uiteraard ook de politieke milieus) bestaan alleen hun resultaten:

3	9	Gem.	F
6	7	6,5	Neen
20	21	20,5	Ja
Verandering (in %)			
+233	+200	+216	
Relatief risico van F voor V			
3,33	3,1	3,2	[1,5]

Merk op wat er met de gemiddelde waarde van het relatieve risico is gebeurd. Dat ligt helemaal niet meer in de buurt van de werkelijke waarde 1,5. Voor de media ligt het nu voor de hand te concluderen dat alle 'serieuze wetenschappers' (namelijk die welke publiceren) het er over eens zijn dat F de kans op V ten minste verdriedubbelt. Het is uiteraard mogelijk dat een van de andere onderzoekers een briefje aan de redactie schrijft. De kans is echter groot dat men hem naar het hoekje van de 'dissidenten' verwijst met de opmerking dat hij toch wel geen 'autoriteit' zal zijn aangezien hij niet over het onderwerp heeft gepubliceerd. Nochtans weten wij dat alle onderzoekers uitermate gewetensvol en competent zijn. Hun onderzoek is over de gehele lijn van dezelfde onberispelijke wetenschappelijke kwaliteit.

Het is duidelijk dat het hierboven beschreven selectieproces gemakkelijk aan de basis kan liggen van ongerustheid of paniek bij het publiek. Het vergt niet veel verbeeldingskracht om te beseffen dat op ons eiland politici, bureaucraten, lobbyisten, activisten en militanten in hun profileringsdrang en hun onverdroten jacht op stemmen, budgetverhogingen en fondsen niet zullen nalaten het gepubliceerde

<sup>16</sup> Niet zo lang geleden dook de term 'Frankenstein voedsel' op in de media. Aanleiding van de berichtgeving was een extrapolatie van de effecten van genetisch gemanipuleerd voedsel op basis van een laboratoriumproef met vijf ratten. Zie o.a. *The Bowditch Group Electronic AgBiotech Newsletter* (August 13, 1998).

‘wetenschappelijk onderzoek’ te gebruiken en te misbruiken voor hun eigen doelen.

*Tweede hypothese:* Laat ons nu veronderstellen dat er in werkelijkheid helemaal geen effect van F op V is. Dat wil zeggen,  $RR=1$ . Zal het wetenschappelijk onderzoek dat dan ook uitwijzen? Zal het publiek dat ook te weten komen? Dat is niet zeker.

Wij veronderstellen dat het eerste onderzoek (van de situatie voor de introductie van F) ongewijzigd blijft. Het onderzoek van de situatie na de introductie van F op het eiland levert nu echter deze resultaten op:

1	2	3	4	5	6	7	8	9	10	Gem.	F
18	7	16	6	14	2	3	6	17	2	10,3	Ja

Het gemiddelde (10,3) stemt goed overeen met de werkelijke waarde van 1 V-lijder per 10.000 inwoners. Wij krijgen dan

1	2	3	4	5	6	7	8	9	10	Gem.	F
8	10	6	10	11	10	16	11	7	13	10,2	Neen
18	7	16	6	14	2	3	6	17	2	10,3	Ja
Verandering (in procenten)											
+125	-30	+167	-33	+27	-80	-81	-46	+143	-85	10,7	
Relatief risico van F voor V											
2,25	0,7	2,67	0,66	1,27	0,2	0,19	0,54	2,43	0,15	1,1	[1]

Eens te meer ligt het gemiddelde relatieve risico ( $RR=1,1$ ) dicht in de buurt van de werkelijke waarde ( $RR=1$ ). Maar ook in dit geval is de variatie van de onderzoeken groot. Er zijn er vier die op een positieve correlatie ( $RR>1$ ) uitkomen en vijf die een negatieve correlatie ( $RR<1$ ) vinden. De gevonden relatieve risico's variëren van 0,15 tot 2,267. Hanteren wij de 0,5-2 norm voor mogelijke publicatie dan vallen de onderzoeken 2, 4, 5 en 8 al bij voorbaat weg:

1	3	6	7	9	10	Gem.	F
8	6	10	16	7	13	10,1	Neen
18	16	2	3	17	2	10,3	Ja
Verandering (in procenten)							
+125	+167	-80	-81	+143	-85	31,5	
Relatief risico van F voor V							
2,25	2,67	0,2	0,19	2,43	0,15	1,31	[1]

Noteer dat drie van de vier positieve correlaties de selectie overleven en dat twee van de negatieve correlaties uit de circulatie verdwijnen. Belangrijk is echter dat er nu een gemiddelde stijging van het aantal V-lijders met 31% geregistreerd wordt. Dat correspondeert met een relatief risico van 1,31—terwijl wij toch weten dat F niets bijdraagt tot het aantal gevallen van V!

Als nu bovendien de uitgevers van wetenschappelijke tijdschriften om welke

reden dan ook beslissen de negatieve correlaties niet te publiceren — bijvoorbeeld omdat zij het a priori onwaarschijnlijk vinden dat F een gunstig effect op V zou hebben — dan wordt de buitenwereld geconfronteerd met deze ‘wetenschappelijke bevindingen’:

1	3	9	Gem.	F
8	6	7	8	Neen
18	16	17	16,25	Ja
Verandering (in %)				
+125	+167	+143	145	
Relatief risico van F voor V				
2,25	2,67	2,43	2,45	[1]

Op basis van die gepubliceerde gegevens zou men kunnen besluiten dat F het risico van V met 145% doet stijgen (een relatief risico van 2,45). Voor de algemene media is dat doorgaans genoeg om paniek te zaaien, vooral als V een ernstige aandoening is. Volgens de definitie van ons model is F in werkelijkheid evenwel geen risicofactor voor V.

Alleen wie toegang heeft tot al het onderzoek — het gepubliceerde en het niet-gepubliceerde — is in staat te ontdekken dat er met F helemaal niets aan de hand is. Hij zou kunnen zien dat het gemiddelde aantal V-lijdens per provincie praktisch onveranderd is gebleven: 10,2 per honderdduizend inwoners voor de introductie van F en 10,3 na de introductie ervan. Wij zien dat onze onderzoekers, hoe gewetensvol en competent zij ook zijn, en het publicatieproces onwerkelijke resultaten kunnen produceren.

Let wel, die vreemde resultaten treden hier op ondanks het feit dat wij verondersteld hebben dat er geen reden was om te twijfelen aan de kwaliteit en de representativiteit van de gegevens waarmee de onderzoekers aan de slag gingen. In de werkelijke wereld is daar vaak reden genoeg voor. Niet zelden wordt onderzoek gedaan niet op het geheel van een populatie maar via steekproeven op een klein deel ervan. Het gebeurt ook dat men een van de te onderzoeken fenomenen of factoren niet rechtstreeks kan tellen of meten. Dan gaat men andere dingen tellen waarvan men weet, beweert of vermoedt dat zij wijzen op de aanwezigheid van de te onderzoeken factor. Zodra men echter dergelijke beweringen of vermoedens introduceert, worden de resultaten uiteraard heel twijfelachtig. Vermoedens en beweringen zijn geen gegevens, zelfs geen weinig betrouwbare gegevens.

Samenvattend:

- Ons voorbeeld laat zien dat zelfs onberispelijk uitgevoerd epidemiologisch onderzoek aanleiding kan geven tot verkeerde en sterk uiteenlopende percepties van het werkelijke relatieve risico.
- Publicatievoorselectie kan zelfs al op het vlak van de wetenschappelijke tijdschriften een opmerkelijk effect hebben. Het effect in de algemene media is vaak nog veel groter.

## Het zogeheten ‘voorzorgsbeginsel’

Vaak wordt het gebruik van ondeugdelijke of weinig betrouwbare statistieken goedgepraat met een verwijzing naar het gezegde ‘voorkomen is beter dan genezen’ (het ‘voorzorgsbeginsel’, Engels ‘precautionary principle’). Voor ons voorbeeld leidt dat tot de argumentatie:

- Er zijn aanwijzingen dat F een risicofactor voor V is.
- Wij weten niet zeker dat F een oorzaak is van V.
- Als F echter een oorzaak van V zou zijn, dan is F een ernstig gevaar.
- ‘Voorkomen is beter dan genezen.’
- Het gebruik van F moet nu al gereguleerd / verboden worden.

Dat is duidelijk een drogreden—en niet alleen omdat de redenering logisch niet klopt. In ons voorbeeld werd alleen de relatie tussen F en V onderzocht. Onderzoek over enig verband tussen F en andere verschijnselen was er niet. Zelfs als F een risicofactor voor V zou zijn dan is toch niet uit te sluiten dat F ook gunstige effecten heeft (bijvoorbeeld de incidentie van andere ziektes vermindert<sup>17</sup>). Het is niet evident dat het beter is V te voorkomen dan die andere ziektes. Voor westerlingen, die weinig last hebben van malaria, kan het een goed idee lijken om verontreiniging van moerassen in tropische gebieden te voorkomen door het gebruik van DDT te verbieden. Of de lokale bevolking het daarmee eens is, is een andere vraag.<sup>18</sup>

Bovendien zijn er veel gevallen waarin ‘voorkomen’ *niet* beter is dan ‘genezen’ (of investeren in een remedie)—bijvoorbeeld omdat het veel duurder is. Een ongevallen- en aansprakelijkheidsverzekering nemen is voor de meeste mensen beduidend rationeler dan thuis in een zetel blijven zitten om ‘ongevallen te voorkomen’.

Het zogeheten ‘voorzorgsbeginsel’ is nu in het EU-verdrag ingeschreven.<sup>19</sup> Het is echter niet meer dan een vuistregel. De Europese Commissie houdt zich dan ook aanbevolen om daaromtrent ‘richtlijnen’ te formuleren. Uitgangspunt is uiteraard dat het ‘principe’ in de praktijk veel ruimer geïnterpreteerd dient te worden dan de tekst van het verdrag aanduidt:

The precautionary principle is not defined in the Treaty, which prescribes it only once - to protect the environment. But *in practice*, its scope is much wider, and specifically where preliminary objective scientific evaluation indicates that there are reasonable grounds for concern that the potentially dangerous effects on the *environment, human, animal or plant*

---

<sup>17</sup> Roken (een positieve risicofactor voor longaandoeningen) is een *negatieve* risicofactor voor de ziekte van Alzheimer en andere aandoeningen van het zenuwstelsel. Graves AB, van Duijn CM et al., ‘Alcohol and tobacco consumption as risk factors for Alzheimer's disease’ *Int. Journal Epidemiology*, 20, 1991; (2 Suppl 2) p.S48-S57

<sup>18</sup> L. Mooney, & R. Bate (eds), *Environmental Health: Third World Problems—First World Preoccupations* (Butterworth, London, 1999); [Washtimes.com/world/20020616-11558965.htm](http://Washtimes.com/world/20020616-11558965.htm) (16 juni 2002)

<sup>19</sup> Article 130r(2) of 1997 Amsterdam Treaty (97/C 340/01) OJ C 340 of 1997-11-10.

*health* may be inconsistent with the high level of protection chosen for the Community.<sup>20</sup>

Wat valt buiten het bereik van op ‘*voorlopige evaluaties*’ gebaseerde ‘*bezorgdheid*’ [van wie?] over ‘*potentieel*’ gevaarlijke effecten die ‘*mogelijk*’ onverenigbaar zijn met een gekozen [door wie?] ‘*hoge*’ graad van bescherming [tegen wat?] van de gezondheid van het milieu, mensen, dieren en planten? En dat zijn dan nog maar de criteria die ‘*in het bijzonder*’ (‘specifically’) gelden! In de interpretatie van de Commissie is het ‘voorzorgsbeginsel’ een vrijbrief voor willekeur in het politieke en administratieve gebruik van statistieken. Voor de ‘rechtvaardiging’ van totalitaire bevoegdheden doet het niet onder voor de beruchte definitie die de WGO aan het begrip ‘ziekte’ (en dus ook ‘ziektebestrijding’) gaf:

Health is a state of *complete physical, mental and social well-being* and not merely the absence of disease or infirmity. The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition.<sup>21</sup>

Wat moeten wij ons bij ‘de hoogste bereikbare standaard van compleet welzijn’ voorstellen? Is de laagste bereikbare standaard van *compleet welzijn* niet genoeg? Het gebruik van nonsensicale retoriek in constitutionele documenten om bevoegdheden ‘af te bakenen’ is ongetwijfeld een bron van veel mentale en sociale spanningen en dus slecht voor ‘de gezondheid’. De bestrijding ervan door de WGO zou niet misplaatst zijn.

## Statistische significantie en betrouwbaarheidsintervallen

Vaak staat bij statistische correlaties een zogeheten ‘p-waarde’ aangegeven. Die waarde wordt gebruikt om te zeggen of de correlatie ‘statistisch significant’ is of niet. De conventionele norm is dat p niet groter mag zijn dan 0,05. Is p groter dan is de correlatie statistisch niet betekenisvol.

De p-waarde is een wiskundig berekende maat — hoe die berekend wordt, valt buiten ons bestek. Zij geeft aan hoe groot *de waarschijnlijkheid is dat de gevonden correlatie louter aan het toeval te wijten is*. Een p-waarde van 0,05 geeft dus aan dat de correlatie waarop die waarde betrekking heeft, met een waarschijnlijkheid van ten hoogste 5% wel aan het toeval te wijten is. Een p-waarde van 0,1 zegt dus dat de correlatie met een waarschijnlijkheid van slechts 90% *niet* aan het toeval te wijten is. De 0,05 norm voor p-waarden is een conventie. Het is voor wetenschappelijke doeleinden een vrij lakse norm.

Vaak worden ook betrouwbaarheidsintervallen (*confidence intervals*) aangegeven. Ook dat zijn louter wiskundig berekende eigenschappen van de gevonden correlaties tussen cijferreeksen. Een betrouwbaarheidsinterval is een interval bepaald door een getallenpaar (x,y), waarin x de benedengrens en y de bovengrens

---

<sup>20</sup> ‘Communication from the Commission on the Precautionary Principle’ van 2 februari 2000.

<sup>21</sup> Constitution of the World Health Organisation (Introduction). [Nadruk toegevoegd.]

van het interval is—bijvoorbeeld 0,1-1,5. Het zegt dat de ‘ware’ waarde van het relatieve risico met een bepaalde waarschijnlijkheid in het opgegeven interval ligt. Zo geeft een 95% betrouwbaarheidsinterval aan dat de ‘ware’ waarde van het relatieve risico zich met een waarschijnlijkheid van 95% in het opgegeven interval bevindt.

Het staat uiteraard niet goed als een statistisch onderzoek een relatief risico van 0,2 vindt terwijl het 95% betrouwbaarheidsinterval voor de onderzochte correlatie 0,3-4,8 bedraagt. Het gevonden risico ligt buiten het betrouwbaarheidsinterval.

Het staat evenmin goed als een onderzoek een ‘positief’ relatief risico vindt ( $RR > 1$ ) en het 95% betrouwbaarheidsinterval een benedengrens heeft die kleiner is dan of gelijk aan 1. Bijvoorbeeld  $RR = 1,3$  (95% BI 0,9-1,7). Wij hebben dan zelfs geen zekerheid dat er inderdaad een positief relatief risico is.

Hetzelfde geldt als een onderzoek een negatief relatief risico vindt ( $RR < 1$ ) en het 95% betrouwbaarheidsinterval een bovengrens heeft die groter is dan of gelijk aan 1. Bijvoorbeeld  $RR = 0,9$  (95% BI 0,5-1,3). Wij hebben dan zelfs geen zekerheid dat er inderdaad een negatief relatief risico is.

Hoe smaller het betrouwbaarheidsinterval hoe beter het statistisch resultaat.  $RR = 0,66$  (95% BI 0,55-0,77) is heel behoorlijk. Een breed interval wijst op gebreken, bijvoorbeeld het gebruik van ‘onzuivere’ of rommelige gegevens. Neem  $RR = 30$  (95% BI 5-180). In dit geval kan het ‘ware’ relatieve risico zes maal groter of kleiner zijn dan het gevonden risico. Dat is een wel erg brede marge.

Statistische significantie is geen kwaliteitsmerk. Zij is een louter wiskundige maatstaf die niets zegt over de wetenschappelijke kwaliteit van het gedane onderzoek of de gebruikte gegevens. Zij is evenmin een aanwijzing van het ‘wetenschappelijk belang’ van het onderzoek. Sommige epidemiologen passen de techniek van het ‘baggeren naar correlaties’ toe. Dat wil zeggen, zij combineren blindweg gegevens tot zij een statistisch significante correlatie vinden. Veronderstel dat wij veel gegevens hebben over eetgewoonten en aandoeningen en ziektes van mensen in Limburg. Rekening houdend met het grote aantal items in het voedsel en met het de complexiteit van het menselijke lichaam kunnen wij een massa verbanden onderzoeken: citrusvruchten / bloedarmoede, citrusvruchten / hoofdpijn, erwtensoep / hoofdpijn, erwtensoep / leverkanker, enzovoort. Met moderne computers is het mogelijk al die verbanden te onderzoeken. Er is een grote kans dat een aantal van die correlaties (binnen het bestand van de beschikbare gegevens) in hoge mate ‘statistisch significant’ is. De kans dat een daarvan enig wetenschappelijk belang heeft is echter uiterst klein.

P-waarden en betrouwbaarheidsintervallen hebben weliswaar geen wetenschappelijke waarde—zij zijn niet meer dan conventionele formele maatstaven—maar dat betekent niet dat men ze zonder meer kan negeren. Veel onderzoek presenteert statistieken die zelfs niet aan die formele eisen voldoen. Daar zit heel wat ‘junk science’ en pseudo-wetenschap tussen.

Het is een goede strategie om onderzoekresultaten die niet aan die voorwaarden voldoen zonder meer te negeren of met meer dan een korrel zout te nemen. Dat

was de strategie van rechter William Osteen in een zaak tegen het *Environment Protection Agency*.<sup>22</sup> Het ging om een betwisting van de door het EPA gebezigde criteria om 'passief roken' als belangrijke risicofactor te beschouwen. De rechter liet geen spaander heel van de verdediging van het EPA. In zijn motivering schreef hij onder andere:

The record and EPA's explanations to the court make it clear that using standard methodology, EPA could not produce statistically significant results with its selected studies. Analysis conducted with a .05 significance level and 95% confidence level included relative risks of 1. Accordingly, these results did not confirm EPA's controversial a priori hypothesis. In order to confirm its hypothesis, EPA maintained its standard significant level but lowered the confidence level to 90%. This allowed EPA to confirm its hypothesis by finding a relative risk of 1.19, albeit a very weak association.

De klagers kregen gelijk. Als echter de wetgever zelf zijn wetten op ondeugdelijke statistische gegevens en methoden baseert, dan staan de rechtzoekenden daar weerloos tegenover. Onzin, incompetentie of gesjoemel? Als het maar gewettigd is! 'Numbers put into an Act of Parliament become enshrined as icons.'<sup>23</sup>

---

Frank van Dun

---

<sup>22</sup> *Flue-Cured Tobacco e.a. v. U.S. Environmental Protection Agency*, in The U.S. District Court for the Middle District of North Carolina (Winston-Salem Division), July 17, 1998, W. Osteen J. (p.77)

<sup>23</sup> John Brignell, *Sorry, Wrong Number. The Abuse of Measurement* (Brignell Associates / European Science and Environment Forum, s.l., 2000), p.140.