

Een van de domeinen van PISA is 'onzekerheid'. Vragen in dat domein gaan onder andere over de interpretatie van getallen in de kansrekening. Dat de interpretatie van '1 op de 342 miljoen' over zeer serieuze zaken kan gaan laten **Michiel van Lambalgen** en **Ronald Meester** zien in hun analyse van het toeval in het geruchtmakend proces tegen Lucy de B.

Wat zeggen al die getallen eigenlijk?

De statistiek in het proces tegen Lucy de B.

Inleiding

In de rechtszaken tegen de van meervoudige moord verdachte verpleegster Lucy de B. speelde statistiek een belangrijke rol. De verdachte had tijdens haar diensten als verpleegster te maken gekregen met een opvallend groot aantal incidenten, en de vraag leek gerechtvaardigd of een dergelijk groot aantal incidenten eigenlijk wel in alle redelijkheid 'toevallig' kan plaatsvinden tijdens de diensten van een enkele verpleegkundige.

De statistiek lijkt de aangewezen discipline te zijn om uit te zoeken of we de gebeurtenissen redelijkerwijs aan toeval kunnen toeschrijven. Het is echter van groot belang om een goed beeld te hebben van wat statistiek wel en niet vermag. Vooropgesteld dient te worden dat statistiek naar de aard der zaak geen onwrikbare getallen en kansen kan opleveren. Uit het overvloedig gebruik van de wiskunde in de statistiek zou men kunnen afleiden dat statistiek leidt tot onaantastbare resultaten. Dit beeld is echter geheel onjuist; de activiteit van een statisticus behelst essentieel meer dan wiskunde. Onder meer hoort daarbij het opzetten van een wiskundig model en het interpreteren van de getallen die berekend zijn op basis van dat model. Beide activiteiten zijn essentieel niet wiskundig, en laten ruimte voor andere keuzes. In dit licht zijn resultaten die voortvloeien uit statistische berekeningen altijd relatief, in twee opzichten: statistici kunnen verschillen in hun modelkeuze, en bovendien ook nog in de methoden die zij acceptabel achten. Het is naar onze mening van groot maatschappelijk belang dat rechters zich van deze tweevoudige subjectiviteit van de statistiek bewust zijn. Als de zaak tegen Lucy de B. dit inzicht zal doen doorbreken, dan heeft zij toch een functie gehad.

In de eerste rechtszaak tegen Lucy de B. is naar onze mening zowel bij de statistische berekeningen zelf, als bij de interpretatie van de uitkomsten, heel veel misgegaan. Enerzijds zijn de gebruikte modellen en de daarbinnen gemaakte berekeningen naar onze mening ondeugdelijk, waardoor de getallen vrijwel betekenisloos zijn. Anderzijds was de interpretatie van de getallen door de rechtbank in haar vonnis van 24 maart 2003 ook aantoonbaar

incorrect. Laten we nu eerst een korte geschiedenis van de statistische kant van de zaak tegen Lucy de B. geven. Bij het oorspronkelijke proces en het daarop volgende hoger beroep heeft het rapport van statisticus prof. dr. Henk Elffers een belangrijke rol gespeeld. In zijn rapportage kwam Elffers tot de conclusie dat, wanneer we:

1. aannemen dat de verdachte dezelfde kans op een incident tijdens een dienst heeft als iedere andere verpleegster;
2. aannemen dat er geen samenhang bestaat tussen incidenten tijdens verschillende diensten;

de kans dat zij minstens het aantal incidenten meemaakt als feitelijk gebeurd is, kleiner is dan 1 op de 342 miljoen. Deze kans is zo extreem klein, aldus Elffers, dat we toeval met een gerust hart kunnen uitsluiten. In haar vonnis van 24 maart 2003 nam de rechtbank het getal van Elffers over, maar uit de motivering bij het vonnis blijkt herhaaldelijk dat de rechters de betekenis van het getal 1 op de 342 miljoen niet begrepen hebben. We geven een drietal voorbeelden. Onder punt 7 van de bewijsoverwegingen schrijft de rechtbank het volgende:

In zijn rapport van 29 mei 2002 komt prof. dr. H. Elffers tot de conclusie dat de kans dat een verpleegkundige bij toeval net zoveel incidenten [...] zou meemaken als verdachte 1 op de 342 miljoen bedraagt.

Onder punt 8 vinden we:

- Uit zijn rapport van 29 mei 2002 blijkt verder dat prof. dr. H. Elffers tevens de navolgende deelkansen heeft berekend:
- a. De kans dat bij toeval 1 van de 27 verpleegkundigen bij 142 uit 1029 diensten betrokken zou zijn bij acht incidenten ... is kleiner dan 1 op 300.000.
 - b. De kans dat verdachte bij toeval ...

Onder punt 11 vinden we ten slotte:

De rechtbank is van oordeel dat uit de door prof. dr. H. Elffers uitgevoerde waarschijnlijkheidsberekeningen, zoals neergelegd in zijn rapport van 29 mei 2002, volgt dat het uitermate onwaarschijnlijk moet worden geacht dat de verdachte de in de tenlastelegging genoemde incidenten bij toeval heeft meegemaakt.

Deze berekeningen geven derhalve aan dat het in hoge mate waarschijnlijk is dat er een verband bestaat tussen het werkzaam zijn van de verdachte en het zich voordoen van bedoelde incidenten.

We citeren zo uitgebreid uit de bewijsoverwegingen om te laten zien dat de rechtbank systematisch formuleringen hanteert zoals ‘kans op toeval’, die principieel moeilijk te verdedigen zijn, en die helemaal niet uit Elffers berekeningen volgen. Elffers beoogt de kans te berekenen dat iets gebeurt als we aannemen dat incidenten gewoon per toeval en zonder samenhang plaatsvinden. De rechtbank meent echter ten onrechte dat Elffers juist de kans op toeval uitrekt. Deze fout wordt vaak gemaakt, en staat bekend als de ‘prosecutor’s fallacy’. Dit is niet alleen maar een kwestie van verkeerde woorden, zoals Elffers ter zitting van het hof wilde doen geloven. De interpretatie van de rechtbank is fundamenteel onjuist en suggereert een veel sterkere bewijskracht van de getallen dan deze werkelijk hebben, waarover straks meer. Bovendien is de conclusie die de rechtbank onder punt 11 trekt (‘deze berekeningen geven derhalve aan ...’) ook nog eens fout, zelfs als men de eerdergenoemde foute interpretatie van de getallen voor lief neemt.

Nadat de verdachte op 24 maart 2003 schuldig bevonden werd, diende de afgelopen maanden het hoger beroep. Wij zijn in dat hoger beroep gezamenlijk gehoord als getuige-deskundigen, opgeroepen door de raadslieden van de verdediging. Wij hebben tijdens ons verhoor geprobeerd uit te leggen waarom er kritiek mogelijk is op het rapport van Elffers, en waarom zijn getal 1 op de 342 miljoen weinig betekenis heeft. In dit artikel zullen wij uitleggen waarom wij dat vinden.

Het artikel is als volgt opgebouwd. In de volgende paragraaf geven we de statistische methode van Elffers weer. Vervolgens zullen we uitleggen waarom deze methode naar onze mening aan kritiek onderhevig is. Wij hebben hier verschillende argumenten voor, die we per paragraaf zullen behandelen. We zullen zowel concrete als meer filosofische bezwaren bespreken. In verband met de betekenis van de uitspraak ‘kans op toeval’ zullen we ook kort ingaan op de zogenaamde Bayesiaanse aanpak van dr. Aart de Vos, zoals hij deze heeft uiteengezet in artikelen in *Trouw*, *NRC Handelsblad* en *NJB*. Ten slotte gaan we in op het probleem van de communicatie tussen de statisticus en de rechterlijke macht, voortvloeiend uit het bijzondere karakter van de statistiek.

De methode van Elffers

We beginnen met de gegevens die we nodig hebben, zoals weergegeven in het rapport van Elffers. De verdachte werkte in twee ziekenhuizen, het Juliana Kinderziekenhuis (JKZ) en het Rode Kruisziekenhuis (RKZ). Binnen het RKZ werkte de verdachte op twee afdelingen, genummerd 41 en 42. Binnen het tijdsinterval van de statistische analyse, werden er op de betreffende afdeling van het JKZ in totaal 1029 diensten gedraaid, waarvan de verdachte er 142 voor haar rekening nam. Tijdens die 1029 diensten vonden er acht zogenaamde incidenten plaats, die alle acht tijdens een dienst van de verdachte plaatsvonden. Op de afdeling werkten er in die tijd 27 verpleegkundigen.

Gedurende dezelfde tijdsperiode werden er in het RKZ-42 een totaal van 339 diensten gedraaid, waarvan de verdachte er 58 voor haar rekening nam. Tijdens die 339 diensten vonden er veertien incidenten plaats, waarvan er vijf tijdens een dienst van de verdachte plaatsvonden.

Ten slotte werden er gedurende de bedoelde tijdsperiode op RKZ-41 een totaal van 366 diensten gedraaid, waarvan de verdachte er slechts een voor haar rekening nam. Tijdens die 366 diensten vonden er vijf incidenten plaats, waarvan er één plaatsvond tijdens die enige dienst van de verdachte.

De aandacht viel op de verdachte door het frequent optreden van incidenten tijdens haar diensten op het JKZ. Daarna is ook onderzoek gedaan naar haar diensten in het RKZ met bovenstaande getallen als resultaat. De vraag die op basis van deze gegevens aan Elffers gesteld werd is: kan dit allemaal toeval zijn? De verdachte maakt immers ogenschijnlijk zeer veel van de gemelde incidenten mee, wellicht meer dan je redelijkerwijs aan het toeval zou kunnen toeschrijven. Om deze vraag te kunnen beantwoorden stelde Elffers een heel eenvoudig wiskundig model op, dat er als volgt uitziet.

We beperken ons eerst tot het JKZ. We weten dat er tijdens de 1029 diensten acht incidenten zijn geweest. Als de kans op een incident voor iedere verpleegkundige altijd hetzelfde is, en incidenten in verschillende diensten niets met elkaar te maken hebben, dan kunnen we die acht incidenten beschouwen als een willekeurige greep van acht diensten uit het totaal van 1029 diensten. Met andere woorden, de diensten waarin een incident optrad zijn voor te stellen als een volstrekt toevallige ‘trekking’ van acht diensten uit het totaal van 1029 diensten. Wanneer we acht diensten toevallig trekken uit het totaal van 1029 diensten, dan kunnen we vrij eenvoudig uitrekenen wat de kans is dat alle acht getrokken diensten een dienst van de verdachte waren. Deze kans blijkt ongeveer 1 op de 9 miljoen te zijn.

Er zit hier echter een addertje onder het gras, aldus Elffers. Wat hier in feite uitgerekend wordt, is de kans dat een van tevoren aangewezen verpleegkundige (namelijk de verdachte) die 142 van de 1029 diensten voor haar rekening neemt, alle acht incidenten meemaakt. Maar dat is niet helemaal fair; ze is alleen maar aan deze berekening onderworpen omdat ze was opgevallen vanwege het frequent voorkomen van incidenten tijdens haar diensten. Daarom is het eerlijker, aldus Elffers, om de kans uit te rekenen dat iemand op de afdeling acht incidenten meemaakt. Aangezien er 27 verpleegkundigen op de afdeling zijn, moet de kans van 1 op 9 miljoen met 27 vermenigvuldigd worden, resulterend in een kans kleiner dan 1 op 300.000, aldus Elffers. Deze vermenigvuldiging met 27 wordt de post hoc correctie genoemd.

Op de twee afdelingen van het RKZ gaat Elffers in eerste instantie analoog te werk, uiteraard met de bijbehorende getallen. Het enige verschil, aldus Elffers, is dat nu de post hoc correctie niet meer nodig is, omdat de verdachte van tevoren al geïdentificeerd was via de gegevens van

het JKZ. Op deze manier verkrijgt Elffers drie kansen: de 1 op 300.000 van het JKZ, en de kansen die betrekking hebben op de afdelingen 41 en 42 van het RKZ. De uiteindelijke kans van 1 op 342 miljoen verkrijgt Elffers door deze drie kansen met elkaar te vermenigvuldigen.

Misschien valt u al iets bijzonders op: om tot die kans van 1 op 342 miljoen te komen, hoeven we kennelijk niet te weten hoe groot de kans is dat een verpleegkundige een incident zal meemaken. Elffers beschouwt dat als een voordeel van zijn methode; in zijn rapport, en ook ter zitting, heeft hij meerdere malen met klem beweerd dat we deze kans niet op een betrouwbare manier kunnen schatten. In plaats daarvan wil hij alleen uitgaan van de 'harde' dienstroosterdata. We zullen straks laten zien dat we zonder enige kennis van deze kans niet kunnen controleren of het model van Elffers wel toepasbaar is, zodat Elffers met deze uitspraken zijn eigen methode ondergraaft.

Op het eerste gezicht lijkt de methode van Elffers eigenlijk helemaal zo gek nog niet. Echter, na enig nadenken beginnen er toch twijfels op te komen. In de volgende paragrafen behandelen we een aantal serieuze problemen van zijn methode.

Op welk niveau moeten we het probleem bekijken?

Elffers beperkt de statistische analyse tot de afdelingen waarop de verdachte werkte. Het is uiterst belangrijk om je af te vragen waarom hij dat eigenlijk doet. Immers, de vraag waar het de rechtbank om te doen is, is of de verdachte zich tijdens haar diensten schuldig heeft gemaakt aan strafbare feiten. In die vraag komt het woord 'afdeling' niet voor. Dat betekent dat Elffers er zelf voor gekozen heeft om slechts de afdeling te beschouwen.

We zullen onze kritiek op de vanzelfsprekendheid van deze keuze toelichten aan de hand van een voorbeeld. Stel, er wordt een loterij uitgeschreven met loten genummerd 1 tot en met 1.000.000. De hoofdprijs blijkt op nummer 223.478 te vallen, en dat lot blijkt gekocht door een inwoner van de Wasstraat in Leiden. Gegeven deze feiten kunnen we nu gaan kijken naar de kans dat iemand in deze straat de hoofdprijs wint. We kunnen een simpel en oncontroversieel model maken waaruit blijkt dat de kans dat iemand in de Wasstraat in Leiden de hoofdprijs wint buitengewoon klein is. Moeten we hieruit concluderen dat de loterij niet eerlijk was? Immers, er is iets gebeurd dat bij een eerlijke loterij een heel kleine kans had. Nee, natuurlijk kunnen we dat niet concluderen. Punt is dat de aandacht pas op de Wasstraat in Leiden werd gevestigd, nadat de trekking geweest was. We hebben pas een model voor de Wasstraat in Leiden gemaakt nadat we gezien hadden dat zich daar iets opvallends had afgespeeld.

Dit fenomeen is berucht in de statistiek. Als men iets opvallends waarneemt, dan is men geneigd om het model en bijbehorende hypothesen aan de data aan te passen; in zulke gevallen wordt de hypothese dus opgesteld nadat we de data hadden gezien. Het probleem van deze dataafhankelijke hypothesen is dat men eerst op grond van de data een idee krijgt voor een hypothese, die men dan ver-

volgens door diezelfde data wil verifiëren. Het is intuïtief duidelijk dat hier iets misgaat, maar binnen de statistiek bestaat er geen volledig bevredigende oplossing voor dit probleem. Kern van de zaak is dat je er rekening mee moet houden dat het van tevoren uitermate onwaarschijnlijk was dat de Wasstraat in Leiden überhaupt in beeld zou komen. Als we de kans gaan uitrekenen dat iemand in Leiden, of iemand in Zuid-Holland de hoofdprijs wint, dan komen daar geheel verschillende getallen uit, en het is belangrijk je te realiseren dat er niet een uniek of voor de hand liggend niveau is waarop je het model moet bekijken. Elk getal is dus relatief aan de keuze voor de schaal van het model.

De (enigszins wrange) analogie met de zaak Lucy de B. is hopelijk duidelijk: de winnaar van de hoofdprijs staat voor de verdachte, en de straat speelt de rol van de afdeling. Elffers beperkt zijn model tot de afdeling waar zich iets opvallends heeft afgespeeld, maar hij vergeet dan dat het feit dat het juist deze afdeling betreft van tevoren uitermate onwaarschijnlijk was. Hij heeft naar de data gekeken, heeft gezien waar zich iets opvallends heeft afgespeeld, en vervolgens uitgerekend dat de kans hierop wel erg klein was. Een andere statisticus zou met net zoveel recht of onrecht in plaats van de afdeling bijvoorbeeld het hele JKZ in zijn of haar analyse mee willen nemen. Onze kritiek is hier dus dat Elffers een willekeurige keuze heeft gemaakt. Wanneer het hele JKZ in de beschouwing meegenomen zou worden, dan zou de methode waarschijnlijk een veel grotere kans opleveren dan de astronomische 1 op 342 miljoen van Elffers. Immers, de post hoc correctie van Elffers dicteert dat we met het aantal relevante verpleegkundigen moeten vermenigvuldigen, en dat zijn er op het hele JKZ veel meer. Op deze manier zouden we de kans uitrekenen dat gegeven alle incidenten iemand op het JKZ zoiets mee zou maken, en die kans is veel groter. Ook kan de kans uitgerekend worden dat een verpleegkundige in Den Haag, in Nederland, Europa enzovoort dit meemaakt – met navenant grotere kansen.

Enkele gedachtenexperimenten

We hebben in de vorige paragraaf gezien dat het niveau van het model van Elffers arbitrair is; er is geen enkele reden om de analyse te beperken tot de afdeling waar de verdachte werkte. Maar als men eenmaal een model gekozen heeft, dan zou het rekenen binnen dat gekozen model niet controversieel moeten zijn. Maar de berekening van Elffers is dat wel. Met de vermenigvuldiging met 27, de post hoc correctie, maakt hij een fout die het antwoord nadelig voor de verdachte beïnvloedt. We kunnen dat als volgt inzien.

Elffers zegt zich in zijn berekening alleen te willen baseren op dienstroosterdata, dat wil zeggen, de diensten van verdachte en andere verpleegkundigen, en de opgetreden incidenten. Het enige dat telt is het aantal diensten dat door de verdachte gedraaid wordt, samen met het totaal aantal gedraaide diensten. Het aantal verpleegkundigen op de afdeling zou dus totaal irrelevant moeten zijn. Im-

mers, de vraag die Elffers gesteld werd is of er reden is om te denken dat zich tijdens de diensten van de verdachte iets opvallends heeft voorgedaan. De enige gegevens die daarvoor nodig zijn, bestaan uit gegevens omtrent de incidenten die zich tijdens, of juist buiten de diensten van de verdachte hebben afgespeeld; het aantal verpleegkundigen waarover de resterende diensten verdeeld worden is totaal irrelevant.

We voeren nu een klein gedachtenexperiment uit. Als de diensten buiten de verdachte niet onder 26 resterende verpleegkundigen verdeeld zouden zijn, maar onder 99 (we noemen maar een willekeurig getal), dan zouden alle dienstroosterdata net zo belastend of ontlastend zijn voor de verdachte. Maar als er 100 verpleegkundigen zouden zijn, dan zou de post hoc correctie van Elffers een vermenigvuldiging met 100, en niet met 27 moeten zijn. Natuurlijk, in de realiteit waren er 27 verpleegkundigen en geen 100, maar dit gedachtenexperiment maakt duidelijk dat de feitelijkheid van de 27 verpleegkundigen geen enkele rol kan en mag spelen in de feitelijke berekening.

Een tweede gedachtenexperiment laat zien dat de volgorde waarin de ziekenhuizen in de berekening worden verwerkt van belang is. Als de verdachte op het RKZ als eerste opgevallen zou zijn, dan zou Elffers volgens zijn methode een post hoc correctie hebben toegepast door te vermenigvuldigen met het aantal verpleegkundigen op de desbetreffende afdeling. Dat waren er vast geen 27, waardoor het getal er weer totaal anders zou hebben uitgezien. Hier zien we duidelijk de invloed van een subjectieve keuze, die niets met de gegevens uitstaande heeft.

Een laatste gedachtenexperiment geeft aan dat ook de spreiding van de diensten over de ziekenhuizen van belang is. Als de verdachte al haar diensten op één afdeling zou hebben gedraaid (in plaats van op drie), waarbij het totaal aantal diensten, het aantal door de verdachte gedraaide diensten, het totaal aantal incidenten en het aantal incidenten tijdens een dienst van de verdachte gelijk zouden zijn aan de reële casus, dan zou er een totaal andere kans uit komen, een veel grotere kans wel te verstaan. Dus het feit dat de verdachte haar diensten spreidt over meerdere afdelingen, maakt haar volgens deze berekening meer verdacht.

Al deze voorbeelden laten zien dat de door Elffers gekozen berekeningswijze zich zeker niet uitsluitend baseert op dienstroosterdata, zoals de rechtbank in haar bewijsmotivering vermeldt, maar dat subjectieve keuzes in de omgang met deze data ook een grote rol spelen.

Is het model eigenlijk wel een goede beschrijving van toeval?

De oorspronkelijke vraag aan Elffers was of het waargenomen aantal incidenten toeval zou kunnen zijn. Nu is het begrip 'toeval' niet eenduidig; het kan op vele manieren geïnterpreteerd worden. Om er als wiskundige naar te kunnen

kijken heeft Elffers de vraag of het toeval kan zijn op zijn eigen manier vertaald naar een wiskundige context. Hij kan ook niet anders, maar het is belangrijk om te begrijpen dat die vertaling allerlei individuele keuzes en subjectieve afwegingen met zich meebrengt. Laten we er één noemen. Elffers vertaalt toeval met de wiskundige veronderstelling dat elke verpleegkundige altijd dezelfde kans op een incident heeft. Maar zoals we gezien hebben is dat helemaal niet vanzelfsprekend. Men zou ook kunnen poneren dat de kans op een incident, gegeven dat verdachte aanwezig is, in alle omstandigheden gelijk is aan de kans gegeven dat ze niet aanwezig is. Deze wiskundige formulering van toeval laat bijvoorbeeld de mogelijkheid open dat er verschil is tussen dag- en nachtdiensten, of de mogelijkheid dat verdachte zelf de ernstig zieke patiënten zou opzoeken. In het laatste geval is de aanname van Elffers onjuist, terwijl er gewoon nog sprake van toeval kan zijn, want ernstig zieke patiënten overlijden altijd vaker, wellicht ongeacht de aanwezigheid van de verdachte.

In dit verband is het goed te wijzen op de getuigenis van prof. mr. R.V. de Mulder. Onder punt 9 van de bewijsoverwegingen lezen wij dat deze deskundige de methode van Elffers correct acht. Bovendien kunnen we onder punt 13 van de bewijsoverwegingen lezen dat volgens De Mulder de mogelijkheid dat de verdachte vaker relatief ernstig zieke patiënten verzorgde, de uitkomsten van de waarschijnlijkheidsberekeningen niet wezenlijk zou beïnvloeden. Deze uitspraak van De Mulder is ons onbegrijpelijk: iemand die doodzieke patiënten verzorgt zal uiteraard meer sterfgevallen of andere incidenten meemaken dan iemand die alleen maar patiënten met een blindarmonsteking tegenkomt. Bovendien kan Elffers in zijn model deze kans helemaal niet berekenen!

Het wiskundig model van Elffers blinkt uit in eenvoud. Op zich is dat prettig, want die eenvoud stelt hem in staat om berekeningen uit te voeren die in een meer gecompliceerd model onmogelijk zouden kunnen zijn. Maar het model is uiteraard een zeer grove versimpeling van de realiteit. Wanneer Elffers dit model beschouwt en tot de conclusie komt dat er geen sprake kan zijn van toeval, dan verwerpt hij niet alleen de hypothese dat incidenten altijd met dezelfde kans plaatsvinden, onafhankelijk van de aan- of afwezigheid van de verdachte. Nee, hij moet daarmee tot de conclusie komen dat het totale model de realiteit blijkbaar niet goed beschrijft. Het totale model bevat echter meer aannamen dan de zojuist genoemde hypothese, in dit geval om de berekening te stroomlijnen, ja zelfs mogelijk te maken. Een extra aanname is bijvoorbeeld, zoals eerder beschreven, dat elke dienst een vaste kans heeft op een incident, onafhankelijk van wat er tijdens andere diensten gebeurt. Die aanname van onafhankelijkheid is niet erg redelijk. Om een eenvoudig voorbeeld te noemen: in een gegeven dienst kan er een reanimatie (dus een 'incident') optreden die dan nog goed afloopt, maar die in de volgende dienst tot overlijden leidt. Ook is de aanname van een vaste kans niet erg realistisch; het hangt helemaal af van het type patiënten dat wordt opgenomen.

Als men nu op grond van het optreden van een onwaarschijnlijke gebeurtenis het model verwerpt, dan verwerpt men het gehele model. Echter, als men alleen weet dat de conjunctie 'A en B en C' onwaar is, dan weet men slechts dat of A of B of C onwaar is, maar niet welk van de drie. In dit geval weet men dus niet of de kernhypothese onjuist is, dan wel een van de idealiserende aannamen. Een verwant principieel probleem met de modelkeuze is de onmogelijkheid om daarbinnen alternatieve hypothesen te formuleren. Kijken we nog eens naar punt 11 van de bewijsmotivering:

Deze berekeningen geven derhalve aan dat het in hoge mate waarschijnlijk is dat er een verband bestaat tussen het werkzaam zijn van de verdachte en het zich voordoen van bedoelde incidenten.

Vanwaar het 'derhalve' hier? Voorzover 'verband' hier wiskundige betekenis heeft, moet het zoiets betekenen als 'de aanwezigheid van de verdachte vergroot de kans op een incident'. Dit suggereert dat er in het model twee hypothesen te formuleren zijn:

1. de hypothese van toeval, zeg zoals door Elffers geformuleerd, en
2. de alternatieve hypothese dat de kans op een incident tijdens een dienst van de verdachte groter is dan tijdens een dienst van de andere verpleegkundigen.

De beslissing tussen (1) en (2) zou dan genomen kunnen worden met de statistische methode van het hypothese toetsen. Dit is echter niet wat Elffers doet.

We merkten al eerder op dat de kans op een incident volgens Elffers niet te schatten is. Wanneer we de alternatieve hypothese willen formuleren, en daarbij dus kans voor de verdachte anders inschatten dan voor de overige verpleegsters, dan rest ons niet veel anders dan twee verschillende kansen te postuleren: één kans (zeg q) voor de overige verpleegsters, en één kans (zeg p) voor de verdachte. Maar nu komt het verrassende: als we uitgaan van de hypothese dat die twee kansen verschillend en onbekend zijn, dan kunnen we de hele methode van Elffers niet meer gebruiken, omdat de kansen die hij wil berekenen niet meer uit te rekenen zijn zonder extra kennis over de kansen p en q . De alternatieve hypothese is in zijn model dus niet te formuleren. Er is een invloedrijke school in de statistiek die deze procedure volstrekt onacceptabel vindt. Wanneer men de hypothese van toeval wil verwerpen ten faveure van de hypothese dat de kans op een incident tijdens een dienst van de verdachte groter is dan tijdens een dienst van de andere verpleegkundigen, dan moeten volgens deze school beide hypothesen in het statistische model geformuleerd kunnen worden. Kan dat niet, dan valt er uit het verwerpen van de hypothese van toeval geen enkele conclusie te trekken.

De interpretatie van de getallen

In de vorige twee paragrafen zijn we ingegaan op problemen met het model zelf; die problemen die te maken hadden met de specifieke keuze van dit model. In deze paragraaf bespreken we algemenere, meer filosofische pro-

blemen die te maken hebben met de interpretatie van elke statistische berekening. We beweren dat zowel Elffers als de rechtbank een onjuiste interpretatie van de berekeningen geven. Dit komt dus boven op alle eerdergenoemde problemen die samenhangen met dit specifieke model. In zijn NJB-artikel doet Elffers na alle berekeningen de volgende uitspraak:

Toeval als verklaring valt niet langer serieus te nemen.

En in *Trouw* schrijft hij:

Het is geen toeval, de rest is aan u.

Dat nu is nogal voorbarig. Elffers kan zijn model als geheel verwerpen, maar toeval als verklaring wordt daar niet door uitgesloten: zoals we gezien hebben kan toeval op een hele andere manier wiskundig worden vertaald, en over die (veel ruimere) vertaling zwijgt de methode van Elffers. Bovendien is het zeer discutabel of het optreden van een gebeurtenis met een hele kleine kans toeval uit kan sluiten, denk aan het voorbeeld van de loterij.

Laten we nu vervolgens kijken naar de interpretatie die de rechtbank aan de uitkomsten van de berekeningen gegeven heeft. Er zijn hier twee problemen.

Ten eerste is de interpretatie veel te realistisch. De rechtbank meent dat deze getallen over de werkelijkheid gaan, terwijl de uitkomsten van dit soort berekeningen altijd relatief zijn aan het gekozen model. We hopen intussen genoegzaam duidelijk gemaakt te hebben dat er vele modellen mogelijk zijn, en daarmee ook vele uitkomsten.

Ten tweede interpreteert de rechtbank de getallen verkeerd door te spreken over de 'kans op toeval'. We zagen dat het eigenlijk precies andersom is: eerst neemt Elffers (op zijn manier) de hypothese aan dat slechts het toeval bepaalt wie de incidenten treft, en hij berekent onder die hypothese de kans op de waargenomen gebeurtenissen. Deze kans blijkt klein te zijn, maar hieruit kunnen we per se niet concluderen dat de kans op de hypothese klein is. Het is zelfs maar de vraag of de 'kans op de hypothese' überhaupt iets is waar we over kunnen spreken; hierover bestaat in de statistiek geen overeenstemming. In het onderhavige geval bestaat de toevalshypothese uit een specificatie van een aantal parameters in het model, namelijk de hypothese dat alle kansen voor alle verpleegsters altijd dezelfde zijn. Elke uitspraak over de grootte van de kans op deze hypothese is een uitspraak buiten het wiskundig model om, en kan alleen maar 'subjectief' geïnterpreteerd worden, waarmee elke grond voor een objectieve waarde ontbreekt.

Er is echter ook een stroming binnen de statistiek die dergelijke subjectieve kansen juist als de aangewezen manier ziet om problemen als de zaak Lucy de B. aan te pakken. De Amsterdamse econometrist dr. Aart de Vos stelde recentelijk voor om een zogenaamde Bayesiaanse analyse van het probleem te maken. Bij een Bayesiaanse analyse is het uitgangspunt radicaal anders: aan het begin wordt een zogenaamde priorverdeling opgesteld over de kans dat een willekeurige verpleegster haar patiënten om-

brengt. Deze priorverdeling zal dan bijvoorbeeld vaststellen dat de kans dat een willekeurige verpleegster haar patiënten ombrengt, gelijk is aan 1 op de 40.000 (we noemen maar wat). Vervolgens wordt naar de data gekeken, en op basis van die data en enige elementaire kanstheoretische regels wordt dan de kans uitgerekend dat na het zien van die data Lucy de B. haar patiënten gedood heeft. De uiteindelijke uitkomst van die berekening hangt natuurlijk zeer sterk af van de prior die gekozen wordt. De Vos komt in de genoemde artikelen tot de conclusie dat de verdachte best met een kans van 80% onschuldig kan zijn, een getal dat in schril contrast staat met de getallen van Elffers.

Is dit nu een redelijke procedure? Wij vinden dat moeilijk in te zien. Hét grote probleem bij een Bayesiaanse analyse is de keuze van de priorverdeling. Als je de prior varieert, dan varieert de uiteindelijke kans net zo hard mee, waardoor het moeilijk is om na te gaan of dergelijke kansen nog wel enige betekenis hebben. Als ze al betekenisvol zijn, dan zijn het ook subjectieve kansen die slechts de mate van geloof in een bepaalde hypothese, in dit geval de hypothese dat er wel of geen toeval in het spel is, uitdrukken. Uiteindelijk is de 80% van De Vos net zo arbitrair als de 1 op 342 miljoen van Elffers. De willekeur van de prior kan wellicht vergeleken worden met de willekeur van de schaal van het model van Elffers. Welk paradigma je ook kiest, keuzes als deze moeten altijd worden gemaakt.

Alternatieven en de communicatie met de rechterlijke macht

Uit bovenstaande analyse kunt u opmaken dat het presenteren van een enkel getal in een statistische analyse een hachelijke en principieel onmogelijke zaak is. Een absoluut getal is nooit te geven: bij elke methode worden persoonlijke, subjectieve keuzes gemaakt, zowel bij de vertaling naar het wiskundig model als bij de correctie die wellicht moet worden toegepast.

We hebben gemerkt dat deze boodschap niet met blijdschap wordt ontvangen. De raadsheren in het hoger beroep spraken zelfs van een 'negatieve benadering': wij zouden slechts anderen bekritisieren zonder zelf met een fatsoenlijk tegenvoorstel – en een ander getal dus – te komen. Dit illustreert het grote probleem in de communicatie

tussen deskundige en rechter. Als rechter wil je natuurlijk graag gewoon een getal hebben, een betrouwbaar getal liefst dat ook nog iets voorstelt. Elffers gaf een getal, zijn inmiddels overbekende 1 op de 342 miljoen. Dat vond de rechtbank prettig, zo bleek ons. Dat kunnen wij ons wel voorstellen, maar de feitelijke keuze is tussen (1) een enkel getal dat niets voorstelt, of (2) een veel breder statistisch perspectief waarin geen eenduidig getal geleverd wordt, maar waarin het hele veld van mogelijkheden, aannames, interpretaties en modellen geschetst wordt, en dat de rechter met zijn gezonde verstand dient te beoordelen. Die keuze lijkt ons niet moeilijk. Een statisticus kan hooguit zeggen: als u dit aanneemt en dit gelooft, dan komt er dat uit; als u daarentegen meer geloof hecht aan die andere aanname, dan komt er dit uit. In het onderhavige geval van de zaak tegen Lucy de B. zijn er nog vele andere methodes beschikbaar die elk hun eigen problemen zullen opleveren, en waarin ook elke keer keuzes moeten worden gemaakt.

Wat een statisticus echter nooit mag doen, is uitspreken dat de gegevens niet door toeval verklaard kunnen worden. Want zoals we al zagen, herbergt deze uitspraak een veelheid van persoonlijke opvattingen in zich, die totaal uit zicht verdwijnen in de schijnbare en misleidende objectiviteit van de uitspraak. Het is allemaal niet zo simpel, en wanneer een statisticus een ferme uitspraak doet met een enkel getal, dan is dat misleidend. We besluiten met een citaat van de eerdergenoemde V. Barnett, dat precies aangeeft waar het om gaat:

In applying a particular technique in a practical problem, it is vital to understand the philosophical and conceptual attitudes from which it derives if we are to be able to interpret (and appreciate the limitations of) any conclusions we draw.

Dit artikel verscheen in september 2004 in TREMA – Tijdschrift voor de rechterlijke macht.

Prof. Dr. M. van Lambalgen, hoogleraar Logica en Cognitiewetenschap aan de Universiteit van Amsterdam

Prof. Dr. R. Meester, hoogleraar Waarschijnlijkheidsrekening aan de Vrije Universiteit

FREUDENTHAL 100
15, 16 & 17 september 2005



www.freudenthal100.nl

NIEUWE WISKRANT

25 jaar, 100 jaar...

Zaterdag 17 september 2005 is het precies 100 jaar geleden dat Hans Freudenthal werd geboren. Het eerste nummer van de vijfentwintigste jaargang van de *Nieuwe Wiskrant* zal geheel in het teken staan van HF100; een dubbelnummer in samenwerking met de *Panama Post*, het tijdschrift voor reken-wiskundeonderwijs aan vier- tot veertienjarigen en de volwasseneneducatie.

Voor de laatste informatie over het HF100 symposium: zie www.freudenthal100.nl