

Statistisch gezien is Lucia de B. onschuldig en zijn er lessen uit de zaak te leren

Thomas Colignatus, 26 juni 2007.

Econometrist, secretaris Samuel van Houten Genootschap, wetenschappelijk bureau van het Sociaal Liberaal Forum; voormalig medewerker Centraal Planbureau, hogeschool-docent statistiek, statistisch modelleur voor het bevolkingsonderzoek naar baarmoederhalskanker, <http://www.dataweb.nl/~cool>

Publiceerbaar. Hiervoor is ook een kortere versie beschikbaar.

De auteur dankt prof. dr. R. D. Gill, hoogleraar Mathematische Statistiek, Mathematisch Instituut, Faculteit Wis- en Natuurkunde, Universiteit Leiden, en dr. P. Grünwald, senior onderzoeker/projectleider learning & statistics, Centrum voor Wiskunde en Informatica, Amsterdam en Research Fellow EURANDOM, Eindhoven, voor commentaar op een eerdere versie. Omdat in deze kwestie vaak de indruk ontstaat dat statistici het onderling oneens zijn hechten Gill en Grünwald eraan dat hier gemeld wordt dat zij de statistische analyse in deze bespreking inhoudelijk onderschrijven. Zie ook Gill's website <http://www.math.leidenuniv.nl/~gill/lucia.html> voor diens visie in extenso over deze kwestie.

In deze bespreking wordt verwezen naar de rapporten die prof. Elffers voor de rechtbank schreef. Er wordt niet uit geciteerd omdat deze rapporten niet aan de openbaarheid zijn prijsgegeven. Mr. S. Franken, advocaat van Lucia de B. heeft deze rapporten aan de auteur ter inzage geven. Daarmee is de relatie van de auteur tot deze partij uitputtend beschreven.

Prof. dr. H. Elffers en prof. mr. R.V. de Mulder hebben de gelegenheid tot reactie gekregen en zien geen aanleiding een reactie aan deze bespreking te verbinden.

Statistisch gezien is Lucia de B. onschuldig en zijn er lessen uit de zaak te leren.....	1
Inleiding.....	2
Het probleem.....	3
1. De deskundige en de vakgenoten.....	5
2. De “data”.....	6
3. Wat is de betekenis van “incident” ?.....	8
4. De statistische attitude.....	9
5. De statistische verantwoordelijkheid voor de data.....	10
6. Wat als de deskundige de verkeerde methode kiest ?.....	11
7. De statisticus als poortwachter.....	12
8. Het optellen van onderdelen	13
9. Communicatie met de vakgenoten en het publiek (1).....	14
10. Communicatie met de vakgenoten en het publiek (2).....	15
11. Keuze van de statistische decisieleer.....	16
12. Omgaan met statistische decisieleer.....	16
13. Second opinion van S.....	20
14. Enkele nieuwe ontwikkelingen.....	20
15. Wetenschappelijke ethiek.....	21
Een dwaling van het recht.....	21
Conclusies.....	23
Literatuur.....	23
Bijlage A: Persbericht NOVA, 19 januari 2005.....	24
Bijlage B: Bijdrage van D aan de studiedag “Omgaan met Statistisch bewijs in strafzaken”	25

Inleiding

Op 18 juni 2004 werd Lucia de B. in hoger beroep veroordeeld tot levenslang plus TBS voor zeven moorden en drie pogingen tot moord. Er zijn argumenten gegeven om de zaak te heropenen (Derksen (2006)) en voor het OM kijkt momenteel de Commissie Posthumus II naar deze zaak. Hoe dit verder ook zij: het zijn rechters die over schuld of onschuld van Lucia de B. moeten oordelen. Deze bespreking laat de schuld of onschuld van Lucia de B. aan poging tot moord, moord of seriemoord derhalve in het midden en op de plaats waar die hoort: de zittingzaal. De auteur is econometrist en kan slechts oordelen over zijn vakgebied, in dit geval het onderdeel statistiek. Hieronder zal het gaan om de interactie tussen recht en statistiek, zal de strafzaak worden bekeken vanuit het gezichtspunt van de statistiek en zal daarbij vooral worden gekeken naar de rol van de statistiek. Het blikveld is evenwel beoogd voor de juridische context opdat er voor deze context lessen worden getrokken indien die er te leren zijn. De statistiek zelf is reeds aan de orde geweest in NJB door Elffers (2003) en in Trema in de besprekingen van Van Lambalgen en Meester (2004) en Elffers (2004b). De onderhavige bespreking richt zich op met name dit andere aspect: de taakverdeling tussen rechter en statistisch deskundige.

In de strafzaak tegen Lucia de B. is prof. dr. Henk Elffers als deskundige op dat terrein opgetreden. Elffers heeft de statistische kant van de zaak nadien op gestyleerde wijze in het openbaar beschreven en meer aspecten toegelicht toen er van de kant van vakgenoten vragen kwamen. Voor een rechter en de samenleving kan het relevant zijn kennis te nemen van kritiek op diens rapport en rol. De rechtbank zelf heeft prof. mr. R.V. de Mulder de methode van Elffers laten controleren, deze heeft de methode onderschreven maar is nadien niet in de openbaarheid getreden.

In voorjaar 2007 oordeelde een collega-statisticus nog: “Achteraf kritiek leveren op iemand’s rapport is meestal makkelijk, maar zelf een uitstekend rapport schrijven waar iedereen het mee eens is is heel erg moeilijk.... Ik ben daarom misschien wat milder (...)” Toch vond ook deze collega dat de deskundige(n) een paar zaken diende(n) te corrigeren. Dit artikel is geschreven vanuit de gedachte dat het inderdaad gemakkelijk is om kritiek te leveren. Daarom is met zorgvuldigheid gekeken naar de bijdragen van de rechtbank-deskundigen en naar wat in de laatste jaren door externe statistici over de zaak is geschreven.

De conclusie van dit artikel is: statistisch gezien blijft Lucia de B. onschuldig en kunnen lessen geleerd worden uit de rol van de statistiek en ‘de’ statistisch deskundige. Hierbij is het eerste een open deur omdat statistiek en rechtspraak zoals gezegd andere dimensies zijn. Maar het blijft nuttig om telkenmale te benadrukken dat statistiek niet misbruikt moet worden en dat statistici daarvoor als eersten dienen te waarschuwen.

In het bijzonder:

- (1) Statistiek dient vooral het onderzoek van het OM. Hoewel gebruik binnen de zittingzaal denkbaar blijft en nooit uitgesloten kan worden (bijv. bij kansspelen of forensische resultaten) dient dan wel behoedzaamheid te worden betracht.
- (2) Voor het OM blijven het belangrijke vragen hoe de data zijn verzameld en hoe statistische beslissingscriteria worden vastgesteld. Data verzameld door rechercheurs uitgaande van verdenking kunnen niet gebruikt worden voor statistische analyses die random selectie veronderstellen. Normen voor statistische significantie voor wetenschappelijke of industriële toepassingen kunnen niet worden toegepast voor misdrijven en het is onjuist dan te verwijzen naar “statistische conventies”.

- (3) Statistiek kan ten enen male slechts kansuitspraken doen. Toeval is m.a.w. nooit uit te sluiten, ook al willen sommige statistici de (in hun ogen) “leken” helpen door toch stellige uitspraken te doen. Die kansen worden ook bepaald door een onderliggende causaliteit, en daarvoor geldt: “Het vraagstuk van de causaliteit wordt niet door de statistiek opgelost, maar door de biologie, de psychologie, de medische, de sociale of economische wetenschap, die zich met de onderzochte verschijnselen bezig houdt” (Rijken van Olst (1969:180)).
- (4) Strafzaken, waarin eenmaal de statistiek in de zittingzaal aan de orde is gekomen en zelfs tot het arrest heeft bijgedragen, bijv. met een gedachte van “dit kan geen toeval zijn”, komen voor integrale heroverweging in aanmerking. Bij herbehandeling voldoet het niet de statistiek slechts terzijde te leggen, of dit zo te stellen, omdat noties, zoals “dit kan geen toeval zijn”, kunnen blijven sluimeren en kunnen bijdragen tot the *conviction intime*. Ter nieuwe zitting dienen de statistische fouten helder benoemd te worden om alle sluimerende gedachten te verwijderen.
- (5) Bij interactie tussen statistiek en andere vakgebieden blijft zorgvuldigheid gewenst. Bij een overlijden waarbij geen (bekende) natuurlijke oorzaak kan worden aangewezen kan nog steeds toeval bestaan. Wanneer een rechter in de zittingzaal geen statistiek wil toestaan zou het niet juist zijn dat ook toeval als (onbekende) natuurlijke doodsoorzaak wordt uitgesloten.
- (6) Statistische rapporten zouden eerder aan externe wetenschappers openbaar gemaakt mogen worden, eventueel na een herschrijving ten behoeve daarvan.
- (7) Het recht lijkt gediend met de figuur van een wetenschappelijk adviseur verbonden aan de juridische macht, die door externe vakgenoten kan worden aangeschreven wanneer zij vragen of zorgen hebben omtrent toepassingen van hun vakgebied binnen de zittingzaal. Momenteel kan men eigenlijk niemand benaderen. Wetenschappers zijn minder geneigd om het OM te benaderen omdat zij geen affiniteit met vervolging hebben, terwijl zo’n benadering ook minder zinvol is wanneer er een uitspraak ligt (die tot de vragen leidt). Rechter zal men ook niet benaderen omdat men de onafhankelijkheid van de rechter respecteert ook al bevat een uitspraak wetenschappelijke dwalingen. Ook de partijen in een geschil zal men niet snel benaderen omdat een wetenschapper niet de indruk van partijdigheid zou willen krijgen en omdat de wetenschap onpartijdig is. Überhaupt zal men aarzelen omdat er wel vragen kunnen ontstaan maar men geen tijd heeft om een gehele zaak in al zijn details te beoordelen. Een wetenschappelijk adviseur verbonden aan het hof die commentaar van externe wetenschappers ontvangt zou e.e.a. kunnen toetsen en intern de juiste kanalen kunnen gebruiken om de informatie naar de juiste plek door te geleiden.

Deze bespreking is als volgt opgebouwd. Eerst wordt een overzicht gegeven van de strafzaak tegen Lucia de B. en de bijdragen van de statistisch deskundigen. Voor de eenvoud noemen we haar voortaan “V” (“verdachte”), Elffers “D” (“deskundige”) en De Mulder “S” (“second opinion”). Gezien zijn publicitaire terughoudendheid kan weinig over de rol van S gezegd worden en de aandacht richt zich in hoofdzaak op D. De controlerende expertise blijkt vooral te komen uit de waarschuwingen van de externe statistische collega’s en we laten zien dat D ten onrechte niet corrigeert. De statistische deskundigen D en S vormen zo poortwachters die voorkomen dat de juiste statistiek gebruikt wordt.

Het probleem

Het volstaat te verwijzen naar de oorspronkelijke bronnen en hun eigen samenvattingen, aangezien die tegenwoordig vaak direct op het internet toegankelijk zijn. De uitspraken van de rechterlijke colleges zijn beschikbaar op www.rechtspraak.nl (2006). Een weerwoord staat op <http://www.luciadeb.nl/> en met name in het boek van Derksen (2006), waarbij de kritiek in het boek overigens al op eerdere datum is gegeven. Sjerps en Coster van Voorhout (ed.) (2005) bespreken het gebruik van statistiek en kansrekening in het strafrecht, met daarin opgenomen Elffers (2005). Gezien de doorlooptijd is dit laatste denkbaar reeds in 2004 geschreven. Dit

artikel vormt hier de beste verwijzing omdat het minder “gefictionaliseerd” is dan Elffers (2003), zie ook Elffers (2005, voetnoot 5). StatOr (2004) is een thema-nummer van het blad van de Vereniging voor Statistiek en Operationele Research (VVS), gewijd aan de kwestie, met daarin een bijdrage van Elffers (2004a) en kritiek van collega’s. Van Zwet (2004) evalueert de verschillende bijdragen van StatOr (2004). Zijn evaluatie is welbeschouwd de meest inzichtelijke en kan als leidraad worden gebruikt. Later in dat jaar verschenen ook Van Lambalgen en Meester (2004) met een antwoord van Elffers (2004b). Meester e.a. (2006) tillen de kwestie naar een internationaal niveau.

De advocaat van V heeft aan de auteur op vertrouwelijke basis twee rapporten ter beschikking gesteld welke D voor het Openbaar Ministerie opstelde, te lezen als één rapport Elffers (2002). Rechtbank en hof hebben deze oorspronkelijk rapporten Elffers (2002) nog niet aan de publiciteit vrijgegeven, ook al heeft D daar om verzocht (en het is te prijzen dat D daartoe heeft verzocht). Gezien de vertrouwelijkheid van deze rapporten wordt er hier niet uit geciteerd maar wordt er slechts naar verwezen.

Sjerps en Coster van Voorhout (2005:vii) geven een goede inleiding tot het algemene strafrechtelijke probleem:

“Strafjuristen zijn op zoek naar zekerheid voor het oordeel over de waarheid van feiten (de materiële waarheidsvinding). Een deel van deze juristen heeft daarbij de verplichting om slechts op basis van *bewijs* tot een veroordelend vonnis in een strafzaak te komen. De strafrechter zal immers slechts op grond van de inhoud van wettige bewijsmiddelen tot de overtuiging mogen komen dat verdachte het hem of haar tenlastegelegde feit heeft begaan. Het is goed te beseffen dat deze wettelijke verplichting vrijwel het enige juridisch normatieve aspect in de bewijsredenering vormt. Voor het overige hangt de waarheidsvinding, als navolging van genoemde verplichting, in hoofdzaak af van logische waarschijnlijkheidsredeneringen met een daarbij behorende dosis aan ‘conviction intime’.” (voetnoten weggelaten).

Waar de rechtbank in eerste aanleg de berekeningen van D expliciet een rol liet spelen bij de bewijsvoering stelt het arrest van het gerechtshof: “Er is geen statistisch bewijs in de vorm van toevalsberekeningen gebruikt.” (samenvatting, LJN: AP2846 d.d. 18-06-2004) Dit blijkt pro forma correct omdat men in het arrest inderdaad zulke berekeningen niet aantreft. Pro forma zijn de berekeningen van D derhalve niet meer relevant voor de rechtsgang zelf. Evenwel valt de conclusie niet te vermijden dat het gerechtshof een en ander wel impliciet een rol laat spelen. In zijn arrest heeft het gerechtshof namelijk (a) regelmatig *toevalsredeneringen* gebruikt om tot een bewezenverklaring te komen, en (b) niet uitgesloten is dat het *oordeel* van de deskundige een rol heeft gespeeld bij de overtuigingsvorming (*conviction intime*) ook al zijn de onderliggende berekeningen niet gebruikt. Het aangehaald citaat van Sjerps en Coster van Voorhout geeft aan dat het recht überhaupt moet oordelen in een wereld van onzekerheid en we zien dit in het oordeel terug.

De kwestie omtrent de (kwantitatieve) statistiek is ook om andere redenen nog geen “dood paard”. (i) De commissie Posthumus II onderzoekt ook de statistiek. (ii) De statistiek zal sowieso in de gedachten een rol spelen en het is goed dat er nog meer helderheid over komt, eventueel voor toekomstige strafzaken. (iii) Voorzover statistiek in deze kwestie waarde heeft pleit zij V eigenlijk vrij. Men zou de statistiek dus niet geheel mogen buitensluiten.

Een belangrijk punt ad (ii) waarbij statistiek in de gedachten nog een rol zal spelen betreft de mogelijke onderlinge beïnvloeding van de verschillende deskundigen. Met name constateert Derksen dat de medische deskundigen hun rapportage mede onderbouwden met een verwijzing

naar de uitkomsten van D dat er “geen toeval” zou zijn. Ook deze kruisinvloed leidt ertoe dat de statistiek meer consequenties heeft gehad dan de uitspraak van het Hof stelt.

Last but not least is er de vorming van de *conviction intime*. Deze is een juridisch noodzakelijk voorwaarde voor een oordeel. Wetenschappers zijn vertrouwd met deze gedachte omdat zijzelf eenzelfde voorwaarde stellen bij het accepteren of verwerpen van een wetenschappelijke theorie. Men dient bij zichzelf te rade te gaan om te bepalen waar men wel of niet van overtuigd is. Het is klaarblijkelijk niet gebruikelijk in de jurisprudentie om de vorming van deze *conviction intime* nader toe te lichten. Deze wordt geacht te blijken uit het gepresenteerde wettelijk bewijs. In dit geval kan het echter zijn dat statistische overwegingen van invloed zijn die samenhangen met gedane berekeningen welke berekeningen expliciet worden uitgesloten. In zo'n geval zou het prudent zijn om ofwel het ontstaan van de *conviction intime* nader toe te lichten ofwel de gehele statistische argumentatie inclusief de berekeningen door te nemen om verkeerde invloeden uit te sluiten.

1. De deskundige en de vakgenoten

Bij de rol van de statistisch deskundige valt op dat vakgenoten kritiek hebben uitgeoefend. Bestudering van de publicaties en het oorspronkelijke rapport van D aan het OM en de rechtbank laat zien dat D in een aantal opzichten redelijk werk heeft verricht. Begrip dient er ook te zijn voor zijn niet al te gemakkelijke positie in het spanningsveld tussen gebrek aan harde data en het risico van een seriemoordenaar. Klaarblijkelijk heeft hij in het begin aan het Openbaar Ministerie opgedragen dat de “data” op een goede wijze verzameld dienden te worden, heeft hij een “post-hoc” correctie toegepast om te corrigeren voor data-selectie, is zijn keuze voor het hypergeometrische model in bepaalde opzichten toelaatbaar, en heeft hij de methode van de statistische decisieleer omtrent het beslissen over hypothesen in ieder geval qua vorm correct toegepast. Waar het hier om gaat is dat “in een aantal opzichten redelijk werk” niet toereikend is waar het moet gaan om “adequaat werk”. Wanneer collega's toereikende argumenten geven dat het werk niet adequaat is dan dient de statisticus in staat te zijn om de kwaliteit van die argumenten te zien, en de moed hebben daarop te corrigeren.

In de latere besprekingen Elffers (2003, 2004ab, 2005) zijn in toenemende mate nuanceringen toegevoegd ten opzichte van de oorspronkelijke rapportage Elffers (2002). Wellicht zal D menen, en velen met hem, dat hij daarmee ingaat op de kritiek van de collega's. Echter, ten eerste ontstaat hierdoor een verkeerd beeld voor het publiek van wat hij de rechtbank oorspronkelijk heeft voorgelegd, ten tweede blijkt bij precieze lezing dat D niet wezenlijk op de bezwaren van de collega's ingaat. Dit zijn twee aspecten die moeilijk over het voetlicht zijn te brengen zolang de oorspronkelijke rapporten niet publiek zijn gemaakt. Maar met enige wijsheid kan het argument denkbaar wel helder worden, juist ook omdat D zijn statistische fouten nooit herroepen heeft en ook in zijn publieke teksten heeft herhaald. D is in die zin dus standvastig dat hij zijn fouten niet herroept, en zijn latere publicaties kenmerken zich door het aanbrenge van nuanceringen en het weglaten van onwelgevallige aspecten.

De rapporten van 8 en 29 mei 2002 bestaan overigens uit 7 respensievelijk 3 pagina's A4 met een groot lettertype en zij richten zich vooral op de berekening. Veel ruimte voor nuance is daarin niet. Het moge duidelijk zijn dat zij daarin afwijken van de vele pagina's die D daarna nog in het publieke domein heeft laten verschijnen. Wat aan deze rapporten van belang is is ook het grote omliggende kader achterin waarin D in twee zinnen zijn conclusie aan de rechtbank meedeelt. Deze uitspraak van D vernietigt alle nuance die daarvoor nog kan hebben bestaan. Zij verschilt welbeschouwd weinig van de uitspraak die D later publiekelijk deed: “Geacht hof, het was geen toeval. De rest is aan u.” (zie Elffers (2004a)).

2. De “data”

De “data” zijn in de publieke bronnen terug te vinden in Derksen (2006) en zien er oorspronkelijk in 2002 uit zoals weergegeven in Tabel 1. Vervolgens komt Derksen zelf tot de correctie in Tabel 2. Waar wij geen toegang hebben tot de procesgegevens laten wij de kwaliteit van zowel de oorspronkelijke “data” als de “gecorrigeerde data” in het midden. Voordat we de “data” überhaupt kunnen gebruiken zijn de volgende vragen relevant:

- (1) Wat zijn de *vragen*, met als doel waarvoor deze “data” zijn verzameld ? Beantwoorden deze “data” aan het doel van de vragen ?
- (2) Gaat het hier om “moorden of pogingen tot moord” of om “medisch onverklaarbare incidenten met de noodzaak tot reanimatie, al dan niet met dood tot gevolg” ? Wat is “medisch onverklaarbaar” ? Is “toeval” geen medische verklaring ? Het maakt natuurlijk wel wat uit of er een dood volgt, en hoe zijn de data dan ? Maar een dood is nog geen moord. Wat te doen met verklaringen van medici dat de diagnostiek of zelfs ook de (post mortem) obductie in een ziekenhuis gangbaar geen rekening houdt met moord of het bepalen van de mogelijkheid van moord ? Moet en kan statistiek in eventuele manco voorzien ?
- (3) Waarom alleen verplegers en geen dokters die eventueel verkeerde medische beslissingen namen ? Wat zijn “diensten”, zijn dag- en nachtdiensten en kinderen en volwassenen gelijk ? Gaat het hier om het einde van een analytische trechter met meer ingangen of is er vanaf het begin sprake geweest van verkokering in de vraagstelling ?
- (4) Hoe zijn de “data” verzameld ? Is men gaan verzamelen nadat de verdenking ontstond en heeft men “incidenten” van niet-verdachten daarbij snel opzij gelegd ?
- (5) Zijn dit de enige “data” of dienen meer data verzameld te worden (afhankelijk van de vraag die gesteld wordt) ?

Tabel 1: Oorspronkelijke “data” gebruikt door D (zie Derksen (2006))

	JKZ			RKZ afd. 41			RKZ afd. 42			Totaal		
	Ja	Nee	Som	Ja	Nee	Som	Ja	Nee	Som	Ja	Nee	Som
V	8	134	142	1	0	1	5	53	58	14	187	201
And.	0	887	887	4	361	365	9	272	281	13	1520	1533
Som	8	1021	1029	5	361	366	14	325	339	27	1707	1734

Aantallen “diensten” door V en anderen, “ja” = incident, JKZ = Juliana Kinder Ziekenhuis, RKZ = Rode Kruis Ziekenhuis, Totaal = som JKZ en RKZ

Tabel 2: “Gecorrigeerde data” door Derksen (2006:140)

	JKZ			RKZ afd. 41			RKZ afd. 42			Totaal		
	Ja	Nee	Som	Ja	Nee	Som	Ja	Nee	Som	Ja	Nee	Som
V	4	138	142	1	2	3	1	57	58	6	197	203
And.	1	886	887	4	359	363	9	272	281	14	1517	1531
Som	5	1024	1029	5	361	366	10	329	339	20	1714	1734

In feite laat zich vaststellen dat er een vermenging van vraagstelling en dataverzameling heeft plaatsgevonden. De zaak is niet begonnen met een “vermoeden van niet-natuurlijke dood” maar omdat een collega-verpleegkundige na de dood van een patiënt naar een teamleidster stapte met de opmerking dat ze een “raar gevoel” had dat L. toch wel erg vaak bij overlijdensgevallen en reanimaties betrokken was. Daarop is de directeur van het JKZ geïnformeerd, die op grond van “statistiek van de koude grond” (zoals hij het zelf noemde) meende te kunnen concluderen dat het

allemaal te toevallig was. Daarop is aangifte gedaan bij de politie. De artsen hebben in de nacht dat die patiënt overleed een verklaring van natuurlijke dood getekend; die is later ingetrokken, nadat de directeur zijn oefening in statistiek had verricht. De collega-verpleegkundige heeft ook verklaard dat zij geen concrete aanwijzing had om te zeggen dat betreffende patiënt (door Lucia) om het leven zou zijn gebracht. In deze context is het OM begonnen met de “data-verzameling”. Statistische dataverzameling is anders dan die welke een OM doet, omdat een statisticus alle gegevens wil hebben terwijl het OM alleen in de verdachte is geïnteresseerd. Derksen (2006:131) meldt uit het dossier: “(...) dat het OM, zodra het zich realiseerde dat een incident niet Lucia betrof, het incident terzijde schoof. Zo schreef op 21 januari 2004 brigadier Spaans dat een vijftal onderzoeken ‘in opdracht van de officieren van justitie (...) niet zijn voortgezet’, omdat ‘ten aanzien van genoemde patiënten niet [is] gebleken dat verdachte enige betrokkenheid had bij het overlijden of een eventueel incident’. [bron]” Zie ook de andere voorbeelden van Derksen. In statistische termen heeft dus dataselectie plaatsgevonden.

In de publicaties van D vinden we verschillende omschrijvingen van welke vraag aan de orde was.

- (1) Elffers (2004b) stelt: “Zou het mogelijk zijn dat deze opeenhoping van incidenten bij verdachte toevallig tot stand komt, en er dus geen aanleiding is tot vervolging over te gaan? Die vraag werd mij door de OvJ gesteld.”
- (2) In Elffers (2005:133) is het “Als incidenten zich opvallend vaak voordoen als een persoon dienst heeft, en zelden of nooit als die persoon geen dienst heeft, kan dat feit dan, na een passende analyse met wiskundig-statistische methoden, bijdragen aan het bewijs dat die persoon schuldig is aan die incidenten?” *Maar*: een vervolging gaan instellen – de persoon als verdachte aanmerken en nader onderzoek gaan plegen om *een bewijs* te krijgen – is wat anders dan iemand schuldig verklaren (“bijdragen aan het bewijs”).
- (3) Elffers (2005:135) geeft een mengvorm: “Een van de diverse bewijsmiddelen die het OM wil aanvoeren is dat het geen toeval kan zijn dat V zovaak bij een incident betrokken kan zijn. Het OM vraagt daarom een statisticus: “Kan dat toeval zijn?”” *Maar*: in deze beschrijving (een mengvorm waarin “opvallend vaak” overgaat in “geen toeval”) gaat het wederom om een “bewijsmiddel” en niet slechts het identificeren van een verdachte.
- (4) In Elffers (2004a) luidt de vraag weer iets simpeler, namelijk “kan het toeval zijn?”, met als het antwoord: “Van belang is in te zien dat het zich een oordeel vormen over de plausibiliteit van deze verklaringen dan wel de verklaring dat mevrouw V de hand in de incidenten heeft gehad aan de rechter is. Als statisticus heb ik de vraag ‘Kan dat toeval zijn?’ beantwoord met een ongeclausuleerd ‘Nee’.”

De inconsistentie in deze publicaties van D omtrent de vraagstelling heeft ongetwijfeld er mee te maken dat iedere publicatie of passage een iets ander doel kan hebben. Echter, steeds gaat het om dezelfde kwestie van V met dezelfde rol van de deskundige ten behoeve van de rechter. In de context van die kwestie mag men wel een consistente vraagstelling verlangen. Een jurist zal menen dat het uiteindelijk gaat om de vraagstelling die het OM cq. de rechtbank aan D heeft voorgelegd en die in het rapport Elffers (2002) wordt behandeld. Voor de jurist zal het niet veel uitmaken wanneer het OM cq. de rechtbank verschillende vragen stelt om uiteindelijk in een convergerend proces tot de definitieve vraagstelling te komen. Voor de statisticus maakt dit alles wel uit, omdat verschillende vragen tot verschillende onderzoeksopzetten leiden. Bovendien kan het niet zo zijn dat het uiteindelijke rapport in 2002 nadien in latere publicaties leidt tot andere omschrijvingen.

Het zijn deels ook dubieuze vragen. Wat oorzaak of gevolg is dient door een vakinhoudelijke deskundige te worden gesteld waarbij de statistiek slechts iets zegt over data. Stel dat een gemankeerd natuurkundige een statisticus verkeerde data geeft waarin objecten zowel omhoog als omlaag vallen. De statisticus rapporteert dat hij 0% correlatie heeft gevonden tussen de

zwaartekracht en de richting waarin objecten vallen. Mag de natuurkundige dan alle verantwoordelijkheid van zich afschuiven en zich beroepen op de statistiek: “Met statistische zekerheid blijkt de zwaartekracht geen rol te spelen.” ? Nee, dat lijkt toch niet. Op dezelfde wijze hebben OM en rechters een taak bij het vaststellen van schuld of onschuld en kan de statisticus slechts iets zeggen over de data. Eventueel kan de statisticus oordelen als iedere burger, met gezond verstand, dat iemand die niet aanwezig was geen dader kan zijn; maar de bespreking gaat hier niet om kwesties van gezond verstand maar om de taakverdeling tussen deskundige en de rechtspraak.

In het artikel in Trema concludeert Elffers (2004b): “(De statisticus) reikt een bewijsmiddel aan. Niet meer, niet minder.” Dat is echter onjuist wanneer statistiek slechts dient voor het OM tot het identificeren van verdachten voor nader onderzoek, aldus wanneer statistiek in principe buiten de zittingzaal dient te blijven omdat (cq. wanneer) statistiek niets zegt over de schuldvraag. Van de verschillende vragen is no. 1 misschien nog het zuiverst. Maar dan krijgt iedere burger het risico vervolgd en zelfs veroordeeld te kunnen worden slechts door een opeenhoping van ongelukkig toeval ergens vaak in de buurt te zijn. Wellicht zijn alle burgers van Amsterdam wel schuldig omdat ze vaak in de stad zijn terwijl daar veel moorden worden gepleegd. Vraag 1 lijkt zuiver maar is het niet. Statistiek kan in dit geval slechts gebruikt worden om onderzoek te vernauwen tot nader te onderzoeken “incidenten” die op hun eigen merites eventueel tot vervolging kunnen leiden. Dat is dus een geheel andere vraagstelling.

3. Wat is de betekenis van “incident” ?

Zou het om pogingen tot moord gaan dan zouden volgens de oorspronkelijke “data” nog 13 pogingen gedaan zijn door andere verplegers waar V niet bij betrokken was – en deze verplegers zouden nu nog vrij rondlopen.

Neem echter aan dat het niet om moorden gaat maar om “onregelmatigheden”. Indien men in de oorspronkelijke “data” 13 incidenten op 1533 diensten van de andere verplegers normaal acht, dan kan men in een binomiaal model de gemiddelde kans opvatten als $p = 13 / 1533$. Met 201 diensten van V kan men de kansen bepalen op het voorkomen van dergelijke onregelmatigheden in haar diensten. Ditzelfde kan men doen met de “gecorrigeerde data” van Derksen (2006:140) met $p = 14 / 1531$. De resultaten van het binomiale model staan in Tabel 3 waarbij “minstens 14” de uitkomst is voor de oorspronkelijke “data” en “minstens 6” de uitkomst is voor de “gecorrigeerde data” (vet afgedrukt). Om mogelijke misverstanden te voorkomen zijn de uitkomsten niet afgerond.

De conclusies zijn als volgt. Bij de oorspronkelijke “data” zal men V gangbaar als een niet-normale verpleegster beschouwen met een kans van 1 op 348 miljoen dit fout te zien. Bij de “gecorrigeerde data” is de kans op die fout 1 op 87 (*géén miljoen*): zou je V in dat geval niet-normaal noemen dan zou je bij 87 geheel normale verplegers er één toch niet-normaal noemen. Van belang is echter dat alle uitkomsten nog steeds kans-verschijnselen zijn. Alles kan nog steeds toeval zijn. De feitelijke (niet-statistische) beslissing om V wel of niet als een normale verpleegster te beschouwen moet genomen worden op grond van andere gegevens. Het is overigens niet bekend wat hier de normen zijn, en wat dan eventueel gedaan wordt. Bovendien betekent het verschijnsel dat iemand een “niet-normale” verpleegster is nog niet dat zij een (serie-) moordnares is – want zoals gezegd gaat het hier om “incidenten” en geen moorden want anders zijn er nog 13 of 14 onopgehelderde moorden van de andere “normale” verpleegsters.

Dit binomiale model maakt het ook mogelijk om een andere kritische vraag te stellen. Wanneer we de uitkomsten van het RKZ als uitgangspunt nemen met in de oorspronkelijke “data” $4 + 9 = 13$ “incidenten” uit $365 + 281 = 646$ “normale” diensten van de andere verplegers, dan zijn de 0

“incidenten” van het JKZ bij de 887 “normale” diensten ook nogal onwaarschijnlijk (1 op de 68 miljoen). Er lijkt dus een duidelijk risico te bestaan dat die “data” van het JKZ vertekend zijn.

Op zichzelf is een kans van 1 op de 348 miljoen niet absurd klein, dat hangt af van de context. Maar zulke kleine kansen kunnen wel aangeven dat er iets aan de hand is, bijvoorbeeld dat de data vertekend zijn.

Hiermee is het statistische probleem beschreven. Wat is nu vervolgens de rol van de deskundige geweest en wat valt daarvan te leren ?

Tabel 3: De kans op minstens X “gevallen” bij V (binomiaal met p ontleend aan anderen)

Zie de tekst voor een berekening met “precies X” en de hypergeometrische verdeling

Minstens X “gevallen” bij V	Oorspronkelijke “data” $p = 13 / 1533, n = 201$	“Gecorrigeerde data” $p = 14 / 1531, n = 203$
5	0.02927792295402542	0.03985763942405506
6	0.007793872043391214	0.01150674575058763
7	0.001791529749499143	0.002872525000988446
8	0.0003614604153810985	0.0006300180956871903
9	0.00006486215381484861	0.0001229779714166934
10	0.00001046412470151469	0.00002159222914333014
11	1.531395710241412 E-6	3.440371792340357 E-6
12	2.048427146794651 E-7	5.011784943720343 E-7
13	2.520532982330792 E-8	6.717763822994271 E-8
14	2.868839166225711 E-9	8.330810485901452 E-9
15	3.03501668241779 E-10	9.604101958160526 E-10

4. De statistische attitude

Van Zwet (2004) brengt de oude les van de wetenschap der statistiek in herinnering dat de statisticus enige verantwoordelijkheid behoudt om erop te letten dat anderen niet met de conclusies aan de haal gaan. Ten aanzien van het Mathematisch Centrum (MC) meldt hij onder meer: “Een andere stelregel was dat het MC een vetorecht behield op de formulering van de statistische conclusies door de klant.” Daarentegen stelt D: “Geacht hof, het was geen toeval. De rest is aan u.” (zie Elffers (2004a)). Er is hier een groot verschil.

In het bijzonder gaat het hierom. Van Zwet: “De argumenten [van de rechtbank] zijn een samenvoeging van de aanwijzingen voor moord en de statistische conclusie van Elffers omtrent de aanwezigheid van LdB. Dit is een staaltje van wanordelijk denken en de statisticus behoort daarop te wijzen.” Het punt is dat “aanwijzingen” geen “bewijs” zijn. “Circumstantial evidence” en de statistische aanwezigheid van een persoon hoeft nog niets te bewijzen.

Van Zwet (2004): “Mijn voornaamste probleem met de gang van zaken bij de rechtbank is dat Elffers zich tegenover de rechter misschien wat al te vrijblijvend heeft opgesteld en dat dit tot misverstand bij de rechter heeft geleid. In Elffers’ visie is de rechter de baas. De getuige geeft advies en dan moet de rechter maar zien wat hij daarmee doet. Zoals ik in mijn inleiding over de goede oude tijd getracht heb duidelijk te maken, heeft men al 50 jaar geleden ontdekt dat dit geen goed model voor statistisch advieswerk is.”

In recente publicaties stelt D dat door de uitspraak van het gerechtshof de statistiek geen rol meer zou spelen. Elffers (2005:134) stelt: “De statistische argumentatie wordt in het arrest niet aangeroerd, en heeft kennelijk niet bijgedragen tot het wettig en overtuigend bewijs”. D miskent dan dat dit alleen om de “vorm van toevalsberekeningen” gaat (de formulering van het hof).

Lezing van de uitspraak van het hof geeft te zien dat statistiek impliciet en materieel nog steeds een rol speelt. Een statisticus dient niet goedgegelovig te zijn. Kritisch lezen is gewenst. Het statistisch oordeel dient onafhankelijk van het recht te blijven.

Had D een andere houding gehad en na de kritiek van vakgenoten zijn fouten gecorrigeerd dan had het gerechtshof eventueel eerder kunnen concluderen dat *juiste* toevalsberekeningen niet omstreden hoeven zijn zodat zij wel degelijk gebruikt hadden kunnen worden – met de mogelijkheid van vrijspraak. Nu de rechtbank berekeningen heeft toegelaten dient bij herbehandeling de mogelijkheid tot correctie te bestaan en is het niet voldoende om eerdere berekeningen, met al hun impliciete associaties, als niet-gedaan te beschouwen.

Voor juristen is deze laatste suggestie een nogal zware eis. In het strafrecht is, om bureaucratie en werkverschaffing te vermijden, aan de rechter de ruimte gegeven om materiaal te negeren zonder dit nader te hoeven te beargumenteren. In dit geval zou desalniettemin een figuur gevonden moeten worden dat er wel degelijk argumentatie ontstaat die toetsbaar wordt.

5. De statistische verantwoordelijkheid voor de data

De tweede belangrijke kwestie betreft de vaststelling van de data. VanZwet (2004): “Toen men in het ziekenhuis verdenking tegen LdB had opgevat, ging men alle gevallen van onverwacht overlijden of reanimatie na. Bij zo’n onderzoek ligt er het levensgrote gevaar op de loer dat de onderzoekers al overtuigd zijn van wat zij moeten vinden. Als er dus een onverklaarbaar overlijden wordt gevonden waarbij LdB aanwezig was dan wordt dit verder onderzocht en vindt zo’n geval al gauw een plaats onder de moorden. Als daarentegen LdB niet aanwezig was dan zal ongetwijfeld de neiging bestaan om dit geval als een natuurlijke dood te classificeren. Dit is een overbekend verschijnsel in de statistiek en ik hoop maar dat iemand ook echt met de onderzoekers gepraat heeft om uit te vinden of zoiets mogelijk kan zijn gebeurd.”

D noemt het dataprobleem maar onttrekt zich aan publieke bespreking: “Zulke twistpunten zijn in het feitelijk proces uiteraard van wezenlijk belang, omdat zij de gegevens waarop de analyse is gebaseerd in twijfel trekken. Voor de ontvouwing van de redenering kunnen we er hier evenwel aan voorbijgaan.” Elffers (2005:134-135)

In geen van de publicaties van D hebben wij een gepaste data-analyse aangetroffen. Omgekeerd zijn er wel aanwijzingen dat hij verdachtmakingen omtrent manco’s in de dataverzameling verre van zich werpt. In het dossier zal D’s opdracht aan de recherche om adequate data ter verzamelen terug te vinden zijn (dat is althans een redelijke aanname) en onder normale condities mag men veronderstellen dat de recherche ook niet gek is. Het punt is echter dat de statisticus de eigen verantwoordelijkheid niet op deze wijze van zich af kan schuiven en naar anderen kan wijzen. Er dient wel degelijk een eigen toets te zijn van de verrichte werkzaamheden. Met name omdat de statistiek vertrouwd is met de door VanZwet genoemde “neiging” en omdat er wel degelijk een vermoeden van “gekte” bestaat (heksenjacht, afschuiven van fouten van dokters op verpleegsters). Met name, en dit mag met nadruk gesteld worden omdat dit juist een statisticus opvalt, omdat een statistische analyse op de aangeleverde data zelf suggereert dat deze data niet homogeen zijn.

In geen van de teksten van D treffen wij een waarschuwing omtrent de kwaliteit van de “data” aan en dat deze met voorzichtigheid zijn te behandelen. Dit ontbreekt met name in het rapport dat aan de rechtbank is aangeleverd.

6. Wat als de deskundige de verkeerde methode kiest ?

Voor de definitie van “incident” kiest D: “dat een kind plotseling in acuut levensgevaar verkeert en een reanimatiepoging, al of niet met succes, wordt ondernomen. Dat is erg opvallend, omdat het een afdeling betreft waar normaal gesproken geen patiëntjes liggen die in levensgevaar verkeren.” (zie Elffers (2005:135)).

(PM. Zie overigens Derksen (2006:22) voor een kritische bespreking van die afdeling en een overzicht voor een langere periode 1996-2000 waarin “normaal gesproken” wel degelijk meer gevallen blijken te bestaan. In de langere periode zijn er bij V’s aanwezigheid juist minder incidenten. D neemt hier een beslissing dat een eerdere periode niet zou zijn mee te nemen omdat de data hierdoor “te heterogeen” worden. Als alternatief kan men stellen dat de data over de korte periode al zo heterogeen zijn dat de verlenging van de periode juist wel gewenst is. Dit zijn punten die in het onderzoek naar heterogeniteit zijn mee te nemen en die V juist kunnen vrijpleiten van verdachtmaking.)

D’s definitie van incident spoort echter niet met de gebruikte methode. Volgens Van Zwet (2004): “(...) is het van essentieel belang dat de rechtbank uitzoekt of, en zo ja, hoeveel moorden er zijn gepleegd en op wie, en wie de groep van mogelijke daders vormen. Dit dient met de daartoe geëigende juridische middelen te geschieden, waarbij de statistische analyse van de dienstrooster gegevens geen rol behoort te spelen. Als niet juridisch bewezen is dat er moorden zijn gepleegd, is de vraag naar de schuld van LdB niet aan de orde. Wie Agatha Christie’s *Murder on the Orient Express* heeft gelezen, zal trouwens ook graag de mogelijkheid van multiple daders door de rechtbank zien uitgesloten. Indien aan deze voorwaarden is voldaan, is de analyse van Elffers overtuigend. Het model is goed te verdedigen en de wiskundige aanpak volgt dan vanzelf en is ook voor leken eenvoudig te begrijpen. Zonder een uitspraak van de rechter over deze zaken is er reden te twijfelen aan de validiteit van de door Elffers gehanteerde methode.”

De methode van D heeft alleen kracht wanneer “incident” staat voor moord. Wanneer D het begrip “incident” ruimer kiest dan wordt niet alleen het probleem van data-selectie groter maar verliest de methode ook aan beslissingskracht, doordat de methode alleen nog maar toestaat een suggestie te doen voor wie “verdacht” is of welke verpleegster “niet-normaal” is (en dus niet een beslissing wie een moordenaar is). Zo’n statistische conclusie zegt nog weinig omdat men slechts op grond van additioneel medisch en forensisch onderzoek moet bepalen of de verdenking ook hout snijdt.

Van Zwet wil niet alleen zien dat vooraf bepaald wordt of er moord is maar ook dat de rechtbank eerst zelf de verdachten aanwijst op grond van andere dan statistische inzichten. In dit geval heeft de rechtbank (of D met goedkeuren van de rechtbank) de verplegers als doelgroep gekozen en betekent de ruime definitie van “incident” dat de methode alleen een hoofdverdachte oplevert – waarna dan nader medisch en forensisch onderzoek dient te volgen om feitelijke schuld vast te stellen. Omdat de redeneervolgorde dan andersom ligt is dan extra zorg nodig dat men het bewijs niet plooit naar een gewenste richting.

D past ook een zgn. “post hoc” correctie toe. Deze methode corrigeert niet voor het bestaan van verkeerde data maar heeft alleen te maken met de mogelijkheid dat de “incidenten” in het JKZ aanleiding waren tot verdenking zodat ze niet gebruikt kunnen worden als bevestiging van de verdenking. Bij de ruime definitie van “incident” is de toegepaste “post-hoc” correctie echter zinloos omdat voor het bepalen van wie “verdacht” is er nauwelijks een bezwaar is om de oorspronkelijke gevallen die tot verdenking aanleiding gaven ook gewoon mee te tellen.

De ruime definitie van het begrip “incident” en de daarbij verkeerd gebruikte methode dekken derhalve niet de stellige conclusie: “Geacht hof, het was geen toeval. De rest is aan u.”

D is consistent in de zin dat zijn conclusie betrekking heeft op “incidenten” (in ruime zin) en het is dus niet zo dat hij stellig zou concluderen tot moord of seriemoord. Maar zijn methode maakt het niet mogelijk om toeval geheel uit te sluiten – zoals hij wel doet. Het is voor een statisticus noodzakelijk om helder aan te geven dat het toeval nog bestaat – omdat zijn uitspraak anders verkeerd kan worden uitgelegd.

Van Zwet (2004) geeft de juiste conclusie – maar geeft ietwat wonderlijk ook weer iets te veel ruimte aan D – mogelijk vanuit gevoelens van collegialiteit: “De statisticus kan hoogstens besluiten dat LdB significant te vaak aanwezig was als er patiënten om niet verklaarde redenen gereanimeerd werden of overleden, maar daarvoor zijn ook andere verklaringen dan moord of doodslag te geven. Elffers is zich hiervan terdege bewust en beschrijft omstandig de rol die hij als statisticus speelt. Alleen al de titel van zijn bijdrage: ‘Geacht hof, het was geen toeval. De rest is aan u’ laat weinig aan duidelijkheid te wensen over. Toch zou ik voor mijn verdere betoog wat preciezer met dit feit willen omgaan dan zijn mededeling aan het hof: ‘Zoek het verder zelf maar uit.’ Of er sprake is van moord – en daarmee bedoel ik voortaan (poging tot) moord of doodslag - dient inderdaad door de rechter op andere gronden te worden beslist.”

“Significant te vaak” betekent hier dat men met vermoedelijk meer succes naar systematische factoren kan gaan zoeken dan denken dat het allemaal maar toevallig is. Maar het sluit toeval nog steeds niet uit. Van Zwet schrijft ook voorzichtig “weinig aan duidelijkheid” en niet “niets aan duidelijkheid”.

PM. Men kan het ook anders formuleren. Sommigen menen dat Van Zwet de verkeerde vraag zou beantwoorden. Een kritische lezer van het artikel van Van Zwet stelt: “De juridische vraag was in hoeverre al deze incidenten erop wezen dat er geen sprake meer was van toeval. Als men op andere gronden al had vastgesteld dat deze incidenten moorden waren, dan had men de statisticus echt niet meer nodig gehad in deze zaak.” Dit is een misverstand. Lezing van het artikel van Van Zwet verheldert dat juist hij zich hiervan bewust is. Hij verheldert (a) de vraag (moorden of toeval), (b) de methode. Vervolgens laat hij zien dat de methode van D past bij de vraag “wie deed het” en niet de vraag naar het toeval van aanwezigheid. De redenering van Van Zwet is: Als er moorden zijn, wie heeft het gedaan? Daar kan de statistiek dan iets over zeggen. Als het geen moorden zijn, dan moet je als statisticus zorgen dat de rechter niet met je conclusie aan de haal gaat, en van “statistisch significante” aanwezigheid gaat concluderen tot moord. Van alle statistici die naar deze kwestie hebben gekeken heeft juist Van Zwet deze omgekeerde vraag gesteld en daarmee het beste verhelderd dat het niet slechts een kwestie van waarschijnlijkheden is maar dat je de verliesfunctie niet mag verwaarlozen (die aanleiding moet zijn om het significantie-niveau te kiezen).

7. De statisticus als poortwachter

Wanneer het om moorden zou gaan dan maakt het onderscheid in soorten diensten (dag/nachtdiensten) weinig uit (zoals Van Zwet ook stelt). Maar wanneer het gaat om “incidenties” dan ligt dit anders. In dat geval zou statistiek iemand niet alleen als verdacht kunnen aanmerken maar ook minder verdacht kunnen maken. Met name dat laatste kan voor de verdediging relevant zijn wanneer er geen concrete bewijzen zijn.

In het rapport voor de rechtbank valt op dat D niet vraagt naar aanvullende data zoals het aantal gevallen per verpleger, de kans op *malpractice* door dokters, data van anderen dan alleen het OM. D wijst voor de rechtbank het gebruik van epidemiologie af.

Hij herhaalt die visie in Elffers (2005:142-143) en gebruikt dan het voorbeeld van de wiegedood van de kinderen Clark als een “waarschuwing (...) tegen het onoordeelkundig gebruik van de epidemiologische vergelijkingsmethode”. Dit is een open deur omdat weinig personen voor “onoordeelkundig gebruik” zullen zijn. D’s algemene stelling is onhoudbaar: “In het algemeen ben ik daarom geen voorstander van de epidemiologische vergelijkingsmethode: de interpretatieproblemen zijn onoverkomelijk.” Deze visie zal door oordeelkundige epidemiologen worden afgewezen.

Bijvoorbeeld zou het eventueel zo kunnen zijn dat het aantal incidenten in de twee ziekenhuizen lager is dan in de rest van Nederland en conformeert het aantal bij V meer met het landelijke gemiddelde ? Bijvoorbeeld dient men rekening te houden met fouten die dokters kunnen maken en het belang die dezen hebben met het afschuiven van schuld op verplegers. Vooreen zaak als deze zou men toch ook graag de context willen weten hoe vaak missers door dokters voorkomen en hoe dit ligt voor JKZ en RKZ. Ook hier zal medisch en forensisch bewijs uiteindelijk de doorslag moeten geven. Bij het bepalen of een persoon als “verdacht” kan worden betiteld zijn deze epidemiologische inzichten voldoende relevant om die bij het statistisch oordeel te betrekken.

Wanneer een rechtbank vertrouwt op een expert als poortwachter voor een breed onderzoeksterrein, zoals in dit geval de data-analyse, dan is dat niet iets wat de rechter expliciet hoeft te vragen maar de deskundige zou altijd uit eigen verantwoordelijkheid kunnen wijzen op andere benaderingen die licht op de zaak kunnen werpen.

8. Het optellen van onderdelen

De statistische expertise is in dit gevallen met name waardevol wanneer het OM en de rechtbank geholpen worden bij het vermijden van subtiele fouten.

Om te verhelderen dat de kwestie subtiel ligt nemen we een voorbeeld van Saari (2001) en vertalen dat naar een medische situatie. Laten er in een ziekenhuis twee Secties zijn, met gelijke aantallen van 30 personen. Van de patiënten is bekend of zij wel of niet een ziekenhuisbacterie hebben (Ja of Nee). Er is een nieuwe test gevonden waardoor je sneller zou kunnen vaststellen of een patiënt inderdaad die bacterie heeft. Die test lijkt bruikbaar voor alle nieuwe gevallen en andere Secties. Maar is de nieuwe test betrouwbaar ? De test wordt uitgeprobeerd op de Secties, met positief of negatief resultaat.- Tabel 4 bevat in onderdeel A de testresultaten, met ook het totaal van de twee Secties. Onderdeel B bewerkt de testresultaten tot een betekenisvolle samenvatting. Een positieve uitslag geeft een andere kans op het ontdekken van de bacterie dan een negatieve uitslag. Een maatstaf voor de kwaliteit van een test is het relatieve risico, de verhouding tussen genoemde twee kansen. Het relatieve risico drukt dus uit hoeveel meer kans je hebt als je een positieve uitslag hebt vergeleken met wanneer je een negatieve uitslag hebt. Het relatieve risico blijkt in beide Secties groter dan één, zodat het middel dus lijkt te werken. Gebruiken ? Nee, uit de optelling van de Secties blijkt het relatieve risico juist onder de 1 te liggen. Juist een negatieve uitslag voorspelt beter dat je de bacterie hebt ! Inderdaad laten de betrouwbaarheidsintervallen (BI) zien dat er nogal wat spreiding is. Conclusie: goed oppassen wanneer data in clusters komen. Deze kwestie staat bekend als de Simpson paradox.

Tabel 4. De Simpson paradox

De optelling van twee Secties leidt tot een tegengestelde conclusie

(A)	Sectie 1			Sectie 2			Totaal		
	Ja	Nee	Som	Ja	Nee	Som	Ja	Nee	Som
Pos.	9	15	24	3	3	6	12	18	30
Neg.	2	4	6	11	13	24	13	17	30
Som	11	19	30	14	16	30	25	35	60

(B)	Relatief risico (RR)	RR > 1 ?	Ondergrens BI	Bovengrens BI
Sectie 1	1.125	Ja	0.759	1.667
Sectie 2	1.091	Ja	0.818	1.455
Totaal	0.923	Nee	0.763	1.116

Keren we terug naar de onderhavige zaak en laten we constateren dat de data worden aangeleverd voor drie ziekenhuisafdelingen. Hoe is dit statistisch verwerkt en wat is de juiste methode? Voor de drie ziekenhuis-afdelingen apart berekent D de kans op het voorkomen van de aantallen gevallen, en deze drie aparte uitkomsten (*p-values*) worden vermenigvuldigd (met een post-hoc correctie). De uitkomst van deze berekening wordt in de uitspraak van de rechtbank aangehaald:

“De rechtbank te ‘s-Gravenhage veroordeelt Lucia de B. tot een levenslange gevangenisstraf voor de moord op vier patiënten en pogingen tot moord op drie patiënten. De rechtbank heeft de uitkomst van een waarschijnlijkheidsberekening mee laten wegen. Uit die berekening bleek dat de kans dat een verpleegkundige bij toeval bij zoveel overlijdensgevallen dan wel reanimatie-incidenten betrokken is, uiterst gering is (1 op 342 miljoen).” (samenvatting op www.rechtspraak.nl) PM. De berekening van D ging overigens uit van 14 “incidenten” en niet 7 (pogingen tot) moorden.

Van Lambalgen en Meester (2004) en Meester e.a. (2006) bekritisieren de methode van de vermenigvuldiging. Deze vermenigvuldiging vermengt het oorspronkelijke probleem met de kans op zo’n afdeling te werken. Derksen (2006) geeft het terechte tegenvoorbeeld van een hypothetische collega W die gedurig op één afdeling had gewerkt en die eenzelfde totaalresultaat had gekregen als V. Bij opgetelde data (zoals bij de Simpson paradox) en verdere toepassing van de methode van D vinden we (hypergeometrisch zonder post-hoc correctie) de kans van 1 op de 4,3 miljoen bij de oorspronkelijke “data” (een factor 100 verschil met 1 op 342 miljoen) en 1 op de 59 (geen miljoen) voor de “gecorrigeerde data”. Het scheelt. Wellicht zijn deze uitkomsten voor een rechtbank ook minder afschrikkend.

(PM. Genoemde totaalsom is overigens reeds gebruikt in Tabel 3 maar dan voor een andere vraag, “minstens X” (en een binomiale benadering) en niet “precies X”.)

9. Communicatie met de vakgenoten en het publiek (1)

In geruchtmakende zaken waarin ook de bijdrage van experts een rol gaat spelen is de communicatie met het publiek van belang. Strafzaken dienen niet slechts de berechting van daders, hetgeen eventueel in het geheim zou kunnen gebeuren, maar openbaarheid draagt ook bij tot het maatschappelijk moreel besef en het gevoel van rechtvaardigheid dat misdadigers hun straf niet ontlopen. Wanneer er vragen zijn omtrent de bijdrage van experts dan is het loffelijk dat deze in het openbaar verheldering geven. Het publiek moet dan wel juist worden voorgelicht.

Waar vakgenoten inmiddels kritiek uitoefenden op de rol van de statistiek in deze strafzaak kwam D niet alleen in de rol om e.e.a. toe te lichten maar ook in de positie dat hij aanleiding zag zich te verdedigen. Het is nuttig om drie punten te noemen waar dit niet zuiver gebeurt.

D beperkt zich tot de kwestie van het JKZ, zie Elffers (2004ab, 2005). Voor het JKZ geldt: (a) het JKZ heeft een extreme uitkomst van 8 versus 0 “incidenten” zodat V heel schuldig lijkt, (b) er zijn dus ook geen “overklaarbare” incidenten / moorden van andere verplegers, (c) de casus bevat de “post-hoc” correctie ten gunste van V zodat D correct overkomt. Het voordeel van een enkele afdeling is ook dat de vermenigvuldiging niet besproken hoeft te worden. Selectie van slechts het JKZ heeft dus het effect dat de bijdrage van de deskundige zo gunstig mogelijk over het voetlicht komt (en gezien de conclusie: voor V zo ongunstig mogelijk).

Vanuit het oogpunt van statistische verantwoording voor de vakgenoten is dit op zijn minst onvolledig. In Elffers (2004b) stelt hij “Ik beperk mij hier even tot de ‘kop’ van de zaak, de incidenten in het Juliana Kinderziekenhuis, omdat daarop reeds de gehele argumentatie naar voren komt.” Elffers (2004b) verwijst naar (2004a) dat eerder gepubliceerd werd. Dit is niet geheel juist, want reeds in Elffers (2004a) beperkte hij zich *ook* tot het JKZ. Elffers (2004b) is bovendien bedoeld als antwoord op Van Lambalgen en Meester (2004) – maar hun argument tegen de vermenigvuldiging van *p-values* krijgt derhalve geen antwoord omdat hij zich tot het JKZ beperkt. D bespreekt die laatste kwestie vervolgens op geen enkele andere plaats. Men kan het als een *cover-up* beschouwen dat D zich beperkt tot alleen het JKZ die V zo slecht mogelijk voorstelt. Op deze wijze komt er niets van de maatschappelijke verantwoording terecht, blijft het publiek gedesinformeerd en kan het gevolg ook zijn dat men gaat denken dat statistiek alleen maar tot verwarring leidt.

Sommigen laten zich daardoor beïnvloeden. Inderdaad neemt Van Zwet (2004) de getallen van alleen het JKZ over. Terwijl ook de oorspronkelijke “data” over RKZ aanleiding geven om die JKZ-“data” dubieus te achten. Het is voortreffelijk van Van Zwet (2004) dat hij toch tot een statistisch juist oordeel komt.

Hiermee is overigens niet gezegd dat D opzettelijk een *cover-up* pleegt en bewust de zaken verkeerd voorstelt. Dat wordt hier niet betoogd. De integriteit van D staat hier niet ter discussie en om te vermijden dat men zulks zou denken kan deze met nadruk onderschreven worden. Een positief punt is bijvoorbeeld dat denkbaar juist doordat D de kwestie in de publiciteit is blijven bespreken hij geholpen heeft dat er aandacht voor is gebleven. Zou D werkelijk de feiten hebben willen verdoezelen dan zou een betere strategie zijn geweest om geheel over de kwestie te zwijgen. Beroepsgroepen met een tuchtcommissie zijn hier mogelijk beter bedreven in. Waar hier van een ‘*cover-up*’ gesproken wordt bestaat deze eruit dat het uiteindelijke effect wordt beschreven, en past de constatering dat D werkelijk denkt dat deze selectie van feiten adequaat is terwijl deze dat niet is, en dat hij de kritiek daarop niet hoort.

10. Communicatie met de vakgenoten en het publiek (2)

Het tweede voorbeeld van onzuiverheid is dat D stelt: “De hele commotie rond het statistisch bewijs lijkt mij deels ingegeven door de foutieve gedachte dat de statisticus de rol van de rechter overneemt. Dat is niet zo: hij reikt een bewijsmiddel aan. Niet meer, niet minder” (Elffers (2004b)). Er is echter geen enkele statisticus die zulks beweerd of gesuggereerd heeft. Door anonieme anderen zulke oordelen in de schoenen te schuiven kan men zelf de vermoorde onschuld spelen, de aandacht afleiden van werkelijke kritiek waar men niet op ingaat.

Als derde voorbeeld van onzuiverheid vinden we in Elffers (2005) een langere bespreking van de verschillende “scholen” in de statistiek. Dit brengt wellicht een nuance aan die in het rapport aan de rechtbank niet te vinden is maar doet verder ook weinig terzake, omdat D niet die statistici noemt die zichzelf “eclectisch” vinden, die niet aan schoolvorming doen, en die bij een probleem de statistische technieken gebruiken die terzake relevant zijn. Elffers (2005) is opgenomen in een

boek dat vooral voor studenten is bedoeld zodat uitwijdingen over verschillende scholen in de statistiek terzake kunnen lijken. Dat kan inderdaad. Maar zulke uitwijdingen bestaan al op vele plaatsen en het zou voor de studenten veel leerzamer zijn geweest om iets te lezen over het vermenigvuldigen van *p-values*.

11. Keuze van de statistische decisieleer

Voor de beantwoording van de vraag “is het toeval?” kiest D voor de statistische decisieleer. Hierin gebruikt men een kansmodel om een beslissing te onderbouwen. Maar, het gebruik van een kansmodel geeft bij voorbaat een positief antwoord dat toeval een rol speelt. Het zijn juist niet-statistici die toeval moeten uitsluiten. De conclusie van D “het is geen toeval” die hij de rechtbank meldde is een non-sequitur op zijn aannames.

Worden de tabellen met de gegevens over de afdelingen gezien als uitkomsten van toevallige trekkingen met homogene (uniforme) kansen dan volgt daaruit een kans dat een verpleger bij nul incidenten aanwezig is 3,5% in de oorspronkelijke “data”, wat betekent dat de kans op 1 of méér 96,5% is, wat zorgwekkend is als het moorden zijn. De kans alle 27 mee te maken is niet nul maar 10 tot de macht -26 . Alleen de kans op 28 of meer is nul, want er waren er maar 27 gevallen.

Een begripsmatig probleem in dit geheel is dat de term “toeval” meerdere vertalingen kent en niet eenduidig is. Een gangbare uitleg is inderdaad die vertaling met homogene kansen; maar ook bij heterogene kansen (de ene uitkomst heeft meer kans dan de andere) bestaat er nog steeds toeval (alleen wat anders verdeeld). Een kernprobleem in deze zaak is dat de “leken-vraag” “is het toeval?” vertaald moet worden naar een statistisch model, en het resultaat weer terugvertaald. Kernpunt blijkt echter te zijn dat zodra je modellen met kansen gaat hanteren er slechts oordelen in termen van kansen mogelijk zijn zodat je niet goed meer kunt zeggen “het is geen toeval”. Je kunt wel berouw krijgen dat je een model met kansen bent gaan hanteren, maar dat berouw kun je niet baseren op de uitkomsten van het kansmodel. Zulk berouw is gebaseerd op het inzicht dat “statistiek brengt ons niet verder” en de basis voor de juistheid daarvan ligt in de aard der zaak en de klaarblijkelijke adequaatheid van de alternatieve deterministische aanpak.

12. Omgaan met statistische decisieleer

Het klassieke leerstuk van de statistische decisie is op zichzelf eenvoudig uit te leggen. Bij het schrijven van dit artikel is ernstig overwogen geen poging te doen om dit nader te bespreken. Al was het maar wegens “In der Beschränkung zeigt sich der Meister”. Aber doch. Een beperking tot enkele essentialia blijft een beperking. Een jurist met weinig affiniteit met getallen doet er ook verstandig aan om dit onderdeel te negeren. Echter, voor een goed begrip van de kwestie is het wenselijk toch ook dit aspect te bespreken.

Een industriële toepassing geeft denklijk het beste voorbeeld. Laat een fabrikant lampen maken met een gemiddelde van 2000 branduren. De fabriek weet uit ervaring hoe de kansverdeling rondom dat gemiddelde ligt. Dit is de status quo. Vervolgens komt de afdeling R&D met een voorstel voor een nieuw productieproces. In een proefopstelling worden lampen gemaakt die langer lijken mee te gaan. Voordat de gehele fabriek omgebouwd wordt wil men meer zekerheid hebben dat de nieuwe lampen gemiddeld minstens 2100 uur zullen branden, waarmee de kosten van de verbouwing kunnen worden terugverdiend. De nul-hypothese is dat het gemiddelde uit de status quo niet verandert. De alternatieve hypothese is dan dat het gemiddelde in de proefopstelling groter is dan 2100 uur. Omdat hier sprake is van gemiddelden en alleen een steekproef kan worden getrokken uit alle mogelijke lampen die ooit in de toekomst geproduceerd gaan worden, is een statistisch model op zijn plaats. De besluitnemer loopt enkele risico's. Stel dat de nulhypothese waar is (de nieuwe lampen branden niet langer dan zoals in de status quo) maar hij of zij kiest toch voor de alternatieve hypothese en het ombouwen van de fabriek: de kans dat dit gebeurt heet kans *alfa*, en wordt het significantieniveau genoemd. Wanneer bijvoorbeeld het significantieniveau 5% is dan zal men bij 100 beslissingen in 5 gevallen een ware nulhypothese ten onrechte verwerpen. Stel dat de alternatieve hypothese waar is (de nieuwe lampen branden inderdaad langer) maar de besluitnemer kiest toch niet voor het ombouwen van de fabriek: de kans hierop wordt *beta* genoemd. Tabel 5 vat de situatie samen. Op grond van het kostenplaatje en de daarbij behorende risico-overwegingen kan de fabrikant de omvang van deze kansen *alfa* en *beta* kiezen. Men is in principe vrij om *alfa* te kiezen maar een lage *alfa* veroorzaakt een hoge *beta* en het gaat dan om het evenwicht. Met de kansverdelingen en de waarden van *alfa* en *beta* kan een steekproefomvang worden bepaald. Op grond van de uitkomst van deze steekproef kan dan een besluit worden genomen met dit gegeven significantieniveau.

Enkele belangrijke principes zijn dan: (1) eigenlijk dient men *alfa* en *beta* voorafgaand aan de steekproef te kiezen, om oneigenlijke invloed van een heimelijk gewenste uitkomst te vermijden; (2) natuurlijk zal de R&D afdeling met lampen blijven experimenteren en komt de vraag naar te kiezen waarden van *alfa* en *beta* later weer terug; men kan dan met afspraken werken om effecten als onder (1) te vermijden; maar als gevolg van een grotere kennis van de materie zou men denklijk op den duur een steeds lagere *alfa* kunnen zien; (3) zolang er geen fysisch bewijs is geleverd dat de lampen noodzakelijkerwijs langer moeten branden blijft alles een beslissing onder onzekerheid; (4) een vraag is of de aannames voor zo'n industriële toepassing zich laten vertalen naar de zittingszaal, en zo ja, hoe dan.

Tabel 5: Tabel van de klassieke statistische deciseleer

De onbekende werkelijkheid versus de feitelijke keuze		De ware situatie (onbekend wat het is)	
		Brandduur = 2000 uur	Brandduur = 2100 uur
Feitelijke keuze	Status quo (nulhypothese)	Goed gekozen	<i>Beta</i> = kans op deze fout
	Verbouwen (alternatief)	<i>Alfa</i> = kans op deze fout	Goed gekozen

Alfa is de kans op het ten onrechte verwerpen van een ware nulhypothese. *Beta* is de kans op het falen een onware nulhypothese te verwerpen. De ware situatie is hier de proefopstelling maar is eigenlijk de projectie wanneer de proefopstelling wordt vertaald naar een geheel nieuw productieproces.

Wat de statistische deciseleer moeilijk maakt is dat statistici vaak zo in hun modellen zitten dat ze het niet goed kunnen uitleggen. Dit is een voorbeeld. In Elffers (2004a, 2005) vinden wij zijn "vertaling" van "gewone-mensen-taal" naar statistische modellering. Dit is gereproduceerd in Tabel 6. "H0" is de nulhypothese die met een bepaalde significantie (d.w.z. kans *alfa*) verworpen kan worden ten gunste van de alternatieve hypothese. Verwerping van H0 betekent in de klassieke statistische deciseleer dus niet dat H0 onwaar is maar dat er een kans *alfa* is dat men *besluit* dat H0 onwaar is terwijl H0 in werkelijkheid toch waar is. Dit negeert deze *alfa* echter.

Tabel 6: D's vertaalschema

Met $H_0 =$ "V en incidenten zijn ongerelateerd" stelt hij:

Het is toeval dat V zoveel incidenten meemaakt (gewone taal) is equivalent aan H_0 geldt (statistische modeltaal)
--

En hij leidt af:

We verwerpen H_0 , dus H_0 geldt niet (statistische modeltaal) hetgeen equivalent is aan Het is geen toeval dat V zoveel incidenten meemaakt (gewone taal)
--

Vervolgens verklaart D: "Sommige statistici menen dat het beter is de conclusie te verwoorden als: het is buitengewoon onwaarschijnlijk dat verdachte zoveel incidenten zou meemaken als het voorkomen van incidenten en het dienstdoen van verdachte ongerelateerd zijn. Met die verwoording heb ik als zodanig geen moeite, maar ik vind het wel een beetje je terugtrekken op het vakjargon: 'ongerelateerd zijn' betekent immers precies dat dan het kansmodel van H_0 geldt, en ik heb er hier boven voor gepleit om dat kansmodel in huis-tuin-en-keuken-taal te beschrijven als "het was toeval". Zeg dat dan ook, want het is toch ook de taak van een deskundige om de betekenis van zijn vak-analyse te vertalen naar het strafproces in casu." (Elffers (2005:146))

Op deze manier wordt het klassieke leerstuk van de statistische decisie verkeerd uitgelegd. Het statistisch verwerpen van een hypothese betekent niet dat toeval is uitgesloten. Die verwerping gebeurt met een bepaalde kans dat de verwerping onterecht is – de *alfa*. Zelfs als D's berekening een "kans 1 op (bijna) oneindig" had opgeleverd dan nog had hij niet aan de rechtbank mogen rapporteren "het is geen toeval": want de redenering is juist op een kansmodel gebaseerd.

Bij de oorspronkelijke "data" is de door D aangehaalde maar verworpen formulering "het is buitengewoon onwaarschijnlijk dat verdachte zoveel incidenten zou meemaken als het voorkomen van incidenten en het dienstdoen van verdachte ongerelateerd zijn" juist wel een correcte formulering, en deze laat zich niet vertalen als "het is geen toeval" (zie sectie 11) maar op zijn hoogst als "het is waarschijnlijker dat de kansen heterogeen zijn". (En zoals bekend, en ook door D gemeld, kunnen er andere oorzaken zijn dan moord, zie ook de bespreking van de definitie van het begrip "incident".)

Kernpunt is dat D nadat hij gekozen had voor de statistische decisieleer bij de uitkomst van zijn berekeningen had moeten mogen ontdekken dat zijn model eigenlijk niet paste voor de vraag "is het toeval?". Dat hij er aanvankelijk voor koos is wellicht niet ongebruikelijk omdat het een statistisch klassieke aanpak is. Maar zeker toen die conclusie getrokken werd had hij moeten beseffen dat het een kans-uitspraak blijft, zodat toeval niet uitgesloten kan worden. Vervolgens is hij er nadien door de vakgenoten op gewezen. Met name Van Lambalgen en Meester (2004) hebben op alternatieve kansmodellen gewezen. Het is opmerkelijk dat D nadien niet heeft gecorrigeerd, terwijl het begrip "toeval" in deze strafzaak zo'n grote rol speelt.

PM 1. De denkfout van D is dus dat hij een equivalentie aanbrengt tussen de gedachte “het is toeval” en het statistisch model van slechts homogene kansen. Men kan bijvoorbeeld ook een equivalentie aanbrengen tussen een Porsche en een fiets – maar ook al zeg je dat, dan is het nog niet in werkelijkheid zo (ook al hebben ze allebei wielen). In dit geval kan de Officier van Justitie met de vraag “is het toeval?” ook impliciet de mogelijkheid van heterogene kansen hebben overwogen – ook al wist hij of zij niet dat hij of zij dit deed. Overigens kan ook een statisticus denken in “gewone mensen taal”, zodat dat onderscheid niet al te scherp is. Ook voor een statisticus heeft “het is toeval” gewone connotaties als voor andere mensen. De extra bijdrage van de statisticus schuilt erin dat hij of zij dergelijke connotaties kan verscherpen naar meer specifieke mogelijkheden. De gedachte “het kan ook toeval zijn” betekent voor de statistisch deskundige slechts het startpunt voor een zoektocht naar het juiste model. Er zal derhalve geen sprake zijn van zo’n equivalentie maar eerder van een hypothetische vertaling, die ook verworpen kan worden.

PM 2. Er is de vervolgvraag: wat is de juiste conclusie wanneer je *wel* accepteert dat er toeval een rol speelt? Hier speelt het grote probleem dat wellicht te weinig van de onderhavige kansen bekend is zodat een uitspraak met onvoldoende zekerheid gedaan kan worden. Weliswaar betekent het verwerpen van een hypothese met een bepaalde *alfa* dat de kans of een foute verwerping beperkt *behoort* te blijven tot *alfa*, maar, wanneer er onvoldoende bekend is hoe die kansen liggen (wanneer je data gemankeerd of onvolledig zijn, we weten veel minder over seriemoordenaars dan over lampen) dan blijft het een schot in de duisternis.

PM 3. Een aspect van dit geheel is dat de uitspraak “het is geen toeval” ook misbegrepen kan worden als “V is een dader”. Dit is een subtiele kwestie. Het blijkt dat D expliciet heeft aangegeven dat dit een verkeerde uitleg is. Hij heeft ook enkele andere mogelijke oorzaken genoemd, die in de uitspraak van de rechtbank inderdaad ook afgevinkt worden (met verder niet te bespreken nauwkeurigheid en doorwerkingen in de uitspraak voor het hof). En inderdaad is het zo dat de (oorspronkelijke) “data” niet uitsluiten dat V een dader is want de definitie van “incident” die hiervoor gekozen is sluit moord niet uit. Desalniettemin vindt Van Zwet (2004) dat de statisticus een bepaalde verantwoordelijkheid behoudt om te voorkomen dat de cliënt met de conclusie aan de haal gaat. Ook al behoudt de rechter de eigen verantwoordelijkheid om een oordeel zich eigen te maken dan nog behoudt de statistisch deskundige de eigen verantwoordelijkheid om een goed advies te geven. Dat D expliciet heeft aangegeven dat de uitspraak “het is geen toeval” niet mag worden opgevat als “V is een dader” pleit hem aldus niet vrij van de kritiek van collegae dat “het is geen toeval” een te gevaarlijke samenvatting is omdat die te gemakkelijk misbegrepen kan worden. In het geval van V en de uitspraak van het hof kan deze gevaarlijke suggestie blijven doorwerken in de *conviction intime*. Hoe dit ook zij, belangrijk blijft dat “het is geen toeval” ook een *verkeerde* statistische conclusie is.

PM 4. Voor de vertaling van de industriële toepassing naar de zittingzaal laat zich nog het volgende opmerken. Ten eerste is er het onderscheid tussen verschillende soorten beslissingen. Voor juridische toepassingen zou men de nulhypothese standaard kiezen in de aanname dat iemand onschuldig is tenzij het tegendeel is bewezen. Met de keuze van $\alpha = 0$ wordt de kans uitgesloten dat een onschuldige veroordeeld wordt. Wanneer het model wordt toegepast op dienstrooster-data en de vraag ontstaat ‘of het toeval is dat iemand zoveel gevallen meemaakt’ is het de vraag of de kwestie van schuld en onschuld dan toch niet op de achtergrond een rol speelt. Zou men zulke connotaties kunnen vermijden dan blijkt de keuze van een *alfa* en *beta* hoe dan ook niet eenvoudig. Met circa 60000 verpleegkundigen en n (serie-)moordenaars in m jaar is wellicht een keuze $\alpha = n / (m \cdot 60000)$ denkbaar – maar ontstaat denkkelijk ook de plicht om al die personen en jaren te toetsen om de keuze van deze waarden te rechtvaardigen. Maar dit gehele probleem vervalst wanneer we vasthouden aan het principe dat zulke statistiek niet in de rechtszaal thuishoort maar slechts bij het OM om verdachten te selecteren die nader zijn te onderzoeken.

PM 5. De verschillende interpretaties van het begrip “toeval” leiden tot complex taalgebruik. Sommige statistici gebruiken dan niet “het is geen toeval (in brede zin)” maar zeggen “het is geen toeval (in enge zin)”. De gedachte is dan: “In je achterhoofd weet je dat je wel ongelijk zou kunnen hebben, maar dat is normaal in het leven, en bovendien weet je dat je niet heel vaak ongelijk zult hebben.” Andere statistici kunnen zo’n laatste formulering billijken voor industriële toepassingen maar wijzen dit af voor toepassingen waarin juist de *benefit of the doubt* cruciaal is.

13. Second opinion van S

De rechtbank heeft er verstandig aan gedaan de methode van D te laten controleren door een andere deskundige. S onderschrijft de methode van D. Daarop krijgt S eenzelfde kritiek van vakgenoten als D gekregen heeft: “In dit verband is het goed te wijzen op de verklaring van prof. mr. R. V. de Mulder als deskundige ter terechtzitting in eerste aanleg. Onder punt 9 van de bewijsoverwegingen in het vonnis van de rechtbank 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 blindedarm ontsteking tegenkomt.” (Lambalgen & Meester (2004))

S is niet in de wetenschappelijke discussie getreden. Hij lijkt het standpunt in te nemen dat het voldoende is dat hij de methode van D onderschrijft en dat D verder naar goeddunken handelt. Mogelijk betekent dit wel dat S niet alles onderschrijft wat D in het openbaar beschreven heeft. Wellicht ook is de verantwoordelijkheid van S minder dan die van D. Anderzijds kan men diens rol crucialer achten omdat het voor een rechtbank verschil uitmaakt of de eigen aangezochte deskundigen elkaar bevestigen of niet. Dat er buiten de zittingzaal nog B1 t/m Bn zijn die anders blijken te denken speelt dan vermoedelijk geen rol meer. Wanneer S zich dan de kritiek op D niet ter harte neemt en niet corrigeert impliceert dan echter wel dat die kritiek ook op hem van toepassing wordt. Maar wellicht heeft S de zaak verder niet gevolgd en is zo’n conclusie van “wie zwijgt stemt toe” te vergaand.

Dit kan een overweging zijn om in de toekomst bij kritiek de rol van de tweede, mogelijk marginaal toetsende, deskundige te herzien. Men kan niet dijken binnen dijken blijven aanleggen om de polder te beschermen tegen eventuele breuk van de hoofddijk, maar wellicht dat de figuur van een wetenschappelijk adviseur verbonden aan de gerechtelijke macht kan helpen, als aggregator, om het commentaar van externe deskundigen te verwerken.

14. Enkele nieuwe ontwikkelingen

(1) Op 19 januari 2007 was er een uitzending van NOVA (zie het persbericht in **Bijlage A**). Op een wat wonderlijke wijze is het lastig om uitspraken gedaan voor televisie adequaat te vertalen naar uitspraken op papier. Een relevant gegeven is dat D bij de “nieuwe gegevens” (d.w.z. de “gecorrigeerde data” van Derksen i.p.v. de oorspronkelijke “data”) anders zou hebben gerapporteerd dan hij deed. Een en ander is eigenlijk niet meer dan logisch. Kernpunt is echter dat reeds de oorspronkelijke “data” tot een correctie aanleiding geven, met name: (a) de data zijn onbetrouwbaar, (b) je kunt niet concluderen “het was geen toeval”.

(2) Ten behoeve van de Commissie Posthumus II heeft professor Derksen een rapport opgesteld met enkele data en berekeningen die een update vormen voor zijn boek waarnaar hier verwezen wordt. Het is niet gepast om hier in dit stadium op in te gaan maar wel kan melding gemaakt

worden van Derksen's stelling dat de hoofdconclusie van zijn boek ongewijzigd blijft. Het is onduidelijk of D en/of S een bijdrage aan deze commissie hebben geleverd.

(3) Recent is er een bijdrage van D aan de "Studiedag omgaan met statistisch bewijs in strafzaken" van 24 mei 2007. De bijdrage van Elffers (2007) is getiteld "Toevallig is statistisch bewijs heel gewoon bewijs". Bij navraag bij D blijkt de tekst niet op papier te staan. Volledigheidshalve is de beschikbare tekst uit de congresfolder in **Bijlage B** opgenomen. In deze folder stelt D: "Het probleem is niet zozeer dat juristen statistiek misinterpreteren, maar veeleer dat menig statisticus geen kaas heeft gegeten van het strafproces." Hier geeft D het probleem verkeerd weer. De kritiek van statistici op zijn werk betrof de statistiek en betrof niet het strafproces. De kritiek luidt ook dat hij niet corrigeert terwijl hij daartoe wetenschappelijk gehouden is. Een andere stelling van D is: "bij een eventuele herziening zal statistische argumentatie geen rol meer spelen". Dit lijkt prematuur. Het is bijvoorbeeld een observatie van De Mulder (second opinion, S) dat de statistiek nog wel een rol kan hebben gespeeld bij de vorming van de *conviction intime*. In zo'n geval kan het onvermijdelijk zijn dat de statistiek in de rechtszaal dient te worden gecorrigeerd om zulke belasting uit het verleden te verwijderen.

15. Wetenschappelijke ethiek

Statistische deskundigen voor een rechtbank vervullen niet alleen een professionele rol maar functioneren ook in de context van een wetenschappelijke ethiek. Statistici kennen niet zoals de medici, farmaceuten of accountants een broederschap met een tuchtcommissie, en er zijn geen regelingen zoals voor consumenten met een of andere gecentraliseerde klachtenregeling. De KNAW heeft een begin van een regeling ten aanzien van de wetenschappelijke ethiek maar die richt zich op onderzoek en niet op praktische toepassingen.

In dit geval is er wetenschappelijke kritiek op D en S uitgeoefend en zij corrigeren niet terwijl het voor de strafzaak wel relevant zou zijn. Men kan oordelen dat zij dan falen ten aanzien van de wetenschappelijke ethiek dat men dient te corrigeren. Maar het is een oordeel dat verder geen consequenties heeft.

Het lijkt niet onbelangrijk dat de gerechtelijke macht zich realiseert dat er beroepsgroepen van mogelijk aangezochte deskundigen zijn die zich minder aan de tucht onderwerpen dan men nu mogelijk veronderstelt.

Een dwaling van het recht

Hierboven hebben we de rol van de statistiek in deze strafzaak geëvalueerd aan de hand van een 15-tal punten en de kern van dit betoog beëindigt daarmee. Is er nog ruimte om iets te zeggen over de rechtsgang zelf? Terwijl een statisticus respect heeft voor de uitspraken van rechters – en vooral als dat respect gevoed wordt door de complexiteit van een zaak, met bijvoorbeeld medische, forensische en psychiatrische getuigenissen – kan tegelijkertijd een onafhankelijk oordeel blijven bestaan, zonder dat men op de stoel van de rechter gaat zitten. Voor een extern waarnemer is het bovendien relevant te zien hoe rechters zelf nog worstelen met de rol van de statistiek en het toeval.

In de uitspraak van de rechtbank (LJN: AF6172 d.d. 24 maart 2003) heette het nog dat "De rechtbank heeft echter alleen de feiten bewezen verklaard in die gevallen waarin geen andere medische oorzaak het incident kan verklaren én er concreet ander bewijs is". Een probleem daarmee is dat ook toeval nog steeds een medische oorzaak kan zijn.

In de uitspraak van het gerechtshof (LJN: AP2846 d.d. 18 juni 2004) heet het: “Alle in aanmerking komende natuurlijke doodsoorzaken konden worden uitgesloten. (...) Er is geen statistisch bewijs in de vorm van toevalsberekeningen gebruikt. Wel is voor het bewijs van belang geacht dat de in het JKZ gepleegde delicten een betrekkelijk korte periode bestrijken en de meeste delicten voornamelijk op een gewone verpleegafdeling hebben plaatsgevonden en hiervoor geen verklaring is gevonden.” Maar ook hier kan toeval de situatie verklaren.

In de uitspraak van de Hoge Raad (LJN: AU5496 d.d. 14 maart 2006) heet het in 3.2: “In twee gevallen heeft het hof vastgesteld dat de dood of het levensbedreigende incident het gevolg is geweest van vergiftiging. Verder was in alle gevallen sprake van een plotseling, medisch onverklaarbaar overlijden of incident, dat zich telkens heeft voorgedaan terwijl of kort nadat de verdachte op de desbetreffende afdeling dienst had.” En In 11.8: “Ten aanzien van de overige bewezenverklaarde levensdelicten geldt dat:

- A. sprake is geweest van een plotseling en onverwacht overlijden of levensbedreigend incident;
- B. sprake is geweest van een medisch onverklaarbaar overlijden of levensbedreigend incident waarbij alle in aanmerking komende natuurlijke oorzaken als boven iedere redelijke twijfel verheven kunnen worden uitgesloten;
- C. het overlijden of het levensbedreigende incident heeft plaatsgevonden op een moment dat de verdachte op de afdeling waar de desbetreffende patiënt lag aanwezig was.”

Maar in deze gevallen heeft de Hoge Raad de oorzaak van toeval niet uitgesloten. “Medisch onverklaarbaar” kan nog steeds “toeval” zijn (cq. dat toeval een medisch acceptabele verklaring kan zijn). Men zegt de “vorm van toevalsberekening” niet in de overweging te betrekken maar doet impliciet toch een beroep op het toeval cq. op de gedachte, nu geheel zonder bewijs, dat de aanwezigheid op het rooster *niet-toevallig* is. De redenering is “wij kunnen het niet verklaren dus u heeft het gedaan want u was er bij – en nu bent u erbij” heeft een nogal middeleeuws karakter terwijl in moderne tijden gangbaar toch schuld *bewezen* moet worden. Het “wettig bewijs” is alleen pro forma “wettig bewijs” omdat de Hoge Raad dan gesproken heeft maar het ontbeert een moderne grondslag van een materiële inhoud.

Derksen (2006) bespreekt de twee gevallen die de Hoge Raad nu bewezen acht en geeft grond te twifelen aan die stelligheid, niet alleen omdat het overlijden nog steeds toeval kan zijn maar ook omdat men juist kan gaan denken aan *malpractice* – met name dat de medici een verkeerd protocol voorschreven. Wat Derksen schrijft lijkt juist aannemelijker.

Vervolgens hanteert de Hoge Raad een “schakelbewijs” dat lijkt op statistiek, maar waarover geen statistisch oordeel is gevraagd. Dit “schakelbewijs” lijkt een *petitio principii* – het vooronderstellen wat bewezen moet worden – want men *zegt* slechts dat die andere gevallen relevant zijn maar men *bewijst* het niet. (Het zijn “incidenten” en slechts door de aanwezigheid van V worden het “moord(poging)en”.) Het lijkt dus belangrijk om de aanhalingstekens rondom het woord “schakelbewijs” te laten staan. Dit roept de vraag naar de dataverzameling wakker, hoe “incidenten” zijn gedefinieerd, en hoe men de kans bepaalt dat iets wat op iets anders lijkt ook hetzelfde is. Weliswaar is in deze rechtsgang de verantwoordelijkheid voor de dataverzameling klaarblijkelijk gelegd bij de recherche en de verhoren in de rechtszaal zelf, en ongetwijfeld hebben rechercheurs en rechters gehoord van verschijnselen zoals data-selectie en *confirmation bias*, doch het lijkt niet onredelijk te veronderstellen dat een kritische instelling van een deskundig statisticus kan helpen om meer helderheid aan te brengen, juist in een situatie waarin de juridische constructie van een “schakelbewijs” wordt overwogen of bewezen moet worden dat zulks toelaatbaar is.

In de huidige situatie van het Nederlandse rechtssysteem kan vooralsnog iedere verpleger, zeg verpleger “Jansen” die toevallig aanwezig is bij een of twee gevallen van *malpractice*, veroordeeld worden voor 3, 10 of 100 andere (toevallige) incidenten waarbij hij of zij slechts

aanwezig hoeft te zijn. In Tabel 1 zijn er nog 13 incidenten van andere verplegers. Heeft het OM voor hen onderzocht of zij niet verdacht kunnen worden voor hun incidenten ?

Voor de statistiek is V “statistisch onschuldig” in de zin dat in geen der gevallen statistisch bewijs gebruikt kan worden. Bij de twee gevallen waarbij de Hoge Raad V schuldig acht aan vergiftiging gebeurt dat op grond van medisch of forensisch bewijs, en dus niet-statistisch bewijs. Derksen (2006) maakt aannemelijk dat de Hoge Raad bij deze laatste twee gevallen dwaalt. Zo’n dwaling past bij het gehele beeld dat de rechtsgang schept. Met 14 of meer gevallen beginnen en er slechts 2 “echte” overhouden begint erop te lijken dat, waar er misschien altijd wel wat mis kan gaan, men altijd wel een stok kan vinden om een hond te slaan.

De rechtsgang kan niet overtuigen en het is te adviseren de kwestie nu eens goed over te doen.

Conclusies

Hierboven is de strafzaak van Lucia de B. bekeken vanuit het gezichtspunt van de statistiek en met name is gekeken naar de rol van de statistiek. Het oogmerk was te kijken of er voor de juridische context lessen konden worden getrokken. Daarbij ging het juist om de rol van de statistische deskundigen. Voor de inzichtelijkheid zijn de conclusies reeds in de inleiding aangekondigd. Er zijn drie punten echter die men zonder het voorgaande wellicht moeilijker had kunnen accepteren, en dat zijn:

(1) Statistisch gezien is Lucia de B. onschuldig. Dit is weliswaar een open deur waar rechtspraak en wetenschap verschillende werelden zijn maar het is nuttig te benadrukken dat wetenschappers daarbij ook hun eigen vakoordeel houden. Een statisticus zal het oordeel van een gerechtelijke instantie gangbaar zonder kritiek registreren maar in het onderhavig geval blijkt de statistiek op een verkeerde wijze gebruikt. Ook al stelt de rechter nu dat geen statistische *berekeningen* zijn gebruikt, dan toch zijn zulke (dwaal-) *redeneringen* wel toegepast en zijn de eerdere fouten niet weerlegd zodat zij wel degelijk nog een rol kunnen spelen.

(2) De rechtsgang kan niet overtuigen en het is te adviseren de kwestie nu eens goed over te doen.

(3) De door de rechtbank aangezochte statistisch deskundigen prof. dr. H. Elffers en prof. mr. R. V. de Mulder falen ten opzichte van de wetenschappelijke ethiek dat men dient te corrigeren wanneer daarop wordt gewezen.

Literatuur

Derksen, T. (2006), “Lucia de B. Reconstructie van een gerechtelijke dwaling”, Veen magazines, zie <http://www.luciadeb.nl/>

Elffers, H. (2002), “Verdeling reanimatie- en overlijdensgevallen in het Juliana Kinderziekenhuis en het Rode Kruisziekenhuis”, Nederlands Studiecentrum Criminologie en Rechtshandhaving, Leiden, 8 mei en “Nadere analyse” 29 mei

Elffers, H. (2003), “Bij toeval veroordeeld? Statistische analyse van dienstroosterdata in het strafproces”, NJB 78/34 p1812-1814

Elffers, H. (2004a), “Geacht hof, het was geen toeval. De rest is aan u.”, StatOr (2004), p4-11, <http://www.vvs-or.nl/stator.htm>

Elffers, H. (2004b), “Dit zeggen al die getallen eigenlijk! Antwoord aan Van Lambalgen en Meester”, Trema, september, p293-294

Elffers, H. (2005), “Dienstroosterdata in het strafproces”, p133-152 in Sjerps, M.J. en J.A. Coster van Voorhout(2005)

Elffers H. (2007), “Toevallig is statistische bewijs heel gewoon bewijs”, bijdrage aan de studiedag “Omgaan met statistisch bewijs in strafzaken”, slechts de congressfolder, zie **Bijlage B**, www.kerckebosch.nl, retrieved from source

Lambalgen, M. van, en R.W.J. Meester (2004), “Wat zeggen al die getallen eigenlijk ? De statistiek in het proces tegen Lucia de B.”, Trema, september, no 7. P286-292, <http://staff.science.uva.nl/~michiell/docs/LuciadeB040205.pdf>

Meester, R.W.J., M. Collins, R. Gill, M. van Lambalgen (2006), “On the (ab)use of statistics in the legal case against the nurse Lucia de B”, ingediend bij “Law, Probability and Risk,” <http://arxiv.org/abs/math.ST/0607340>, zie ook <http://www.math.leidenuniv.nl/~gill/lucia.html>

Rechtspraak (2006), <http://www.rechtspraak.nl/Actualiteiten/Dossiers/Lucia+de+B/>
Begindatum 24 maart 2003, Datum laatste toevoeging 13 juli 2006

Rijken van Olst, H. (1969), “Inleiding tot de statistiek. Deel III”, Van Gorcum

Saari, D.G. (2001), "Decisions and elections", Cambridge

Sjerps, M.J. en J.A. Coster van Voorhout(ed) (2005), “Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht”, Kluwer

StatOr (2004), “Thema. Statistiek in de rechtszaal”, juni, jrg 5 no 2, <http://www.vvs-or.nl/stator.htm>, met bijdragen van H. Oud, H. Elffers, M. Sjerps, W. van Zwet

Zwet, W. van (2004), “Niets nieuws onder de zon”, StatOr (2004) p21-28

Bijlage A: Persbericht NOVA, 19 januari 2005

From: Sabine Baak <Sabine.Baak@novatv.nl>
Date: Jan 19, 2007 10:39 PM
Subject: persbericht Lucia
To: usnova@tmo.blackberry.net
Cc: sfranken@bfkw.nl (...)

http://www.novatv.nl/index.cfm?ln=nl&fuseaction=videoaudio.details&reportage_id=4874

STATISTICUSELFFERS OVER NIEUWE CIJFERS IN ZAAK LUCIA DE B:

"ALS IK DIT HAD GEWETEN ZOU IK TOT EEN ANDER OORDEEL ZIJN GEKOMEN"

Statisticus Henk Elffers, de belangrijkste deskundige in de rechtzaak tegen de Haagse verpleegkundige Lucia de B, zegt vanavond in NOVA dat hij tot een andere conclusie zou komen, als de nieuwe statistische gegevens kloppen. Een getuige-deskundige die terugkomt op zijn verklaring omdat hij niet alle gegevens had. In de Puttense Moordzaak was het reden tot herziening door de Hoge Raad.

Lucia de B. Veroordeeld voor een levenslange gevangenisstraf omdat ziekenhuisdirectie en justitie wisten dat het geen toeval kon zijn dat zij aanwezig was bij zoveel onverwachte sterfgevallen. Statisticus Henk Elffers berekende al in het politiekonderzoek hoe klein de kans was dat het toeval was: 1 op 342 miljoen. Inmiddels zijn er op alle vlakken grote twijfels over de veroordeling van Lucia. Ook de berekening van Elffers ligt zwaar onder vuur, al speelt die letterlijk geen rol in het arrest van het Haagse hof.

Nu zijn er ook nog nieuwe gegevens beschikbaar, die verregaande gevolgen hebben voor het cijfer van Elffers. Een kans van 1 op 48 of zelfs 1 op 9 zou volgens statisticus Richard Gill waarschijnlijker zijn. Het gerenommeerde wetenschappelijk tijdschrift Nature schrijft deze week een artikel over de problemen met het gebruik van statistiek in de rechtszaal, met de zaak van Lucia de B als voorbeeld van hoe het niet moet.

Vanavond reageert Henk Elffers in NOVA voor het eerst op de kritiek op zijn statistische berekening. Hij erkent dat de nieuwe gegevens gevolgen hebben voor de betrouwbaarheid van zijn rapport. Elffers: "Andere gegevens, andere analyses, dat lijkt me duidelijk.... Als er andere gegevens waren geweest zou er natuurlijk een ander geval uitgekomen zijn. Dat staat buiten kijf". En: "Als deze gegevens inderdaad net zo hard zijn als de bestaande gegevens dan zou ik met een ander oordeel zijn gekomen". Elffers sluit zelfs niet uit dat hij uitkomt op de berekening van collega statisticus Richard Gill.

Mark Buchanan, journalist van Nature, vanavond in NOVA: "There's controversy over how the statistics were used, how the numbers were gathered and how it was interpreted by the Court. And how the statistics influenced the opinion of many observers".

Elffers geeft ook toe dat zijn presentatie van de cijfers voor verwarring heeft gezorgd. Elffers vergeleek in zijn oorspronkelijke berekening de kans op toeval op de afdeling van Lucia in het Juliana Kinderziekenhuis, met de kans op twee vergelijkbare andere afdelingen. "Elffers vermenigvuldigde die kansen en dat levert automatisch een kleinere kans op, ten onrechte. Het vermindert de kans op toeval maal 1000", zo beweren wetenschapsfilosoof Ton Derksen en statisticus Richard Gill. Derksen schreef een boek over de bewijslast tegen Lucia de B, en dat leidde uiteindelijk tot het lopende onderzoek van de Commissie Posthumus 2.

En, Elffers geeft voor het eerst toe dat zijn berekening en argumentatie wel degelijk een rol heeft gespeeld "... in de hoofden..." van de rechters bij het Haagse hof. Tot nu toe hield hij vast aan het feit dat nergens in het arrest letterlijk verwezen wordt naar hem of zijn berekening.

Bijlage B: Bijdrage van D aan de studiedag "Omgaan met Statistisch bewijs in strafzaken"

Verantwoord gebruik van getallen als resultaat van (forensisch) onderzoek

Donderdag 24 mei 2007

Jaarbeurs Congrescentrum (Beatrixgebouw), Utrecht

Studiecentrum Kerckebos, www.kerckebos.nl

Toevallig is statistisch bewijs heel gewoon bewijs

- Statistisch bewijs onderscheidt zich niet van ander bewijs
- De Nederlandse rechter heeft dienstroosterargumentatie bestempeld als onvoldoende als geen direct bewijs aanwezig is, en als overbodig als dat wel het geval is
- De Haarlemse crècheleidster BK is in drie instanties vrijgesproken na presentatie door het OM van een statistisch dienstroosterargument
- De Haagse verpleegster LdB is niet op grond van statistische evidentie veroordeeld, bij een eventuele herziening zal statistische argumentatie geen rol meer spelen
- Het probleem is niet zozeer dat juristen statistiek misinterpreteren, maar veeleer dat menig statisticus geen kaas heeft gegeten van het strafproces

Prof. Dr. H. Elffers

Onderzoeker aan het Nederlands Studiecentrum Criminaliteit en
Rechtshandhaving (NSCR)

Hoogleraar rechtspsychologie Universiteit Antwerpen