STAtOR

THEMA VEILIGHEID

periodiek van de VVS jaargang 12 nummer 3-4, november 2011

STAtOR

Jaargang 12, nummer 3-4, november 2011

STAtOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VVS). STAtOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operationele research. Verschijnt 4 keer per jaar.

Redactie
Joaquim Gromicho (hoofdredacteur), Ana Isabel Barros, Johan van Leeuwaarden, Mirjam Moerbeek, Gerrit Stemerdink (eindredacteur), Hilde Tobi. Vaste medewerker: Fred Steutel

Kopij en reacties richten aan
Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, telefoon 020-5986010, mobiel 06-55886747, <j.a.dossantos.gromicho@vu.nl>.

Bestuur van de VVS
Voorzitter:
prof. dr. Jacqueline Meulman <president@vvs-or.nl>
Secretaris:
dr. Irene Klugkist <bestuur@vvs-or.nl>
Penningmeester:
dr. Ad Ridder <penningmeester@vvs-or.nl>
Studentlid:
Maarten Kampert (Bsc) <student@vvs-or.nl>
Overige bestuursleden:
prof. dr. Fred van Eeuwijk (BMS), prof. dr. ir. Stan van Hoesel & dr. John Poppelaars (NGB), dr. Eric Cator (SMS), dr. Michel van de Velden (ECS), dr. Andries van der Ark (SWS).

Leden- en abonnementenadministratie van de VVS
VVS, Postbus 244, 6700 AE Wageningen, telefoon 0317 - 419572, fax 0317 - 421364, <admin@vvs-or.nl>.
Raadpleeg onze website over hoe u lid kunt worden van de VVS of een abonnement kunt nemen op STAtOR of op een van de andere periodieken.

VVS-website
www.vvs-or.nl

Advertentieacquisitie
Nikki Bisschop & Joren Brunekreef,
Lange Nieuwstraat 6, 3512 PH Utrecht, 06-55874175,
<adverteren.stator@vvs-or.nl>. STAtOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak
Pharos | M. van Hootegem & C. Oomen

Druk
Drukkerij Zoeteweij, Yerseke

Uitgever
© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

Inhoud

Geachte lezer,

Voor u ligt ons jaarlijkse themanummer, ditmaal een dubbelnummer dat handelt dat over 'veiligheid'. Veiligheid is een vast onderdeel van ons leven en is nu, meer dan ooit, een actueel onderwerp. Het is een breed begrip, hetgeen er begrijpelijkerwijze toe leidt dat dit nummer een zeer veelzijdige inhoud biedt.

Zo treffen we maar liefst drie artikelen aan over gevaren in de medische zorg. Karin de Booij laat ons zien hoe het gevolgde beleid invloed heeft op de lengte van een diagnosetraject, Bart Veltman en Arnoud Kuiper brengen de onvermijdelijke gevolgen van het terugdringen van het aantal spoedeisende eerstehulpposten in kaart (dit artikel sluit direct aan bij een actuele intentie van de overheid) en Joris van de Klundert kijkt naar risico van het overlijden als gevolg van medisch handelen.

Ook het beleid gevolgd door de gemeentelijke overheid heeft grote invloed op de veiligheid, maar wellicht ook – op een minder evidente wijze – op de perceptie van veiligheid. Tom de Leeuw laat ons zien hoeveel subjectiviteit zich achter objectieve criteria kan verschuilen.

De Waterwet schrijft voor dat de dijkbeheerders iedere 6 jaar toetsen of de primaire Nederlandse waterkeringen voldoen aan de veiligheidsnormen. Deze waterkeringen moeten bestand zijn tegen extreme belastingen vanuit Noordzee, IJsselmeer, Rijn of Maas. Robin Nicolai, Ton Vrouwenvelder, Karolin Wojciechowska en Henri Steenbergen laten zien hoe de hydraulische belasting op de waterkering berekend kan worden, daarbij rekening houdend met veel onzekerheden.

Ook onze dagelijkse winkelbezoeken zijn mogelijke momenten van gevaar, en dan bedoelen we niet het gevaar van onnodige impulsaankopen. Recente calamiteiten in winkelcentra zijn helaas het bewijs van écht gevaar dat kan optreden. Daarom, maar ook om over de eigendommen van de winkelier te waken, is toezicht door beveiligingscamera's nodig. Een probleem is dan of het mogelijk is al die visuele informatie tijdig genoeg te analyseren om adequaat te kunnen reageren. Léon Rothkrantz laat ons zien wat er mogelijk is om dit proces te automatiseren.

Bij de opsporing buigt de politie zich over zaken als de vermissing van een persoon, de vondst van een levenloos lichaam, de opsporing van een gezochte verdachte, enzovoort. Bram Wisse, Sicco Pier van Gosliga en Gerard Bijsterbosch laten zien hoe een statistische benadering tijdens het opsporingsproces ondersteuning kan bieden bij de gelijktijdige beoordeling van meerdere mogelijke hypothesen. En om in de sfeer van de misdaad te blijven: hoe zit het nu eigenlijk met die omstreden geuridentificatieproef, waarin speurhonden van de politie verdachten van een misdrijf aanwijzen? Geurt Jongbloed en Frank van der Meulen vertellen u alles hierover! Ronald Meester behandelt de statistische aspecten van DNA-bewijs, daar zit meer aan vast dan menigeen denkt.

Dan ten slotte de nog-niet-begane misdaad. Analyse van sociale media kan een waardevolle indicatie bieden voor mogelijke dreigementen, zoals Eric Boertjes, Almerima Jamakovic en Stephan Raaijmakers ons laten zien.

Wij wensen u een veilig gevoel toe bij het lezen van dit nummer maar dan wel met de nodige spanning, veroorzaakt door de wetenschappelijke nieuwsgierigheid die de STAtOR-lezer eigen is!

Uw STAtOR redactie

illustratie: Celine Oomen

OVERLEEF IK MIJN DIAGNOSTISCH TRAJECT?

Planningssystemen in de zorg: een fragiel evenwicht

Karin de Booij

Goede zorg begint met het snel stellen van een goede diagnose. Als in een vroeg stadium wordt ontdekt wat er aan de hand is, kunnen behandelingen vaak effectiever, minder zwaar en minder duur zijn. Tijdens het Congres Operations Research in de gezondheidszorg in 2009 – georganiseerd door de Erasmus Universiteit Rotterdam – werd door aanwezigen juist de lengte van het diagnostisch traject als zorgelijk gezien, waarbij met name de radiodiagnostische scans als de bottleneck werden aangewezen. In Nederlandse zorginstellingen zijn toegangstijden van weken tot maanden voor CT- en MRI-scan geen uitzondering (www.kiesbeter.nl). Daarbij komt dat patiënten vaak meerdere keren terug moeten komen naar het ziekenhuis.

Op dit moment worden verschillende methoden toegepast om enerzijds de doorlooptijd van het diagnostisch traject te verkorten en anderzijds het aantal ziekenhuisbezoeken te verminderen. Maar hoe effectief zijn die methoden eigenlijk (op de lange termijn en voor de gehele patiëntenpopulatie) en zijn ze verder te optimaliseren? In mijn afstudeeronderzoek heb ik drie veelbelovende methoden verder onder de loep genomen en gekeken hoe deze methoden geanalyseerd, en waar mogelijk geoptimaliseerd, kunnen worden. Opvallend zijn de grote verschillen tussen de methodes, een oorzaak van het fragiele evenwicht van planningsalgoritmen in de zorg.

Methode 1: Inloop

De meeste planningssytemen voor MRI en CT scan maken gebruik van een afsprakensysteem waarbij slots volgens een *first come first served* basis worden gevuld. Ze leiden onder een slecht voorspelbare scanduur, patiënten die niet komen opdagen en de noodzaak van het reserveren van slots voor een van te voren onbekend aantal spoedpatiënten. In tegenstelling tot het traditionele afsprakensysteem is het idee van inloop waarbij patiënten na een polikliniek bezoek direct plaats kunnen nemen in de wachtkamer van bijvoorbeeld MRI- of CT-scan. Als alle diagnostische afdelingen inloop zouden toestaan, zouden theoretisch gezien alle onderzoeken op dezelfde dag uitgevoerd kunnen worden. De vraag is echter in hoeverre inloop de wachttijden in de wachtkamer, de bezettingsgraad van de scans en de gemiddelde overwerktijd beïnvloedt.

In Matlab zijn verschillende scenario's van inloop gesimuleerd op basis van *discrete event simulation*. Uitgangspunt in het onderzoek is de MRI-afdeling van het VU medisch centrum in Amsterdam, maar het principe is generiek. De simulatie is gebaseerd op de volgende input:

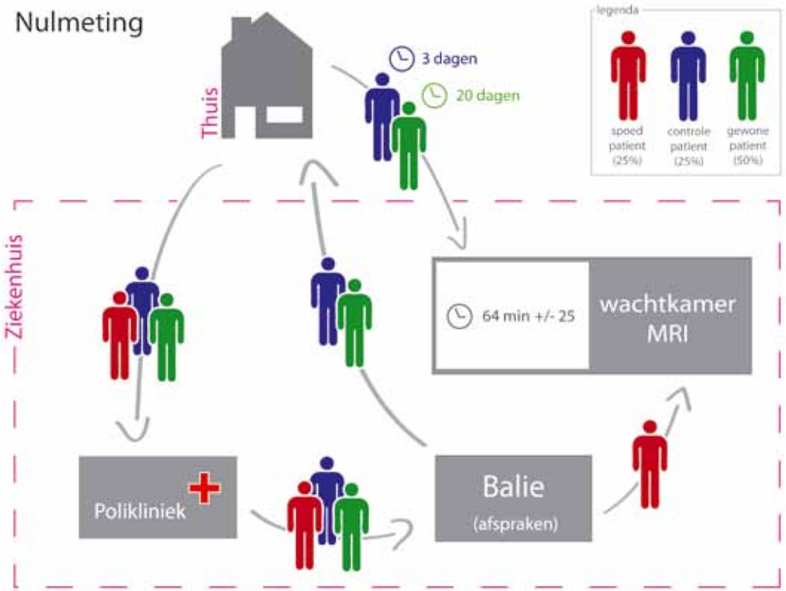
- aanvragen: inhomogeen Poisson proces met per uur en per dag een andere gemiddelde aankomstfrequentie tussen 3 en 12 patiënten per uur;
- scanduur: Pareto-verdeling met gemiddelde 47 minuten en standaarddeviatie 17 minuten;
- verhouding spoedeisende/controle/normale patiënten.

Nulmeting

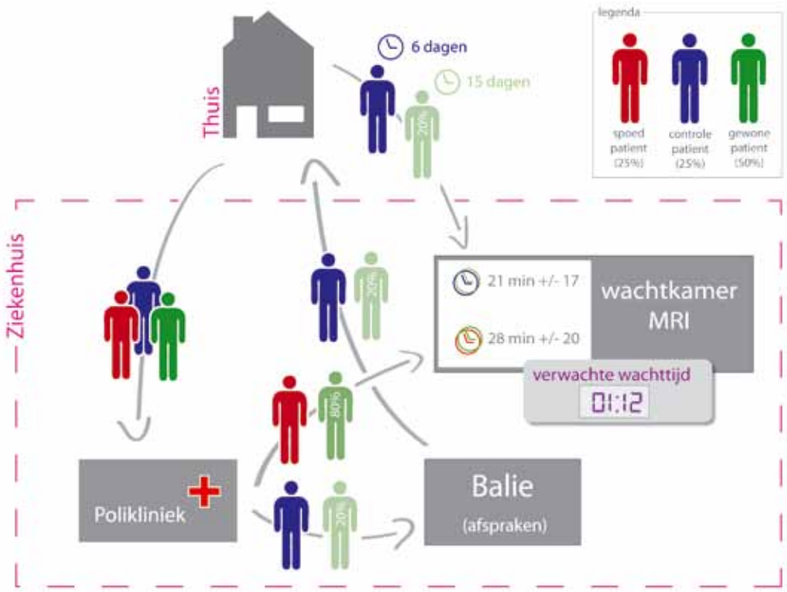
Figuur 1 geeft inzicht in de huidige situatie die gebruikt is om te komen tot een nulmeting. In het onderzoek is uitgegaan van een patiëntenpopulatie met 25% spoedeisende patiënten (deze patiënten moeten op dezelfde dag worden onderzocht) en 50% normale patiënten (zij moeten zo snel mogelijk worden onderzocht maar niet perse op dezelfde dag). 25% van de patiënten komt voor een controle afspraak die 3 à 6 maanden van te voren geboekt kan worden.

In de nulmeting maakt iedereen een afspraak voor een MRI-scan en krijgen de spoedeisende patiënten een afspraak op dezelfde dag. Voor deze patiënten worden altijd de laatste twee slots van de dag gereserveerd en wordt de rest behandeld in toevallig niet bezette slots of in overwerktijd. 7,5% van de patiënten komt niet opdagen voor hun afspraak waardoor de MRI leeg staat.

De prestaties van de nulmeting worden op vier dimensies gescoord. Gemiddeld moet een normale patiënt 20 dagen wachten op een MRI-scan, is de bezetting van de MRI-scan 84%, de gemiddelde overwerktijd 21 minuten en wordt er zo strak gepland dat er een gemiddelde wachttijd van 64 minuten in de wachtkamer is (zie figuur 1 voor de schematische weergave en figuur 3 voor de resultaten).



Figuur 1. Simulatie traditioneel afsprakensysteem MRI (gemiddelde doorlooptijd en wachttijd +/- standaarddeviatie)



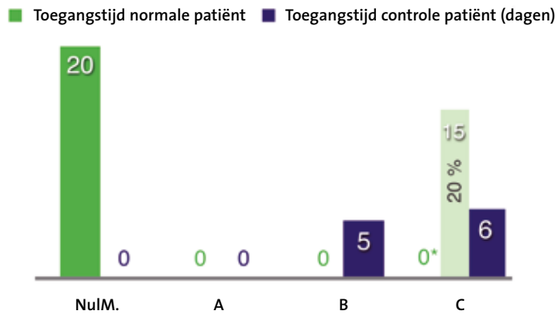
Figuur 2. Simulatie van scenario C met keuze voor inloop of afspraak (gemiddelde doorlooptijd en wachttijd +/- standaarddeviatie)

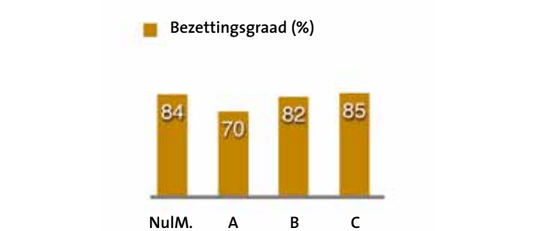
Scenario A: Volledige inloop
Ten opzichte van de nulmeting is gekeken wat er gebeurt wanneer volledige inloop wordt toegepast en iedereen dus op dezelfde dag geholpen wordt. Wat hier opvalt is het extreem hoge gemiddelde en standaarddeviatie van overwerktijd en wachttijd. Daarnaast is de bezetting van de MRI-scan vrij laag, wat verklaarbaar is omdat om 8.00 's ochtends nog vrijwel niemand binnen komt lopen en de MRI dan dus leeg staat (zie figuur 3).

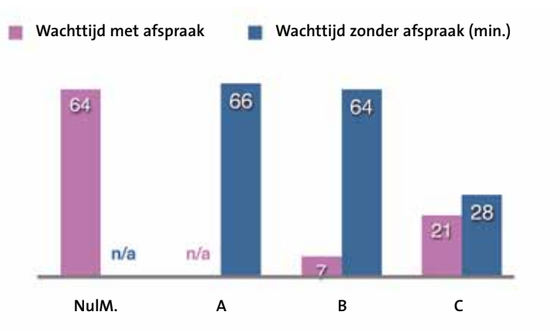
Scenario B: Inloop met afspraken voor controle patiënten
In scenario B maken controle patiënten wél een afspraak en mogen dat alleen doen voor zorgvuldig geselecteerde slots op momenten met lage inloop (dus bijvoorbeeld om 8 uur 's ochtends). Voor de andere patiënten blijft de situatie ongewijzigd. De bezettingsgraad is nu een stuk hoger dan in situatie A, namelijk 82% (zie figuur 3 voor de resultaten). De wachttijd in de wachtkamer en de overwerktijd kennen een onacceptabel hoge standaarddeviatie.

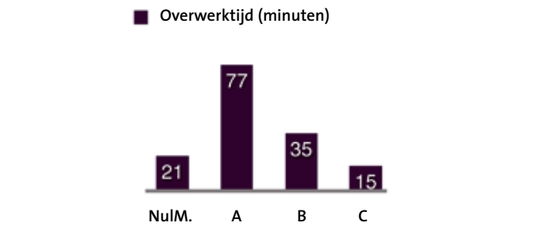
Scenario C: Keuze voor inloop of afspraak
In scenario B kregen controle patiënten een afspraak op momenten van lage inloop. In dit laatste scenario krijgen nu ook de normale patiënten de mogelijkheid om een afspraak te maken, opnieuw alleen op momenten van lage inloop. Zij krijgen bij binnenkomst op de MRI-afdeling de verwachte wachttijd te zien (deze is afhankelijk van het # patiënten in het systeem) en kunnen besluiten om in de wachtkamer plaats te nemen of om een afspraak te maken en later terug te komen. In figuur 2 is dit scenario schematisch weergegeven. Onderzoek in het Antoni van Leeuwenhoek Ziekenhuis leerde dat patiënten liever anderhalf uur wachten dan later terugkomen (Deetman, 2008).

De resultaten laten zien dat 20% van de normale patiënten bij aankomst in de wachtkamer

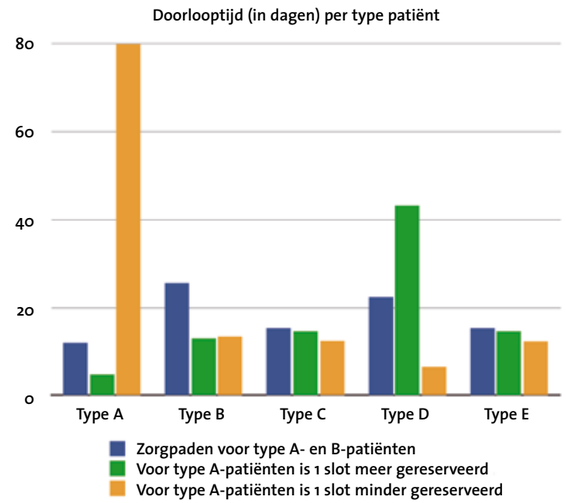








Figuur 3. Simulatie resultaten van verschillende scenario's van inloop



Figuur 4. Doorlooptijd gesimuleerde zorgpaden wanneer 1 slot meer of minder op de MRI wordt gereserveerd

een verwachte wachttijd van meer dan 1,5 uur ziet en daarom een normale afspraak plant. Hij moet hier 15 dagen op wachten. De overgebleven 80% van de patiënten wordt dezelfde dag onderzocht. Dit scenario presteert op alle fronten beter dan de nulmeting en heeft zelfs overal een lagere standaarddeviatie. Bovendien is deze methode relatief makkelijk te implementeren en door de keuzevrijheid erg patiëntvriendelijk. In figuur 3 zijn de resultaten van de nulmeting en de verschillende scenario's weergegeven.

Methode 2: Zorgpaden

Een tweede methode die op dit moment erg populair is, zijn de zogenaamde zorgpaden. Het bekendste voorbeeld van een zorgpad is waarschijnlijk de mammapoli waarbij een zorginstelling anticipeert op een x aantal patiënten met de verdenking van borstkanker per week. Van te voren wordt voor deze groep bijvoorbeeld op donderdagmiddag ruimte gereserveerd op alle noodzakelijke diagnostische faciliteiten. Zorgpaden leiden theoretisch tot efficiënter georganiseerde zorg en minder ziekenhuisbezoeken voor een homogene patiëntengroep. Maar als er voor zorgpaden slots gereserveerd worden, in hoeverre leidt dit dan tot een suboptimum? Wat zijn de gevolgen op de doorlooptijden voor de gehele patiëntenpopulatie?

Om hier een antwoord op te vinden werd met behulp van Matlab een zeer versimpelde versie van de werkelijkheid gesimuleerd. In deze simulatie wordt een fictieve patiëntenpopulatie met vijf typen patiënten verondersteld van gelijke grootte, die elk een beroep doen op een deel van de radiodiagnostische faciliteit van het ziekenhuis. Voor twee typen patiënten worden zorgpaden gedefinieerd en slots gereserveerd. Voor de andere drie groepen worden geen slots gereserveerd. Figuur 4 laat het grote gevaar van deze aanpak zien: wanneer 1 slot meer of minder wordt gereserveerd kan de toegangstijd voor een patiëntengroep die gebruik maakt van dezelfde diagnostische faciliteit al exploderen. Deze simulatie onderstreepte het wankele evenwicht en daarmee de noodzaak van constante monitoring en een iteratief feedback systeem wanneer gebruik wordt gemaakt van zorgpaden.

Methode 3: Offline plannen

Ziekenhuizen gebruiken vaak een online planningsmethode, wat betekent dat op het moment dat een aanvraag binnenkomt de afspraak meteen gepland wordt. Bij een offline-methode worden afspraken verzameld en pas later ingepland waarna de patiënt wordt teruggebeld (met als risico dat de geplande afspraak niet haalbaar is voor de patiënt). Theoretisch gezien presteren offline-methodes beter omdat zij bij het inplannen van afspraken gebruik maken van meer informatie. De vraag is echter hoeveel beter een offline-methode zou werken in het geval van het inplannen van afspraken voor de radiodiagnostische faciliteit.

Om hier een antwoord op te vinden wordt opnieuw uitgegaan van dezelfde fictieve patiëntenpopulatie met vijf groepen patiënten van gelijke grootte. Er worden twee strategieën onderzocht. In de eerste strategie maken patiënten afspraken voor de noodzakelijke onderzoeken. Patiënt A gaat dus eerst naar de MRI en dan naar de ECHO om een afspraak te maken. Stel dat er op een dag 10 MRI-aanvragen binnenkomen dan zullen in een offline-methode deze afspraken aan het eind van de dag in de eerste 10 vrije slots worden ingepland. In een online-methode zal dit ook gebeuren met als verschil dat, als er dezelfde dag nog een gaatje is, de patiënt dezelfde dag geholpen kan worden. In strategie 2 wordt de restrictie toegevoegd dat type A- en C-patiënten hun afspraken op hetzelfde dagdeel moeten maken.

Het model is geformuleerd als een integer lineair programming-probleem en doorgerekend met een ILP solver. Voor strategie 1 presteert de online-methode zoals verwacht iets beter dan de offline-methode. Voor strategie 2 geeft de offline-methode voor patiënten van type A en C een totale doorlooptijd die ongeveer 20% korter is. De vraag is echter of deze verbetering het verlies in flexibiliteit waard is.

Conclusie

Voor alle drie de methoden valt op dat kleine veranderingen grote invloed kunnen hebben op de prestaties van het planningssysteem. Het is een zoektocht naar balans die eigenlijk alleen gevonden lijkt te kunnen worden in flexibele systemen die kunnen omgaan met fluctuerende patiëntenstromen en scanduren, patiënten die niet komen opdagen en spoedpatiënten. Wanneer voor zorgpaden slots worden gereserveerd op een hoog bezette diagnostische faciliteit is constante monitoring en feedback noodzakelijk. Offline *scheduling* verliest in de praktijk een deel van zijn kracht doordat het niet gebruik kan maken van slots die op dezelfde dag vrijkomen en doordat het minder patiëntvriendelijk is. Volledige inloop werkt niet omdat het geen antwoord heeft op de piekmomenten. Pas wanneer inloop op piekmomenten wordt gecombineerd met zorgvuldig geselecteerde afspraakslots wordt het voldoende flexibel om de balans terug te vinden.

Dit artikel is gebaseerd op de master thesis 'Towards a patient friendlier radio-diagnostic track' van Karin de Booij (2010), Vrije Universiteit Amsterdam.

Literatuur

Deetman, J. (2008). *Reducing throughput time of the radiodiagnostic track*. Master Thesis, Universiteit van Twente / NKI – AvL.

Gilles, R. (2007). *Same day access: mission (im)possible?* Master Thesis, University of Twente / NKI- AvL.

Murray, M., Berwick, D. M. (2003). Advanced Access: Reducing Waiting and Delays in Primary Care. *The Journal of the American Medical Association* 289(8), 1035-1040.

Karin de Booij *is afgestudeerd in operations research aan de Vrije Universiteit Amsterdam. Sinds 2010 is ze consultant bij Casemix, een onderzoeks- en adviesbureau in de gezondheidszorg.*
E-mail: <Karin.de.booij@casemix.nl>.

SPOEDEISENDE HULP

Zijn 47 SEH's echt voldoende?

Er is een maatschappelijke discussie gaande over de wijze waarop spoedeisende hulp in Nederland is georganiseerd. De bereikbaarheid van en aanrijtijd tot een post voor spoedeisende hulp is daarin een terugkerend aandachtspunt. In reactie op enkele recente publicaties, waarin voornamelijk gekeken wordt naar de risicogevallen waarvoor de aanrijtijd te groot is, roepen wij op tot enige nuance in de discussie. Een herorganisatie van de spoedeisende hulp heeft immers invloed op vrijwel heel de bevolking, in plaats van op een zeker aantal grensgevallen.

Arnoud Kuiper & Bart Veltman

Aanleiding: kamervragen over onnodig beroep op spoedeisende hulp

In april van dit jaar bevestigde minister Schippers van Volksgezondheid, in reactie op Kamervragen van de VVD, dat ze geen bezwaar heeft tegen het terugbrengen van het aantal posten voor spoedeisende hulp, de SEH's (Dorresteijn, 2011). In het licht van de stijgende zorgkosten en de wens om de beperkt beschikbare middelen zo in te zetten dat zij zoveel mogelijk ten goede komen van de zorgbehoevende zelf, is dit een zinvolle afweging. Begrijpelijkerwijs roept het direct de vraag op naar het aantal SEH's waar onze maatschappij dan wel behoefte aan heeft.

De NOS heeft die vraag opgepakt en onderzoek laten doen naar het minimaal aantal benodigde SEH's. In dit onderzoek is rekening gehouden met de, door de minister verwoorde, conditie dat een ambulance binnen 45 minuten na melding met de patiënt op een SEH moet kunnen zijn. Het onderzoek geeft aan dat 47 SEH's dan volstaan. Dit is een sterke vermindering ten opzichte van de ruim 100 SEH's die ons land momenteel heeft. Zoals de NOS zelf schrijft op 24 april: 'Helft spoedeisende hulp kan dicht' (Brink & Parre, 2011).

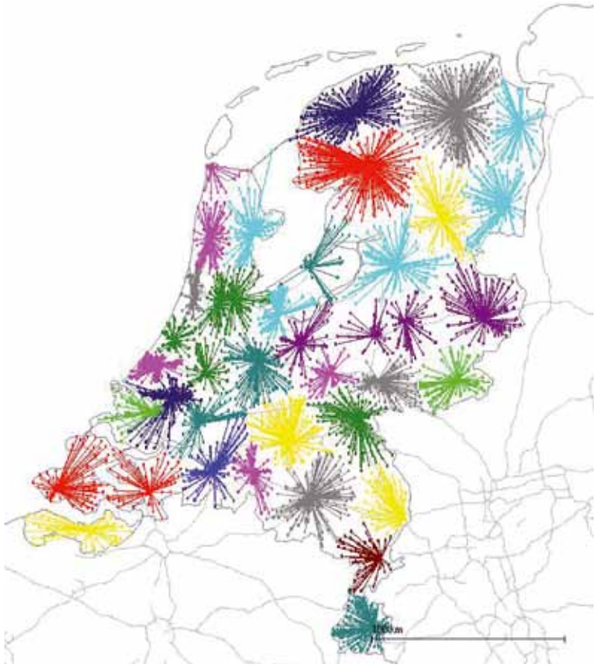
Nuance wenselijk

Nederland heeft genoeg aan 47 SEH's! Het is een heldere, eenvoudige uitspraak. Naast de kracht die ervan uitgaat, is deze eenvoud ook verleidelijk. Zij roept op om in actie te komen het aantal SEH's te verminderen. In de gezondheidszorg is daadkracht en actie zeker noodzakelijk, echter, hoe verleidelijk ook, iets meer nuance in de besluitvorming over een eventuele vermindering van het aantal SEH's zou geen overbodige luxe zijn.

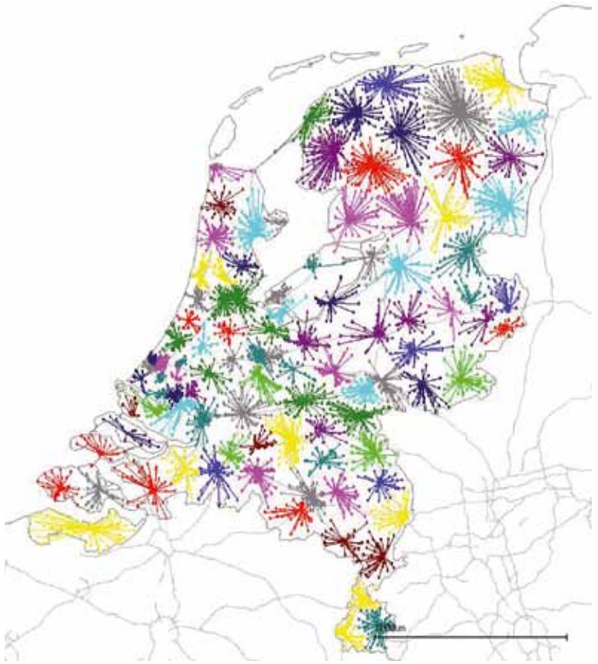
De eenvoud van de uitspraak versluiert namelijk informatie die voor de besluitvorming mogelijk relevant is. Het belangrijkste daarin is misschien wel dat het maar zeer de vraag is of één enkel criterium, de bereikbaarheid, maatgevend is voor het aantal SEH's dat onze maatschappij wenst. Maar zelfs als dat zo is, is op dat vlak van bereikbaarheid enige nuance toe te voegen.

Enerzijds suggereert de uitspraak dat 47 (zevenenveertig) SEH's volstaan, een precisie die slechts schijn is. Er is teveel onzekerheid in het spel om een dergelijk precieze uitspraak te rechtvaardigen, zoals:

- de verkeersdrukte en effecten die dat heeft op bereikbaarheid van de plaats van melding;



Figuur 1. Scenario met 40 SEH's



Figuur 2. Scenario met 100 SEH's

SEH's	0:10	0:12	0:14	0:16	0:18	0:20	0:22	0:24	0:26	0:28	0:30
40	51,0%	61,8%	72,1%	79,9%	86,1%	92,5%	95,2%	96,(%	98,3%	99,0%	99,3%
50	57,9%	69,1%	78,9%	85,2%	89,7%	94,7%	96,7%	97,6%	98,7%	99,1%	99,4%
60	63,5%	74,3%	83,8%	89,6%	92,8%	96,4%	97,7%	98,3%	99,1%	99,4%	99,6%
70	67,3%	77,7%	86,5%	92,4%	95,3%	97,9%	99,0%	99,5%	99,8%	99.9%	99.9%
80	71,2%	81,5%	89,0%	94,4%	96,8%	98,7%	99,4%	99,6%	99,8%	99.9%	99.9%
90	75,0%	84,4%	90,7%	95,5%	97,4%	98,8%	99,4%	99,6%	99,8%	99.9%	99.9%
100	77,5%	86,2%	91,9%	96,2%	98,0%	99,2%	99,6%	99,7%	99,8%	99.9%	99.9%

Tabel 1. Percentage van de bevolking dat bereikbaar is binnen een gegeven aantal minuten bij een gegeven aantal SEH's

- de benodigde tijd voor een ambulance om op de plaats van melding te komen, volgens ministeriele richtlijn maximaal 15 minuten;
- de benodigde tijd voor het ambulancepersoneel op de plaats van melding in het verlenen van de eerste, direct benodigde hulp;
- de in aanmerking komende locaties voor een SEH.

Los van het gegeven dat uit het eerdergenoemde onderzoek onduidelijk is hoe deze factoren in de berekeningen zijn meegenomen, is het aantal gewenste SEH's eerder een maatschappelijk te maken keuze dan een meetbaar feit.

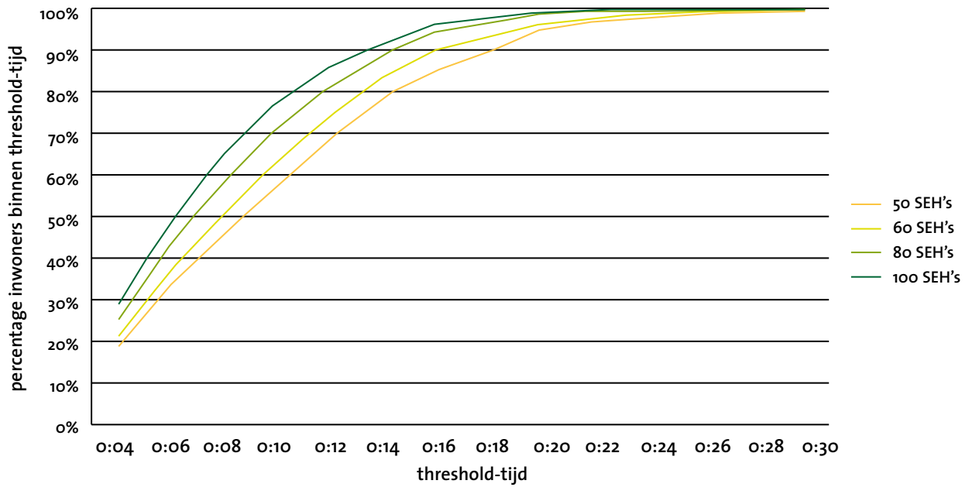
Anderzijds zijn er, naast de uiterste grens van 45 minuten, alleen al op het vlak van bereikbaarheid meer effecten te melden. Geïnspireerd door het signaal van de ministers, hebben we zelf onderzoek gedaan naar de effecten van een verandering van het aantal of de locaties van SEH's. De minister zelf roept, naast haar instemming met het terugbrengen van het aantal SEH's, op tot een goede landelijke spreiding en een doelmatige herinrichting. Hierbij wil de minister dat zowel situaties van overaanbod als zorgschaarste in beschouwing worden genomen zodat, in het bijzonder, de dunbevolkte gebieden minder kwetsbaar worden.

Dat doet zij niet voor niets. Simpel gezegd, zullen voor iedere SEH die gesloten wordt, spoedgevallen bestaan die meer tijd nodig gaan hebben om de dan dichtstbijzijnde SEH te bereiken. Daarmee geeft zij aan meerdere aspecten van belang te vinden bij de herinrichting van SEH's:

- het respecteren van de 45 minuten grens;
- het minder kwetsbaar maken van dunbevolkte gebieden door SEH-locaties beter te spreiden over het land.

Effecten voor risicogevallen en alle andere mede-burgers

In ons onderzoek zijn alle ziekenhuislocaties (in 2008 een kleine 200 in totaal, volgens gegevens van het RIVM) als potentiële SEH-locatie meegenomen. Vervolgens is voor verschillende mogelijk gewenste aantallen SEH's telkens berekend waar de SEH's moeten liggen om de bereikbaarheid zo goed mogelijk te houden. Wij hebben daarna de gewenste aantallen SEH's in stapgroottes van 10 laten oplopen, variërend van 40 tot 100; zie de resulterende locaties op de kaarten in figuur 1 en 2. Op die wijze kan voor verschillende tijdsintervallen berekend worden welk percentage van de bevolking binnen dat tijdsinterval na melding op een SEH kan zijn. Hierbij is rekening gehouden met de, eveneens door de minister opgelegde, randvoorwaarde dat een ambulance binnen 15 minuten na melding op de plaats van melding moet kunnen zijn. Effectief betekent dit dat er dan in het ongunstigste scenario nog maximaal 30 minuten over zijn om van de plaats van melding bij een SEH te komen.



Figuur 3. Percentage bevolking bereikt binnen threshold-tijd voor een gegeven aantal SEH's

De nuance die dit brengt, komt bijvoorbeeld mooi tot uiting in een vergelijking van het scenario met 100 SEH's (vergelijkbaar met de huidige situatie) en een scenario met 50 SEH's (vergelijkbaar met het eerder genoemde minimum aantal van 47 SEH's). Dan blijkt dat deze halvering in aantal SEH's betekent dat voor circa 20% van de bevolking de (dichtstbijzijnde) SEH ineens niet meer binnen 10 minuten bereikbaar is, waar die dat eerder wel was. Daarnaast groeit het aantal mensen dat niet binnen 30 minuten op een SEH kan zijn (vanaf plaats van melding) met ongeveer 0,5%, ofwel ruim 80.000 individuen. Zie ook figuur 3 en tabel 1 over *percentage bevolking bereikt*.

Conclusie

Of we een dergelijke impact, maatschappelijk gezien, betreuren of accepteren lijkt ons juist het onderwerp te mogen zijn van de discussie en de besluitvorming over het aantal en de locaties van de SEH's, naast mogelijk andere criteria. Ons pleidooi betreft de nuance te behouden in die discussie, ook al gaat dat ten kosten van de noodzakelijke daadkracht en actie.

Literatuur

Brink, R. van den & Parre, H. van der (24 april 2011). *Helft Spoedeisende hulp kan dicht*. NOS. <http://nos.nl/artikel/235398-helft-spoedeisende-hulp-kan-dicht.html>.

Dorresteijn, M. van (12 april 2011). *Schippers vindt concentratie SEH's geen bezwaar*. Zorgvisie. <http://www.zorgvisie.nl/Ondernemen/Schippers-vindt-concentratie-SEHs-geen-bezwaar.htm>.

Arnoud Kuiper *is afgestudeerd in de toegepaste wiskunde aan de Universiteit vanTwente en is senior consultant logistics bij de ORTEC Consulting Group.*
E-mail: <Arnoud.Kuiper@ortec.com>.

Bart Veltman *is partner en managing consultant bij de ORTEC Consulting Group.*
E-mail: <Bart.Veltman@ortec.com>.

Hoe veilig is het kwantificeren van VEILIGHEID?

West-Kruiskade, Oude Westen, Rotterdam

Tom de Leeuw

Veiligheid als maatschappelijk vraagstuk

Sinds begin jaren negentig van de vorige eeuw wordt er in Nederlands beleid gesproken over integrale veiligheid. De eerste aanzet daartoe was het begrip 'kleine criminaliteit' in het rapport *Samenleving en Criminaliteit* uit 1985. Deze term is in de loop van de jaren negentig hervertaald naar 'overlast', wat in het Angelsaksische discours *quality of life offenses* wordt genoemd. In het grotestedenbeleid vanaf 1994 werd vervolgens ook gesproken over leefbaarheid en veiligheid en zijn de eerste integrale veiligheidsmetingen ontwikkeld. De vraag is hier wat die *meetbare*

veiligheid ons sindsdien heeft opgeleverd. De Rotterdamse Veiligheidsindex is in 2002 voor het eerst verschenen om, onder het mom van *new public management*, beleid in meetbare doelstellingen uit te drukken en er verantwoordelijke ambtenaren en professionals op af te rekenen. Om korte metten te maken met de leefbaarheids- en veiligheidsproblemen had het toenmalige stadsbestuur behoefte aan een instrument dat de effecten van zijn veiligheidsbeleid kon gaan uitdrukken in 'harde feiten'. Na de verkiezingsoverwinning van Leefbaar Rotterdam in 2002 werd er een nieuw meetinstrument ontwikkeld dat sindsdien de veiligheidsontwikkeling van de stad in kaart brengt.

De Veiligheidsindex als methode

De scores in deze veiligheidsindex zijn opgebouwd uit drie verschillende soorten gegevens per veiligheidsthema, te weten: éénderde *'objectieve' gegevens* (meldingen en aangiften van criminaliteit en overlast uit registratiesystemen van ordediensten) en tweederde *gerapporteerde onveiligheidsgevoelens* uit een jaarlijkse steekproef van 13.000 geënquêteerde burgers. Om de objectieve en subjectieve gegevens onderling vergelijkbaar te maken, worden er z-scores berekend. Dit is een maat voor de gemiddelde afwijking van het rekenkundig gemiddelde. Vervolgens worden die z-scores gecorrigeerd voor de gestandaardiseerde ontwikkeling van het rekenkundig gemiddelde ten opzichte van het basisjaar 1999, zodat de z-scores tussen jaren vergeleken kunnen worden. Per veiligheidselement wordt het gemiddelde van deze genormaliseerde z-scores berekend voor de objectieve en de subjectieve gegevens en worden deze opgeteld om tot een indexscore per veiligheidselement te komen. Het veiligheidsindexcijfer van de wijk als geheel bestaat uit het optellen van de genormaliseerde rekenkundige gemiddeldes van die acht afzonderlijke veiligheidselementen aangevuld met vier *contextgegevens*, zoals de economische waarde van de woningen, de etnische samenstelling van de bevolking, het percentage uitkeringsgerechtigden en de tevredenheid over de buurt. Vervolgens worden deze gecorrigeerde rekenkundige gemiddeldes gecorrigeerd voor de invloed die ze volgens een correlatieanalyse zouden hebben op het veiligheidsgevoel. Deze totaalscore van een wijk wordt vervolgens omgezet in een indexscore op een schaal van 1 tot 10, waarbij onveilige wijken (onder de 3,9), probleemwijken (tussen de 3,9 en de 5), bedreigde wijken (tussen de 5 en de 6), aandachtswijken (tussen 6 en 7,1) en veilige wijken (boven de 7,1) worden onderscheiden (Gemeente Rotterdam, 2010a).

De Veiligheidsindex als politieke constructie

In plaats van een 'objectief' instrument is de veiligheidsindex een optelsom van beleidskeuzes. Gerapporteerde subjectieve onveiligheid drukt zwaarder dan de geregistreerde meldingen en aangiften op het veiligheidsindexcijfer, maar ook een sociaal-economische lagere waarde van de wijk tendeert dientengevolge naar een lager veiligheidsindexcijfer. Daarnaast bestaat veiligheid hier slechts uit acht veiligheidselementen (diefstal, drugsoverlast, geweld, inbraken, vandalisme, overlast, schoon en heel, verkeer) waarvoor registratiesystemen zijn geraadpleegd en waarover vragen zijn gesteld aan respondenten. De enquêtevragen vallen uiteen in de mate waarin die acht thema's als een buurtprobleem worden gewaardeerd en de mate waarin men daar slachtoffer van is geweest. Deze twee uiteenlopende vragen worden echter niet los van elkaar weergegeven in de uiteindelijke rapportage, net zoals ook de 'objectieve' en de subjectieve scores niet afzonderlijk worden besproken. De veiligheidsindex levert, door alle rekenkundige bewerkingen van de ruwe gegevens, één cijfer op dat de ontwikkeling van de wijkveiligheid weergeeft.

De vraag is wat we op basis van dit veiligheidsindexcijfer eigenlijk weten over veiligheid, gezien enkele methodologische kanttekeningen. Een van de gemiste kansen lijkt de onderlinge vergelijkbaarheid van de resultaten uit de verschillende jaren. Er wordt een min of meer vaste methode gehanteerd per jaar, maar de steekproef bestaat jaarlijks niet uit dezelfde respondenten. Hierdoor wordt het aantal variabelen dat invloed kan hebben op de beoordeling van veiligheid vergroot. Ook is het onduidelijk wat de kenmerken zijn van respondenten uit de steekproef en wat voor invloed deze hebben op hun gerapporteerde veiligheidsgevoel, terwijl de uitkomsten van de enquête daar statistisch goed voor te controleren zijn. Een kritiekpunt van respondenten is de suggestieve waarde van de enquêtevragen. Last hebben van hondenpoep en duiven worden gekoppeld aan veiligheid. Jongeren zien als buurtprobleem draagt bij aan een somberder beeld van veiligheid en slachtoffer zijn van drugsoverlast zou ook geïnterpreteerd kunnen worden als het enkel zien van drugsdealers en -gebruikers. Zonder afbreuk te willen doen aan de waarde van statistische bewerkingen om orde in de chaos van gegevens aan te brengen, roept deze reductie van complexiteit in de veiligheidsindex vragen op.

Etnografische inzichten in de complexiteit en contextualiteit van veiligheidsbeleving

Uit drie jaar etnografisch onderzoek in de wijk het Oude Westen blijkt dat statistische analyses moeilijk recht kunnen doen aan de complexiteit van dagelijkse realiteiten (De Leeuw & Van Swaaningen, 2011). Deze wijk had tussen 2001 en 2008 het laagste veiligheidsindexcijfer van de stad en scoort ook nu nog een onvoldoende (Gemeente Rotterdam, 2010b). Echter, al snel bleek uit mijn interviews en (participerende) observaties dat de veiligheidsbeleving zeer divers was en dat het grootste deel van de bewoners en professionals zich niet herkende in het veiligheidsindexcijfer. Na jaren van veiligheidinterventies zijn allerlei actoren het er over eens dat de wijk een sterke veiligheidsverbetering heeft doorgemaakt in de context van haar verleden. Er zijn bijna geen drugsgebruikers meer, het aantal dealers is sterk afgenomen en de grip op criminaliteits- en overlastrisico's is sterk toegenomen door nieuwe interventiemogelijkheden. Toch blijken respondenten in de steekproef nog steeds een laag veiligheidsgevoel te rapporteren, voornamelijk rond drugs- en jongerenoverlast. Uit interviews blijkt dat er op die thema's grote verschillen bestaan in perceptie die niet terugkomen in de veiligheidsindex, omdat daarin geen plaats is voor een *toelichting* op onveiligheidsgevoelens en de vragen uitsluitend gericht zijn op de *aanwezigheid* van dergelijke gevoelens. Op deze manier weten we nog steeds niet *waarom* een bewoner zich veilig of onveilig voelt en blijft het onduidelijk waar beleidsmatig op ingegrepen moet worden om deze gevoelens te verbeteren. Relatieve verbeteringen komen dus niet terug in de meting als elke dealer er nog één te veel is en zaken als slachtofferschap, onveiligheidsgevoel en irritatie in de bevolkingsenquête door elkaar gebruikt worden.

Methodologische verrijkingsmogelijkheden van onderzoek naar veiligheid

Statistisch onderzoek leent zich niet voor specifieke uitspraken over de diversiteit en complexiteit van betekenissen van veiligheid op een lager abstractieniveau zoals dat van individuen in een wijk als het Oude Westen. Dat is dan ook niet een verwijt dat ik hier wil maken aan de statistiek, want zij kan daarom juist goed gecombineerd worden met etnografisch onderzoek (Maltz, 1994).

Het punt is juist dat de veiligheidsindex in dat opzicht weinig van de potenties van de statistiek benut om specifiekere antwoorden te genereren op complexe vragen. De vraag is dus hoe we enerzijds meer gebruik kunnen maken van de mogelijkheden van de statistiek in het onderzoek naar veiligheid, en anderzijds hoe we in dat gebruik meer kunnen zoeken naar een symbiose met de etnografie. Statistische analyse zou meer een startpunt mogen zijn voor etnografisch veldwerk, zodat vanuit (afwijkingen op) algemene patronen meer ingezoomd kan worden op achterliggende betekenissen. Tegelijkertijd zouden etnografische resultaten ook meer statistisch gevalideerd kunnen worden door op basis van die bevindingen nieuwe variabelen te maken die het veiligheidsgevoel blijken te beïnvloeden. Een voorbeeld is het vragen naar de woonduur van respondenten en de mate waarin zij actief zijn in de wijk. Statistiek zou niet alleen beschrijvend maar ook meer verklarend ingezet kunnen worden door bijvoorbeeld regressieanalyses tussen uiteenlopende variabelen en veiligheidsgevoelens uit te voeren. Dit soort analyses wordt beleidsintern wel gedaan maar niet gecommuniceerd in de jaarlijkse rapportage van de veiligheidsindex. De indexcijfers per wijk en per veiligheidselement geven slechts verschillen aan van enkele tienden op een schaal van 1 tot 10, waardoor ze weinig houvast bieden om hier specifieke conclusies aan te verbinden. Berekeningen aan de hand van minder bewerkte ruwe data bieden daarentegen meer kans op minder afgevlakte resultaten.

Methodologisch tegenwicht aan de politisering van veiligheid

Kortom: de manier waarop statistisch materiaal nu gepresenteerd wordt in de veiligheidsindex geeft algemene lijnen en weinig inzichten in de intensiteit en mogelijke betekenissen van cijfermatige veranderingen. De praktische toepassing van het statistisch instrumentarium blijkt sterk aangepast aan de politieke belangen die het meetinstrument dient. Het is in de eerste plaats een beleidsinstrument in plaats van een wetenschappelijk meetinstrument dat alle mogelijkheden van de statistiek en etnografie heeft afgetast. Er was in 2002 behoefte aan een instrument dat op pragmatische wijze de beleidsinzet zichtbaar kon maken toen veiligheid hét politieke thema werd (Tops, 2007). Het risico dat wijken als het Oude Westen hierdoor vertekend uit naar voren kunnen komen en dat dit reële consequenties heeft, in de vorm van beleidsinterventies die worden ingezet als gevolg van het veiligheidsindexcijfer, wordt op de koop toegenomen. Het kwantificeren van veiligheid is dus op zichzelf geen onveilige exercitie, maar wel met de wetenschap dat cijfers ook maar relatief zijn en verder geduid dienen te worden in hun specifieke context, iets waartoe de etnografie in staat is.

Literatuur

Gemeente Rotterdam (2010a). *Methodologische verantwoording Veiligheidsindex 2010*, geraadpleegd op <www.rotterdam.nl/veilig>.

Gemeente Rotterdam (2010b). *Veiligheidsindex 2010*, geraadpleegd op <www.rotterdam.nl/veilig>.

Leeuw, T. de & Swaaningen, R. van (2011). Veiligheid in veelvoud: beeld, beleid en realiteit in Rotterdams Oude Westen. *Tijdschrift voor veiligheid* (10)1, 26–42.

Maltz, M. D. (1994). Deviating from the mean: the declining significance of significance. *Journal of research in crime and delinquency* (31)4, 434–463.

Tops, P. (2007). *Regimeverandering in Rotterdam: hoe een stadsbestuur zichzelf opnieuw uitvond.* Amsterdam: Uitgeverij Atlas.

Tom de Leeuw is als promovendus verbonden aan de sectie criminologie van de Erasmus Universiteit Rotterdam.
E-mail: <deleeuw@frg.eur.nl>

VOORWAARDELIJKE DINSDAGSKINDEREN

'There are three kinds of lies: lies, damned lies, and statistics.' Dit is een adagium dat aan allerlei beroemdheden wordt toegeschreven. Statistici/kansrekenaars hebben bij het 'publiek' nog altijd een slechte naam: ze geven verkeerde antwoorden, onbegrijpelijke antwoorden of helemaal geen antwoorden. Een bekend voorbeeld is het nog steeds niet uitgewoede drie-kastenprobleem, waarbij statistici voornamelijk elkaar voor de voeten lopen. Maar, het gaat vaak om onzinnige vragen en het betreft meestal voorwaardelijke kansen. Hieronder een treffend voorbeeld.

Het juninummer van *Pythagoras* – wiskundetijdschrift voor jongeren – opent met het artikel 'Kansrekening is een verraderlijk vak'. Vervolgens wordt een voorbeeld gegeven van deze verraderlijkheid. Het voorbeeld heet 'De kinderen van Femke' en behandelt de volgende vraag:

> 'Femke heeft twee kinderen, waarvan minstens één zoon die op dinsdag is geboren. Wat is de kans dat Femke twee zonen heeft?'.

Deze vraag werd ook gesteld tijdens een symposium ter ere van de beroemde puzzelaar Martin Gardner (1914 - 2010). Er kwam een fout antwoord. 'Vet kicken', natuurlijk. Maar ook hier geldt het spreekwoord dat één dwaas meer kan vragen dan honderd wijzen kunnen beantwoorden.

De eenvoudige vraag: 'Femke heeft twee kinderen, waarvan minstens één zoon. Wat is de kans dat Femke twee zonen heeft', is eenvoudig, zonder veel rekenwerk, te beantwoorden: 1/3. Het antwoord op de wat ingewikkelder vraag van hierboven kan ook eenvoudig worden gevonden, maar niet helemaal zonder rekenwerk. Je intuïtie laat je hierbij in de steek. Hoe zou iemand ook kunnen weten dat zijn buurvrouw met twee kinderen 'minstens één zoon heeft die op dinsdag is geboren'?

De gezochte kans zal moeten worden geïnterpreteerd als een voorwaardelijke kans. Het enige bruikbare hulpmiddel hierbij is de definitie: $P(A|B) = P(AB) / P(B)$.

We geven de gebeurtenissen: het eerste, resp. tweede kind is een zoon aan met $Z(1)$ en $Z(2)$, en de gebeurtenissen: het eerste resp. tweede kind wordt op een dinsdag geboren met $D(1)$ en $D(2)$. Deze vier gebeurtenissen zijn onderling onafhankelijk en voor alle Z's en D's geldt $P(Z) = 1/2$ en $P(D) = 1/7$. Gevraagd wordt:

$$P[Z(1)\cap Z(2) \mid \{Z(1)\cap D(1)\} \cup \{Z(2)\cap D(2)\}].$$

Toepassing van de gegeven definitie van voorwaardelijke kans levert het in *Pythagoras* vermelde antwoord: 13/27.

Het intuïtief verrassende is dat de toevoeging 'op dinsdag' verschil maakt. Als je de toevoeging 'op dinsdag' vervangt door 'in januari', dan wordt de kans 23/47. Specificeren we dat tot 'op 1 januari', wordt de kans bijna 1/2, namelijk 729 / 1459.

Dit roept een vraag op. Het is bekend dat $P(A|B)=P(A)$ als A en B onafhankelijk zijn. Het is ook bekend dat A en B onafhankelijk zijn, als $P(B)=0$ is. Mijn vraag luidt: geldt ook dat $\lim P[A|B(n)]=P[A]$, als $\lim P(B(n))=0$ voor n naar oneindig? Antwoord: nee.

Zo kan een onzinnige vraag toch nog een echte vraag oproepen. Toch zou *Pythagoras* er goed aan doen om zijn 'jongeren' op het nut en het vermaak van de kansrekening/statistiek te wijzen en niet via onzinnige voorbeelden op de 'verraderlijkheid' ervan.

Fred Steutel is emeritus hoogleraar kansrekening aan de TU Eindhoven.
E-mail: <f.w.steutel@tue.nl>

Omgaan met onzekerheden in het waterveiligheidsbeleid

Robin Nicolai, Ton Vrouwenvelder, Karolina Wojciechowska & Henri Steenbergen

Nederland is wereldwijd vermaard om haar expertise op het gebied van waterbouw. De strijd tegen het water wordt al gevoerd sinds het begin van de jaartelling. Zeeuwen waren de eersten die hun vruchtbare land tegen stormvloeden beschermden door ringdijken te bouwen om polders (Rooijendijk, 2009). De dijken bleken lang niet altijd bestand tegen stormvloeden. Diepe geulen doorsneden de dijken, maar de kennis ontbrak om die te dichten. Alleen de cisterciënzer monniken konden dat. Maar na het herstellen van de dijken waren er nieuwe grotere stormvloeden, die nieuwe betere maatregelen vereisten. Langzamerhand werd er steeds meer kennis verworven over het bouwen van waterkeringen. Het sluitstuk tot nu toe vormen de Deltawerken, die in 1986 werden voltooid met de oplevering van de Oosterscheldekering.

De huidige waterveiligheidsfilosofie is mede te danken aan de ideeën van wiskundige David van Dantzig. De operationele research (OR) technieken in zijn artikel *Economic decision problems for flood prevention* vormden de wetenschappelijke onderbouwing voor de veiligheidsnormen van de Deltacommissie uit 1960. De Deltacommissie was direct na de watersnoodramp van 1953 ingesteld om te onderzoeken hoe de getroffen gebieden voortaan beter te beschermen. De destijds opgestelde veiligheidsnormen vormen nu nog de basis voor de wettelijke veiligheidstoetsing van de waterkeringen. Dit artikel laat aan de hand van een vorig jaar uitgevoerde studie zien hoe OR en statistiek binnen de wettelijke toetsing worden toegepast.

Waterveiligheid in Nederland

Een groot deel van Nederland wordt bedreigd door overstromingen vanuit de Rijn en de Maas, vanuit het IJsselmeer of vanuit de Noordzee. Om Nederland te beschermen tegen zulke overstromingen is het opgedeeld in dijkringgebieden: gebieden omringd door een reeks van aaneengesloten primaire waterkeringen (dijken, duinen of kunstwerken) en hoge gronden. Iedere dijkring heeft een wettelijk vastgesteld veiligheidsniveau. Deze fungeert als norm en is uitgedrukt in een overschrijdingsfrequentie (zie figuur 1). De totale lengte van de primaire waterkeringen is ongeveer 3.600 kilometer.

De Waterwet vereist dat de waterkeringen bestand zijn tegen omstandigheden, i.e. waterstanden en golven, waarvan de overschrijdingsfrequentie groter is dan de norm. De hoogte van de norm hangt af van de (economische) waarde van het gebied en de herkomst van de overstroming (kust, meer of rivier). Voor de kustregio's is het veiligheidsniveau meestal gelijk aan 1/4000 en 1/10000 per jaar gekozen. Voor de dijkringen

Figuur 1. Veiligheidsnormen van dijkringgebieden in Nederland

langs de rivieren Rijn en Maas is de norm 1/1250 of 1/2000 per jaar. De norm van enkele kleine dijkringen langs de Maas in Limburg – de Maaskaden – is gelijk aan 1/250 per jaar. Voor iedere dijkring worden waterstanden en golven bepaald met een overschrijdingsfrequentie die overeenkomt met de norm. Deze *maatgevende* waterstanden en golven worden Hydraulische Randvoorwaarden (HR) genoemd.

De HR spelen een belangrijke rol in de wettelijke toetsing van de primaire waterkeringen. Bij de toetsing wordt voor verschillende faalmechanismen nagegaan of de waterkering bestand is tegen de *maatgevende omstandigheden*. De huidige veiligheidsfilosofie voorziet *nog niet* in een volledig probabilistische betrouwbaarheidsanalyse van de waterkering, die als uitkomst een overstromingskans kent. De belasting op de waterkering is gemodelleerd als de verwachtingswaarde van een stochast. De sterkte van de waterkering is een constante. Als compensatie voor onzekerheden zijn aan de sterktekant wel veiligheidsmarges ingebouwd.

Traditioneel wordt er bij de berekening van de HR alleen rekening gehouden met onzekerheden als gevolg van natuurlijke variabiliteit zoals zeewaterstand, meerpeil, afvoer en wind. Onzekerheden in andere invoerparameters, statistische gegevens en fysische modellen – samen kennisonzekerheden – worden verdisconteerd in een verwachtingswaarde of aanname. Ze maken (nog) geen deel uit van het modelinstrumentarium voor het afleiden van de HR. *Uitsluitend,* maar niet alle, natuurlijke onzekerheidsbronnen worden volledig probabilistisch meegenomen ('uitgeïntegreerd') om tot de verwachtingswaarde (lees HR) te komen.

De belangrijkste argumenten om uit te gaan van verwachtingswaarden in plaats van het volledig meenemen ('uitintegreren') van alle (kennis) onzekerheden, zijn:

- Meenemen van kennisonzekerheden in de afleiding van de HR zou een aanpassing betekenen van de manier waarop waterkeringen worden getoetst. Mogelijk kan dit worden opgevangen door aanwezig conservatisme aan de sterktekant, maar dit is niet zeker.
- De complexiteit van de modellen neemt toe naarmate meer onzekerheden worden meegenomen.
- De benodigde hoeveelheid invoergegevens en modelleerwerk neemt enorm toe. Een belangrijke vraag hierbij is of er genoeg informatie is om deze onzekerheden 'voldoende betrouwbaar' te kunnen modelleren.

Bij beleidsmakers bestaat behoefte aan inzicht in het effect van kennisonzekerheden op de HR. Enerzijds wil men weten wat een volledige probabilistische faalkansanalyse oplevert. Anderzijds is het voor het prioriteren van onderzoek naar kennishiaten belangrijk om te weten welke kennisonzekerheden de HR het meest beïnvloeden. In 2010 is door de auteurs een onderzoek uitgevoerd naar het effect van kennisonzekerheden op de HR. Het doel was om de grootte van het effect te bepalen en aan te geven welke onzekerheidsbronnen de grootste invloed hebben op de HR. Dit artikel is een beknopte samenvatting van deze studie.

Modellering hydraulische belastingen

Dreigend hoogwater en extreme golfcondities worden veroorzaakt door hoge afvoeren, hoge windsnelheden, hoge zeewaterstanden of hoge meerpeilen of combinaties hiervan. Het modelinstrumentarium heeft als invoer een aantal natuurlijke kansvariabelen: wind, zeewaterstand, meerpeil en afvoer. Deze basisstochasten hebben een bepaalde kansverdeling, die meestal op een beperkte set meetgegevens (bijvoorbeeld jaarmaxima van rivierafvoer of zeewaterstand) is gebaseerd. De onzekerheid in de parameters van de kansverdeling kan door korte meetreeksen best groot zijn. Na ongeveer 120 jaar meten is de hoogst waargenomen afvoer op de Rijn bij Lobith bijvoorbeeld 12.600 m^3/s. Om de dijken langs de Rijn te toetsen is echter de afvoer met een overschrijdingsfrequentie van 1/1250 per jaar nodig. Uit een statistische analyse van extreme Rijnafvoeren volgt uiteindelijk een *maatgevende* Rijnafvoer van 16.000 m^3/s. Het bijbehorende 95%-betrouwbaarheidsinterval loopt naar schatting van 13.500 tot 18.500 m^3/s. Het modelinstrumentarium houdt, conform het uitgangspunt van de huidige aanpak voor de HR, geen rekening met deze statistische onzekerheid.

Fysische modelberekeningen leiden voor verschillende combinaties van de basisstochasten tot de lokale waterstand en de golfcondities op de waterkering. Deze berekeningen worden apart gemaakt en de uitkomsten worden opgeslagen in een database. De fysische modellen voor de waterbeweging en golfgroei zijn net als ieder model een benadering van de werkelijkheid. De uitkomsten bevatten een bepaalde onzekerheid. Hierbij komt nog dat de invoerparameters van de fysische modellen, zoals bijvoorbeeld de bodemligging, als deterministische parameters worden verondersteld. In werkelijkheid vertoont de bodemligging een natuurlijke variatie in ruimte en in tijd, die binnen het huidige beleid niet volledig is meegenomen.

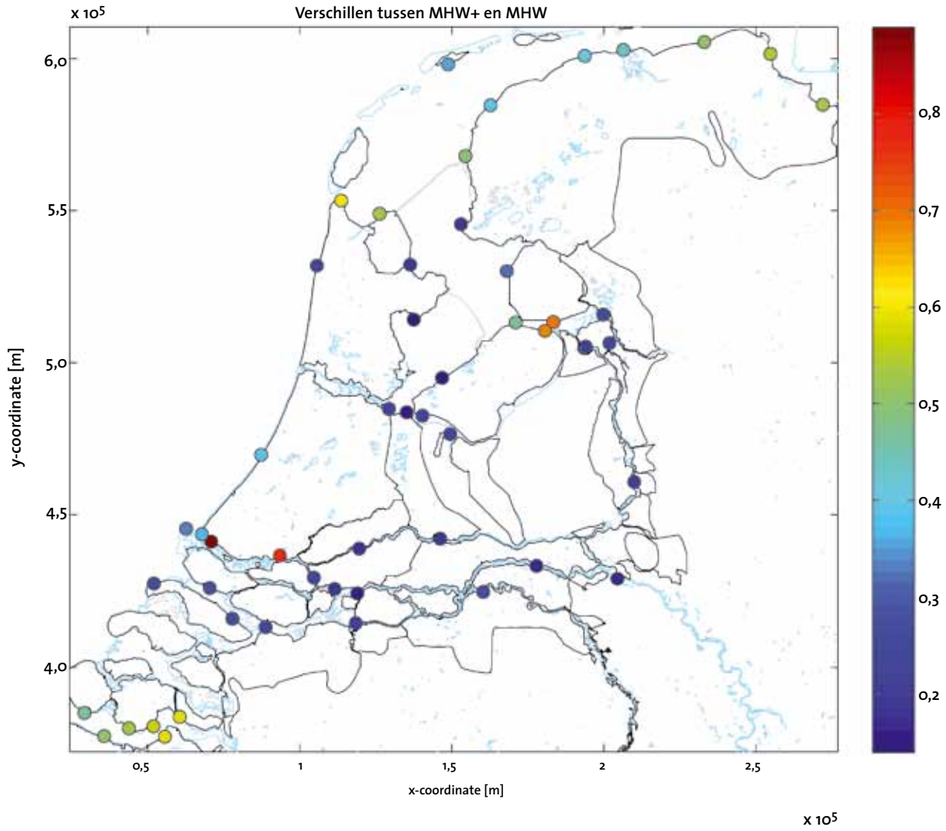
De probabilistische berekening van de HR combineert de vooraf berekende fysische grootheden met de kansverdelingen van de basisstochasten. De HR volgen uit de *meest waarschijnlijke* waarden of verdelingen van de basisstochasten met een gezamenlijke overschrijdingsfrequentie gelijk aan de norm. In de vooraf opgestelde database worden de lokale waterstand en golfcondities opgezocht die optreden bij deze waarden van de basisstochasten. De waarden van de basisstochasten en de HR vormen samen het *ontwerppunt* in deze betrouwbaarheidsanalyse.

Modellering kennisonzekerheden

In de in 2010 uitgevoerde studie is eerst een lijst opgesteld met onzekerheden, die niet in het

	WATERSYSTEEM				
ONZEKERHEIDSBRON	BOVENRIVIEREN	BENEDENRIVIEREN	KUST	MEREN	VECHT- EN IJSSELDELTA
Modelonzekerheid lokale waterstand	X	X		X	X
Statistische onzekerheid in rivierafvoer	X	X			X
Statistische onzekerheid in windsnelheid	X	X	X	X	X
Statistische onzekerheid in IJsselmeerpeil				X	X
Statistische onzekerheid in zeewaterstand		X	X		
Bodemligging	X	X	X	X	X
Strijklengte	X	X			X
Fout in golfparameters	X	X	X	X	X

Tabel 1. Kennisonzekerheden voor verschillende watersystemen



Figuur 2. Effect van additionele onzekerheden op maatgevende waterstanden

modelinstrumentarium zijn opgenomen. Hierbij is onderscheid gemaakt tussen onzekerheden voortkomend uit de fysische modellering, statistische onzekerheden (vanwege beperkte hoeveelheid gegevens) en overige invoerparameters (zoals bijvoorbeeld de bodemligging). Voor iedere onzekerheidsbron is op basis van bestaande literatuur of *expert judgement* een kansverdeling geschat. In de meeste gevallen zijn gemiddelde en standaarddeviatie ingeschat en is een normale kansverdeling verondersteld. De statistische onzekerheid in de *maatgevende* Rijnafvoer is bijvoorbeeld als een normale kansvariabele met gemiddelde 16.000 m^3/s en standaarddeviatie 1.276 m^3/s (anders gezegd: een variatiecoëfficiënt van 8%) gemodelleerd. Niet alleen de onzekerheid rondom de maatgevende Rijnafvoer, maar de onzekerheid rondom alle mogelijke afvoeren moet worden meegenomen. Standaard gaat het model uit van een gegeven relatie tussen de Rijnafvoer en de overschrijdingsfrequentie van bepaalde afvoerwaarden: de 'werklijn'. De statistische onzekerheid in de Rijnafvoer is daarom nu gemodelleerd door een kansvariabele met gemiddelde 1 en variatiecoëfficiënt 8% te vermenigvuldigen met de afvoer op deze werklijn.

Tabel 1 toont de kennisonzekerheden die aan het model zijn toegevoegd voor de berekening van hoogwaterstanden en golfcondities. Deze onzekerheidsbronnen zijn volledig probabilistisch verwerkt in de berekening van de waterstanden en golfcondities.

Resultaten en conclusies

De uitkomsten van de berekeningen met de additionele stochasten voor ongeveer 50 waterkeringen in Nederland laten zien dat de waterstanden en golfcondities hoger/zwaarder uitpakken dan het referentieniveau (huidige methode). Dit is niet verrassend. Toevoegen van extra onzekere factoren leidt bijna altijd tot hogere verwachtingswaarden, omdat de hogere stochastwaarden zwaarder doorwegen in het eindresultaat. Figuur 2 toont het effect op de waterstand. Het effect van kennisonzekerheden op de waterstand is over het algemeen beperkt tot enkele decimeters, maar dit is wel significant.

Opvallend is dat het toevoegen van kennisonzekerheden resulteert in een 0,8 m hogere waterstand bij Rotterdam. Hier is nog nader onderzoek gewenst, maar het lijkt erop dat een eenvoudige wijziging van het sluitregime van de Maeslantkering dit effect al kan beperken.

Met name de zeewaterstand- en afvoerstatistiek en de onzekerheid in de lokale waterstand blijken zwaar door te wegen in de berekende waterstanden. De onzekerheid rondom de genoemde factoren is erg groot en dit werkt vrijwel rechtstreeks door in de maatgevende waterstand. Golfcondities worden veelal beïnvloed door de grote statistische onzekerheid in de windsnelheid.

De studie die is uitgevoerd is niet uitputtend of volledig. Niet alle kennisonzekerheden zijn toegevoegd. De resultaten geven echter wel inzicht in de belangrijkste kennisonzekerheden voor de HR. Statistische onzekerheidsbronnen hebben het grootste effect op de HR. Het is wel zo dat het bepalen van statistische onzekerheid niet eenduidig gebeurt. Het verdient aanbeveling om hier objectieve methoden voor te ontwikkelen. Indien wordt gekozen om overstromingskansen en risico's volledig probabilistisch te bepalen, moeten kennisonzekerheden worden meegenomen. Een tweede aanbeveling is het uitvoeren van een modelanalyse waarbij alle onzekerheden en verborgen veiligheid, ook aan de sterktekant, integraal worden meegenomen. Het volledige model moet de basis vormen voor het uitvoeren van de economische kosten-baten optimalisatie, zoals in 1956 door Van Dantzig uitgevoerd.

Literatuur

Dantzig, D. van (1956). Economic decision problems for flood prevention. *Econometrica* 24, 276-287.

Rooijendijk, C. (2009). *Waterwolven. Een geschiedenis van stormvloeden, dijkenbouwers en droogmakers.* Amsterdam/Antwerpen: Uitgeverij Atlas.

Robin Nicolai is werkzaam als adviseur bij HKV lijn in water. Hij is in 2008 gepromoveerd op een onderzoek naar onderhoudsoptimalisatie van complexe systemen. Momenteel richt hij zijn aandacht op onzekerheden in waterveiligheid.
E-mail: <r.p.nicolai@hkv.nl>.

Karolina Wojciechowska werkt als adviseur / onderzoeker bij HKV lijn in water. Zij schrijft een proefschrift over de toepassing van beslismodellen in operationeel overstromingsmanagement.
E-mail: <wojciechowska@hkv.nl>.

Ton Vrouwenvelder en Henri Steenbergen zijn beiden werkzaam als senior onderzoeker constructieve veiligheid bij TNO. Zij houden zich onder meer bezig met het beoordelen van zowel nieuwe als bestaande bouwconstructies alsmede het ontwikkelen en toepassen van geavanceerde rekentechnieken voor het ontwerpen en toetsen van waterkeringen.
E-mailadressen: <ton.vrouwenvelder@tno.nl> & <henri.steenbergen@tno.nl>.

OPROEP

OM KANDIDATEN TE NOMINEREN VOOR DE VvS+OR THESIS AWARD 2011

Ter bekroning van een uitzonderlijke afstudeerprestatie aan een Nederlandse instelling voor wetenschappelijk onderwijs of het hoger beroepsonderwijs looft de VvS+OR al sinds lange tijd een scriptieprijs uit: de VvS+OR Thesis Award. Ook dit jaar roept de VvS+OR op voor nominaties voor deze prijs. De prijs bestaat uit een oorkonde en een geldbedrag van 1000 euro.
Genomineerd kunnen worden studenten die tussen september 2009 en September 2011 zijn afgestudeerd en die nog niet eerder zijn genomineerd. Vanaf dit jaar wordt geen onderscheid meer gemaakt tussen een bachelor- of een master-thesis.

Hierbij worden supervisors opgeroepen om een uitmuntende afstudeerthesis te nomineren voor de VVS+OR Thesis Award 2011. Reglementen en het nominatieformulier zijn te downloaden op de website van de VvS+OR (<www.vvs-or.nl>). Het nominatieformulier dient tezamen met de afstudeerthesis in pdf-formaat te worden opgestuurd naar de VVS+OR op thesisaward@vvs-or.nl. Daarnaast dient de uitgeprinte versie van het nominatieformulier en de thesis opgestuurd te worden naar:

dr. Mark van der Loo
Secretaris VVS+OR Thesis Award 2011
Room A3069
Dept. of Methodology and Quality Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague, The Netherlands

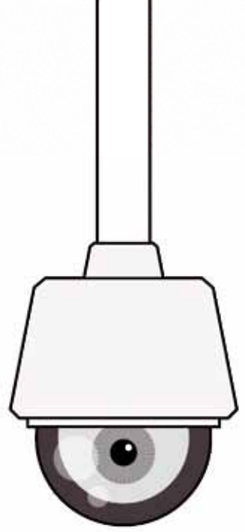
De nominatie dient per e-mail en post binnen te zijn voor 1 december 2011.

De indiening van een nominatie dient vergezeld te gaan van een aanbevelingsbrief van de supervisor van de genomineerde. In deze brief dienen in ieder geval aan te orde te komen:

- De beschrijving van de master thesis als een originele bijdrage aan een onderwerp uit de statistiek of operations research, of als een inventieve toepassing van theoretische concepten uit de statistiek en/of operations research, alsmede
- de overige kwaliteiten van de genomineerde.

Namens de VvS+OR,

Prof. dr. Jacqueline Meulman, President
Dr. Mark van der Loo, Secretaris Jury Thesis Award 2011



SLIMME BEWAKINGSCAMERA'S

Op straten, in gebouwen, stadions, supermarkten, aankomst- en vertrekhallen van vliegvelden, op stations en in de metro, overal hangen camera's waarmee bewakingsdiensten mensen observeren om de veiligheid in deze openbare ruimten te kunnen garanderen. Het analyseren van de videobeelden door mensen is een kostbare aangelegenheid en op grote schaal nauwelijks uitvoerbaar. Het ontwikkelen van slimme camera's waarmee videobeelden automatisch en real time geanalyseerd kunnen worden is op dit moment een hot topic in de onderzoekswereld.

Léon Rothkrantz

In de meeste supermarkten zijn op dit moment camerasystemen geïnstalleerd. Het voornaamste doel is ongewenste bezoekers, personen met ongewenst gedrag of ongewenste gebeurtenissen zo snel mogelijk te detecteren. Het gaat hierbij onder andere om proletarisch winkelen, zakkenrollers en agressie ten opzichte van het personeel. Maar de camera's kunnen ook gebruikt worden als digitale winkelassistent. Om kosten te besparen wordt er bezuinigd op het aantal winkelassistenten. Vooral in de stille uren of in de piektijden kan het gebeuren dat mensen vergeefs wachten op assistentie of dat er onbemenste kassa's zijn. Het camerasysteem kan dit observeren en een waarschuwing genereren (figuur 1).

Bij grote supermarkten worden de beelden van de camera's in de controlekamer bekeken door speciaal opgeleid bewakingspersoneel. De beelden worden ook tijdelijk opgeslagen zodat ze indien nodig achteraf bekeken kunnen worden. Het dagelijks bekijken van de beelden is niet alleen een geestdodende activiteit maar ook een kostbare aangelegenheid. Bij veel onderzoeksinstituten wordt geprobeerd om 'slimme' camera's te ontwerpen waarmee de beelden automatisch

Figuur 1. Klant vraagt via de camera om assistentie van de filiaalchef

geanalyseerd kunnen worden en ongewenst gedrag of gebeurtenissen gedetecteerd kunnen worden. Het achteraf analyseren van de beelden is zinvol om een reconstructie van een diefstal te maken maar is minder effectief om daders op heterdaad te kunnen betrappen. Het *real time* analyseren van beelden is het ultieme doel (Zajdel 2007; Datcu 2007).

In veel gevallen gaat het niet om geïsoleerde camera's, maar om een netwerk van camera's. Het is de bedoeling dat de camera's onderling gaan communiceren over geobserveerde personen. Bij de ingang van een supermarkt kunnen personen geïdentificeerd worden die op de lijst van veelplegers staan en speciaal gevolgd worden door de winkel. Op het moment dat er een vergrijp plaatsvindt, is het wenselijk de dader te volgen op weg naar de uitgang en voordat hij het pand verlaat te kunnen aanhouden. Het volgen van personen in drukke winkels waarbij personen gemakkelijk kunnen opgaan in de massa is geen eenvoudige aangelegenheid (figuur 2). In de meeste gevallen proberen daders het pand zo snel mogelijk maar wel zo onopvallend mogelijk te verlaten. In het geval van meerdere daders die gestolen spullen aan elkaar doorgeven wordt het automatisch volgen nog complexer.

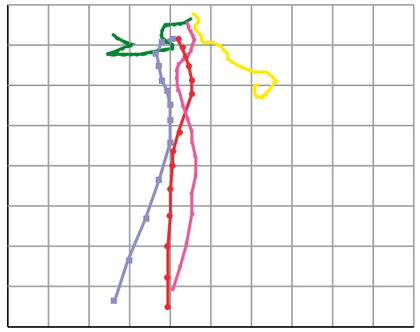
Onderzoeksdoel

In dit artikel concentreren we ons op het ongewenste gedrag in een supermarkt (Popa, 2011). Van het winkelend publiek wordt verondersteld dat ze zich op een bepaalde manier gedragen. Bij binnenkomst is het de bedoeling een boodschappenwagentje of -mandje te nemen. Gewenste producten worden daarin gedeponeerd en niet in de eigen tas. Het is ook niet gewenst om door de winkel te rennen en tegen mensen en producten aan te botsen. Het automatisch detecteren en semantisch interpreteren van gedrag is een complexe aangelegenheid. Vooral omdat er vele gedragingen en gedragsvarianten mogelijk zijn.

Iedere gedragsanalyse start met het automatisch detecteren en volgen van objecten. Het detecteren van personen kan plaatsvinden met gezichtsherkenning (Popa, 2010). Bij de meeste fotocamera's is software geïnstalleerd waarmee (lachende) gezichten herkend kunnen worden door het plaatsen van een rechthoekje rond de gezichten. Het herkennen van gezichten is nog verre van perfect. Doordat de gezichten te ver weg van de camera zijn, gedeeltelijk bedekt worden, mensen niet in de camera kijken of door slechte belichtingsomstandigheden, is het niet altijd mogelijk de gezichten en daarmee personen te detecteren.

Bewegende objecten worden eerst gelokaliseerd en vervolgens getraceerd door het vergelijken van opeenvolgende frames van de video-opnames. Het getraceerde pad kan vervolgens geanalyseerd worden: is het pad een rechte lijn of een slingerende beweging en zitten er veel tempoversnellingen in het pad? Op basis van deze parameters kan een semantische interpretatie worden gegeven (zie figuur 3), zoals: is er sprake van een besluiteloze shopper, een funshopper, een gehaaste shopper of een verdachte shopper.

Figuur 2. Het volgen van winkelend publiek



Figuur 3. Gevolgde paden van winkelbezoekers

Helaas werken object-detecterings-trackings algoritmes nog verre van perfect. Personen kunnen alleen met een zekere kans *p* herkend worden. Uiteraard is deze kans afhankelijk van de plaats waar de camera bevestigd is, het blikveld van de camera en het tijdstip waarop de waarneming plaatsvindt, en natuurlijk het gedrag van de te detecteren persoon. Om deze waarnemingskansen te bepalen is het noodzakelijk video-opnames te maken van het winkelend publiek, deze data te annoteren en vervolgens te analyseren.

In supermarkten is er sprake van een netwerk van camera's waarmee het winkelend publiek langs verschillende routes geobserveerd kan worden. Een cruciale vraag hierbij is, hoe groot is de kans dat personen langs een bepaalde route niet waargenomen worden. Uiteraard is dat afhankelijk van de waarnemingskans van de verschillende camera's langs de route. Maar verder moeten we rekening houden met feit dat het langs bepaalde routes veel drukker kan zijn dan langs andere routes. Het gaat er dus niet alleen om de kans te bepalen dat een persoon niet gedetecteerd wordt die eenmalig een bepaalde route loopt. Maar als we alle waarnemingen van personen langs alle mogelijke routes op een dag verzamelen, hoe groot is dan de kans dat personen langs een bepaalde route niet gedetecteerd worden. In tabel 1 zien we een voorbeeld van observaties langs twee routes a, b van punt A naar punt B waarvan de helft gedetecteerd wordt en de andere helft niet. De detectiekansen van de camera's zijn langs beide routes gelijk, maar het is duidelijk dat over het cohort van alle opnames, de kans op non-detectie langs pad b groter is.

	GEDETECTEERD	NIET-GEDETECTEERD
Aantal shoppers langs route a	1	1
Aantal shoppers langs route b	49	49

Tabel 1. Aantallen waarnemingen langs twee verschillende routes met detecties

Samenvattend formuleren we de volgende onderzoeksvragen:

1. Hoe groot is de kans op detectie van personen door een camera op een bepaalde locatie en op een bepaald tijdstip?
2. Hoe groot is de kans dat een persoon dat zich langs een bepaalde route van ingang naar uitgang beweegt volledig gemist wordt door het surveillance systeem van meerdere camera's?
3. Wat zijn de meest kwetsbare routes?

Vraag 1 kunnen we beantwoorden aan de hand van voldoende video-opnames door visueel na te gaan hoeveel personen al dan niet worden waargenomen. Om de vragen 2 en 3 te kunnen beantwoorden moeten we de kansen van tijdreeksen gaan berekenen, bestaande uit observaties langs een bepaalde route. Beschouwen we de routesegmenten als toestanden dan beweegt een persoon zich van de ene toestand naar de andere met een zekere waarschijnlijkheid.

In iedere toestand vinden er een of meerdere observaties plaats van al dan niet gedetecteerde personen, eveneens met een zekere waarschijnlijkheid. De waarschijnlijkheid van een rijtje opeenvolgende observaties kan berekend worden met behulp van (hidden) Markov modellen.

Probabilistisch redeneren

In verband met de privacy van geobserveerde personen en locatie geven hebben we opnames en analyses uitgevoerd in een gesimuleerde winkel (zie figuur 4). De winkel is opgedeeld in sectoren (*region of interests*) die geobserveerd worden door verschillende camerasystemen. De kans dat een winkelbezoeker in sector S_i wordt gedetecteerd noteren we met $p(S_i)$ en met $1\text{-}p(S_i)$ de kans dat hij niet wordt gedetecteerd. We noteren met a_{ij} de kans dat een bezoeker van sector S_i naar sector S_j loopt .

De kans op non-detectie van een bezoeker die een pad van sector S_i naar S_j volgt kunnen we als volgt berekenen:

$$P_{nd}(O_i,O_j) = (1\text{-}p(S_i)) \times a_{ij} \times (1\text{-}p(S_j))$$

We herkennen hierin een Markov model, waarbij zoals bekend de kans van een reeks observaties berekend wordt als product van de kans van een waarneming in een bepaalde toestand vermenigvuldigd met de transitiekansen tussen opeenvolgende toestanden. Op deze wijze kan de waarschijnlijkheid van ieder rijtje observaties worden berekend.

Figuur 4. Mock-up van een winkel met bijbehorende plattegrond

Resultaten

Aan de hand van de opnames in de gesimuleerde winkel waren we in staat de onderzoeksvragen 1, 2 en 3 te beantwoorden. Op basis van de resultaten met betrekking tot vraag 1 hebben we de positie van de camera bijgesteld en andere parameterinstellingen genomen. De dode hoeken waarbij het winkelend publiek uit het beeld verdwijnt, zijn zo veel mogelijk verwijderd. De resultaten van de vragen 2 en 3 hebben er toe geleid dat er langs de drukke routes een camera van hogere kwaliteit is geïnstalleerd dan langs de andere routes. Vervolgens is het *trackings* algoritme geoptimaliseerd aan de hand van de opnames. Ten slotte hebben de analyses van trajecten geleid tot een aantal voor winkeliers interessante grafieken over consument-en-productinteractie. Inmiddels zijn we gestart met de opnames en analyses van *real life*-opnames waarbij de beschreven modellen eveneens goed bleken te werken.

Conclusies

Dit artikel geeft een beschrijving van een systeem van slimme bewakingscamera's geïnstalleerd in een supermarkt. Probabilistische modellen zijn gebruikt om de gevolgde paden en gedrag van winkelende bezoekers te analyseren. De resultaten hebben geleid tot een aanpassing van de camerasystemen en optimalisering van de gebruikte software. De analyse van de paden van de winkelbezoekers hebben geleid tot voor winkeliers interessante resultaten. Zo bleek uit de analyse van trajecten dat sommige delen van de winkels en producten nauwelijks aandacht kregen van de bezoekers wat geleid heeft tot een andere opstelling van het assortiment. Bezoekers die assistentie nodig hadden en getraceerd werden omdat ze veel heen en weer liepen konden geholpen worden. Verder is er een lijst opgesteld van verdachte en ongewenste gedragingen waar we in verband met de vertrouwelijkheid van het onderzoek niet nader op kunnen ingaan.

Het artikel is gebaseerd op onderzoeksresultaten van de PhD studenten Mirela Popa en Iulia Lefter binnen een onderzoeksproject met de partners Philips, TNO, TU Delft en de NLDA. We danken Philips voor het beschikbaar stellen van hun labfaciliteiten.
Meer achtergrond informatie is te vinden op de website <http://mmi.tudelft.nl/~leon>.

Literatuur

Collins, R., Lipton, A. & Kanade, T, (2000). Introduction to the special section on video surveillance. *IEEE Transactions pattern analysis and machine intelligence* 22(8), 745–746.

Datcu, D., Yang, Z. & Rothkrantz, L. J. M. (2007), Multimodal workbench for automatic surveillance applications. *Multimodal surveillance: Sensors, algorithms, and systems*. Boston: Artech House Publishers, pp. 311–338.

Haritaoglu, L., Harwood, D., Davis, L. S., (2000). Real-time surveillance of people and their activities. *IEEE Transactions pattern analysis and machine intelligence*, 22(8), 809–830.

Popa, M. C., Rothkrantz, L. J. M., Datcu, D., Wiggers, P., Braspenning, R. & Shan, C., 2010. A comparative study of HMMs and DBNs applied to facial action units recognition. *Neural network world* 6, 737–760.

Popa, M. C., Rothkrantz, L. J. M., Yang, C., Wiggers, P & Shan, C. Analysis of Shopping Behavior based on Surveillance System. In: Dimirovski G, editor. IEEE International conference on Systems and Man-Machine Interaction and Cybernetics (SMC 2010). Instanbul: Kudret Press, 2010, pp. 2512–9.

Zajdel, W., Krijnders, J. D., Andringa, T. C. & Gavrila, D. M., (2007). Cassandra: audio-video sensor fusion for aggression detection. *Proceedings IEEE Conference on advanced video and signal based surveillance AVSS*, pp. 200–205.

Léon Rothkrantz heeft wis-en natuurkunde gestudeerd aan de Universiteit van Utrecht en Amsterdam en psychologie aan de Universiteit van Leiden. Sedert 1980 werkt hij bij de TU Delft en vanaf 2008 als deeltijdhoogleraar aan de Nederlandse Defensie Academie.
E-mail: <LJM.Rothkrantz@NLDA.NL>.

ZO MOET HET GEGAAN ZIJN

De noodzaak van alternatieve verklaringen en bekritiseerbare analyses bij zoekzaken binnen de opsporing

Recente justitiële dwalingen in onder andere de zaak Ina Post, de Schiedammer parkmoord, de Puttense moordzaak en de zaak Lucia de Berk laten zien dat het werk van een opsporingsteam vatbaar is voor het fenomeen tunnelvisie: het gevaar dat te veel wordt vastgehouden aan één scenario/hypothese. Een statistische benadering tijdens het opsporingsproces, gericht op gelijktijdige beoordeling van meerdere mogelijke hypothesen, biedt hierbij hulp. TNO heeft hiervoor het Hypothesis Management Framework (HMF) ontwikkeld. HMF is een nieuwe methode waarmee de waarschijnlijkheid van verschillende hypothesen op basis van de beschikbare bewijzen simultaan en kwantitatief wordt ingeschat. Hiermee kan het risico op tunnelvisie worden verkleind en worden conclusies herleidbaar en bekritiseerbaar.

Bram Wisse, Sicco Pier van Gosliga & Gerard Bijsterbosch

Bij zoekzaken binnen de opsporing buigt de politie zich over zaken als de vondst van een levenloos lichaam, de opsporing van een gezochte verdachte of de vermissing van een persoon. Bij de analyse van de beschikbare informatie in zoekzaken spelen onzekerheid en onvolledigheid een prominente rol. Opsporingsteams moeten zich ondanks deze onzekerheid en onvolledigheid een onderbouwd beeld vormen van de zaak en dit beeld kunnen overbrengen op betrokkenen zoals de teamleider en de leider van het opsporingsonderzoek (officier van justitie). Bij het vormen van dit beeld maakt de analist (impliciet) allerlei aannames en inschattingen over de onzekerheid en interpretatie van de beschikbare informatie, en loopt hij/zij het risico op valkuilen als tunnelvisie: het gevaar dat teveel wordt vastgehouden aan één scenario/hypothese en men niet voldoende in staat is om open te staan voor alternatieve verklaringen. De rapportage van de commissie Posthumus, ingesteld naar aanleiding van de justitiele dwaling in de Schiedammer parkmoordzaak, heeft tot speciale aandacht voor het fenomeen tunnelvisie geleid.

Analyse van concurrerende hypothesen

De behoefte om rationeel en methodologisch verschillende hypothesen te beoordelen bestaat al lang. In de jaren 70 ontwikkelde voormalig CIA-medewerker Richards Heuer hiervoor de methode Analysis of Competing Hypotheses (ACH) (Heuer 1999; 2005). Deze methode bestaat uit een achtstappenplan waarbij de analist expliciet alle redelijkerwijs mogelijke verklaringen dient te identificeren en met elkaar te laten ‘concurreren’. Dit in plaats van de plausibiliteit van mogelijke verklaringen één voor één onafhankelijk te beoordelen. De kern van de methode wordt gevormd door een matrix waarin de diagnostische waarde van ieder bewijs beoordeeld wordt voor *alle* hypothesen. De methode legt de nadruk op het ontkrachten van de mogelijke verklaringen. Dit gebeurt door één voor één hypothesen te verwerpen waarvoor voldoende conflicterend bewijs beschikbaar is.

De ACH-methode van Heuer wordt onderwezen als onderdeel van de leergang Recherchekunde, afstudeerrichting criminaliteitsanalyse aan de School voor Recherche van de Politieacademie. Binnen de door de Politieacademie gehanteerde methode worden hypothesen slechts verworpen als hiervoor een beargumenteerde onderbouwing gegeven kan worden; bijvoorbeeld doorslaggevend DNA-materiaal of een sluitend alibi. Wat niet expliciet en beargumenteerd kan worden uitgesloten, blijft als optie open.

ACH heeft als kwalitatieve methode twee belangrijke beperkingen. In de eerste plaats worden impliciet alle hypothesen bij aanvang van het onderzoek even waarschijnlijk geacht. Hoewel dit ‘eerlijk’ klinkt hoeft dit niet altijd wenselijk te zijn. Stel dat een analist werkt aan een vermissingszaak van een jonge vrouw. Zonder over verdere informatie te beschikken, zal een analist een loverboy-scenario minder plausibel achten dan een wegloop-scenario. Het laatstgenoemde scenario komt immers veel vaker voor dan het eerstgenoemde. Bij ACH worden beide scenario’s bij aanvang even waarschijnlijk geacht, ook wan-

neer een scenario zeer zeldzaam is. Ten tweede kan bij een onevenwichtige verzameling bewijs het beschikbare bewijs een grotere invloed hebben dan wenselijk.

Er zijn verscheidene pogingen ondernomen om deze beperkingen tegen te gaan en de ACH-methode te kwantificeren door de waarschijnlijkheid van de hypothesen op basis van het beschikbare bewijs uit te drukken in een kans. Omdat deze pogingen te eenvoudig of praktisch niet toepasbaar worden geacht heeft TNO de HMF-methode ontwikkeld (Gosliga en Van de Voorde 2008; Wisse en Gosliga 2010). De HMF-methode is oorspronkelijk voor inlichtingenanalisten van Defensie opgesteld, maar in dit artikel wordt de methode gepresenteerd aan de hand van een casus uit het (politie) opsporingsdomein.

Hypothesis Management Framework

De HMF-methode ondersteunt de gelijktijdige evaluatie van mogelijke hypothesen, op basis van subjectieve statistiek. HMF kan worden gezien als een probabilistische ACH-benadering. De methode bestaat uit een stappenplan. Hierin wordt eerst een (kans)model opgesteld voor mogelijke hypothesen en (mogelijk) beschikbaar bewijs. Vervolgens wordt het model kritisch geëvalueerd en de uiteindelijke bevindingen gerapporteerd. Een HMF-model kent drie typen variabelen: hypothesen, indicatoren en statements. *Hypothesen* zijn te evalueren mogelijke verklaringen. Bij de vermissing van een jonge vrouw zouden dit bijvoorbeeld 'weggelopen', 'loverboy', 'zelfmoord' of 'omgebracht' kunnen zijn. Een *indicator* is een (in principe) observeerbaar fenomeen dat diagnostische waarde heeft voor tenminste één van de hypothesen. Een *statement* tenslotte is concrete informatie (met bronvermelding) waarmee het optreden van een of meerdere indicatoren wordt bevestigd of ontkracht. Een voorbeeld van een indicator in een vermissingscasus is 'spullen meegenomen voor een overnachting', waarbij een bijbehorend statement kan zijn: 'bij huisbezoek door rechercheur Jansen op 8 augustus bleken toiletartikelen en kleding niet te ontbreken'. Er wordt dus zowel met de aanwezigheid als met de afwezigheid van bewijs gerekend.

Een HMF-model wordt gekwantificeerd door drie typen schattingen. In de eerste plaats wordt de waarschijnlijkheid van de verschillende hypothesen uitgedrukt in een kans, nog voordat beschikbare informatie in beschouwing wordt genomen. Hiervoor kunnen eventueel beschikbare statistieken over vergelijkbare zaken worden gebruikt.

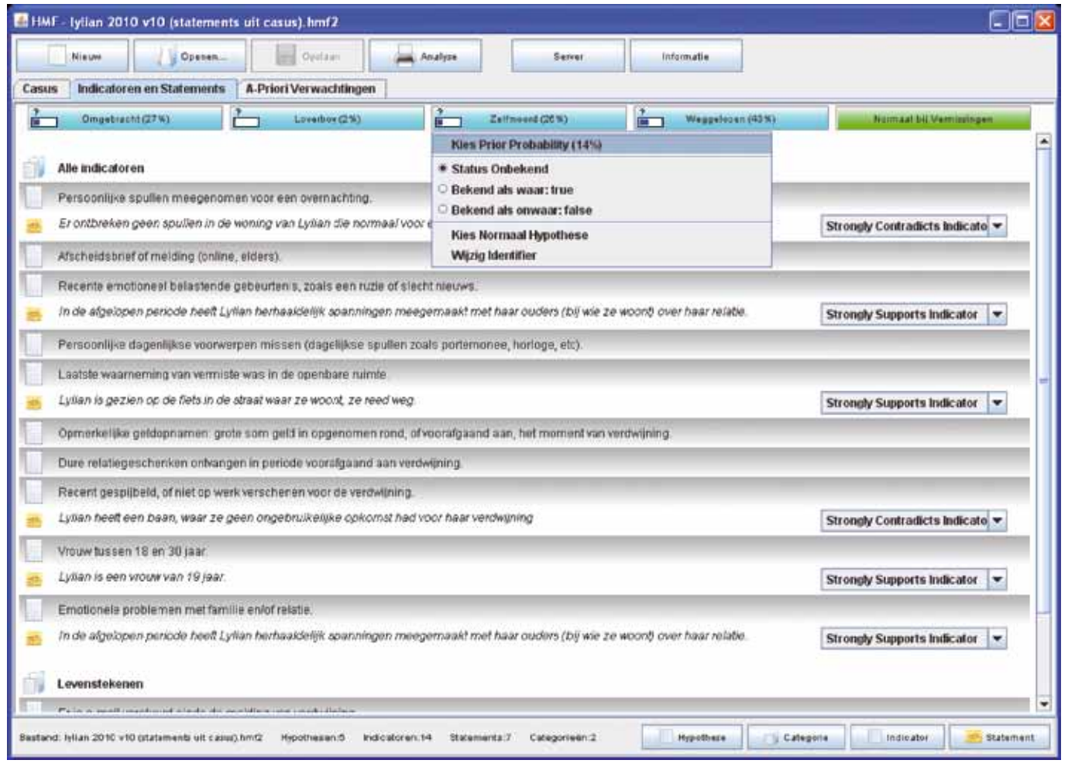
De volgende schatting heeft betrekking op de indicatoren. Hiervoor dient de likelihood geschat te worden dat de indicator optreedt, gegeven het al dan niet optreden van elk van de hypothesen. Voorbeeld: de likelihood van het optreden van de indicator 'spullen meegenomen voor overnachting' gegeven dat de hypothese 'Weggelopen' waar is zal groot zijn, bijvoorbeeld 80%. Het is echter ook mogelijk dat deze indicator optreedt terwijl de vermiste *niet* is weggelopen. Deze zgn. 'false positive' kans wordt ook meegenomen in de beoordeling. Als laatste moet voor elk van de statements de likelihood geschat worden voor het statement, gegeven het al dan niet optreden van de indicator. Voorbeeld: de likelihood van het statement 'bij huisbezoek door rechercheur Jansen op 8 augustus bleken toiletartikelen en kleding niet te ontbreken' gegeven dat de indicator 'spullen meegenomen voor overnachting' waar is, zal laag zijn, bijvoorbeeld 10%.

Op basis van deze drie typen schattingen wordt de waarschijnlijkheid van elk van de hypothesen iedere keer geüpdatet nadat nieuw bewijs in de vorm van een statement wordt toegevoegd aan het model. De volgorde waarin de statements aan het model zijn toegevoegd maakt uiteraard niet uit voor de gevonden waarschijnlijkheid van de hypothesen. Bij het updaten van de waarschijnlijkheden van de hypothesen wordt gebruik gemaakt van een onderliggend Bayesian Belief Netwerk (BBN). Een HMF-model is in feite een BBN waarbij specifieke eisen worden gesteld aan de modelstructuur van het BBN. Hierdoor kan het model met beperkte inspanning aangepast en uitgebreid worden, waardoor relatief eenvoudig hypothesen, indicatoren en statements toegevoegd en verwijderd kunnen worden.

Door gebruik te maken van indicatoren én statements kan onderscheid gemaakt worden tussen concreet bewijs, de statements, en interpretatie/gebruikswijze van bewijs, de indicatoren. De HMF-methode is geïmplementeerd in een prototype softwareapplicatie die de onderliggende BBN-modelstructuur automatisch genereert en waarmee de berekeningen geautomatiseerd uitgevoerd kunnen worden.

Vermissingscasus

Om de toepasbaarheid van HMF te onderzoeken is een casus uitgewerkt die is gebaseerd op een geanonimiseerde vermissingscasus van een 19-jarige vrouw. In tabel 1 staan samenvattende statistieken van zowel een HMF-analyse als een ACH-analyse van de casus. De rij 'HMF zonder bewijs' betreft de a-priorikansen (verwachtin-



Schermafbeelding van de HMF-software

gen vooraf zonder meenemen van de bewijslast) op de hypothesen in het HMF-model. Deze zijn gebaseerd op een studie door Foy (2004) naar de oorzaak van vermissingen in Australië. HMF-gebruikers dienen uiteraard wel altijd in te schatten of en in welke mate gebruikte statistieken representatief zijn voor de huidige casus.

Na de invoer van het bewijs uit de casus zien we in de rij 'HMF met bewijs' dat de hypothese 'Weggelopen' nu minder waarschijnlijk wordt geacht door de analisten (wijzelf in dit geval), terwijl de waarschijnlijkheid van de minder wenselijke hypothesen 'Zelfmoord' en 'Omgebracht' is toegenomen. Dit zien we ook terug in de rij *ACH inconsistency score*: er is veel meer inconsistent bewijs voor 'Weggelopen', uitgedrukt in een *inconsistency score* van 3, dan voor 'Zelfmoord', score 1, en 'Omgebracht' waarvoor geen inconsistent bewijs is in de casus. Echter, waar bij ACH de hypothese 'Weggelopen' het vaakst inconsistent is met het beschikbare bewijs, is 'Weggelopen' bij de HMF-analyse nog wel steeds de meest waarschijnlijke hypothese, waarschijnlijker dan de andere drie hypothesen bij elkaar. HMF biedt dus een veel duidelijker inzicht in de absolute waarschijnlijkheid van de hypothesen.

Door deze casus slechts aan de hand van samenvattende statistieken te bespreken doen we beide methoden wel te kort. In de praktijk bieden de methoden vooral een raamwerk dat de kritische evaluatie van bewijsmateriaal ondersteunt. Daarnaast zal men bij ACH vooral ook naar de diagnostische waarde van bewijs kijken, en niet slechts naar de hoeveelheid inconsistent bewijs.

Conclusie

Een HMF-model stelt de analist in staat op elk moment gedurende het onderzoek de waarschijnlijkheid van de hypothesen kwantitatief in te schatten op basis van het dan beschikbare bewijs. Op ieder moment in het onderzoek kan een dan zeer onwaarschijnlijk geachte hypothese dus waarschijnlijk worden als nieuw bewijs daarvoor aanleiding geeft. Doordat onwaarschijnlijke hypothesen actief in beschouwing blijven, wordt het risico op tunnelvisie verminderd. Daarnaast bestaat de mogelijkheid indicatoren in het model op te nemen waarvoor nog geen informatie beschikbaar is. Door een gevoeligheidsanalyse op deze en reeds beschikbare indicatoren uit te voeren, kan de analist bepalen van welk type informatie hij/zij op dat moment het meest verwacht te leren; voor welke indicatoren zou het zinvol kunnen zijn om extra informatie te verzamelen. Zo kan de HMF-methode helpen bij de prioritering van het verzamelen van aanvullende informatie.

Door de toepassing van HMF worden aannames en overtuigingen expliciet gemaakt en conclusies herleidbaar en bekritiseerbaar. Hoewel het gebruik van HMF meer tijd en inspanning vergt dan het gebruik van ACH, biedt de methode duidelijk meerwaarde. Voor vaker voorkomende zaken kunnen modellen van hypothesen en indicatoren als templates worden ontwikkeld, waarmee naast tijdswinst ook kennis kan worden vastgelegd.

ANALYSEMETHODE	HYPOTHESE			
	WEGGELOPEN	LOVERBOY	ZELFMOORD	OMGEBRACHT
HMF zonder bewijs	63%	1%	14%	13%
HMF met bewijs	47%	1%	17%	20%
ACH-inconsistency score	3	2	1	0

Tabel 1. Samenvattende statistieken vermissingscasus van HMF- en ACH-analyse

Literatuur

Foy, S. (2004). *Profile of Missing Persons in New South Wales*. Thesis. Charles Sturt University, Faculty of Arts, School of Policing Studies.

Gosliga, S. P. & I. van de Voorde (2008). Hypothesis Management Framework: a Flexible Design Pattern for Belief Networks in Decision Support Systems. In *Uncertainty in Artificial Intelligence: Proceedings of the Twenty-Fourth Conference*. Helsinki: AUAI Press.

Heuer, R. J. J. (1999). *Psychology of Intelligence Analysis.* Washington, D.C.: Center for the Study of Intelligence, CIA.

Heuer, R.J.J. (2005). *How Does Analysis of Competing Hypotheses (ACH) Improve Intelligence Analysis?*, available from: http://www.pherson.org/Library/H15.pdf, Pherson Associates, LCC.

Wisse, B. W. & S. P. van Gosliga (2010). *Kwantitatieve Hypothesevorming: Hypothesis Management Framework*. TNO-DV 2010 A101. Den Haag: TNO.

Bram Wisse studeerde bedrijfswiskunde & informatica aan de VU in Amsterdam. Sinds zijn afstuderen doet hij onderzoek bij TNO op het gebied van (militaire) operations research en promoveert hij in de subjectieve statistiek aan de University of Strathclyde Business School in Glasgow.
E-mail: <bram.wisse@tno.nl>.

Sicco Pier van Gosliga behaalde zijn MSc in Kunstmatige Intelligentie aan de University of Edinburgh. Thans is hij werkzaam bij TNO en promoveert hij aan de UvA in Amsterdam op gedistribueerde Bayesiaanse netwerken.
E-mail: <sicco_pier.vangosliga@tno.nl>.

Gerard Bijsterbosch behaalde zijn MCI Rechercheknde aan de Politieacademie te Apeldoorn. Hij is werkzaam als docent-onderzoeker aan de School voor Recherche van de Politieacademie.
E-mail: <gerard.bijsterbosch@politieacademie.nl>.

JAARLIJKSE LUNTEREN BIJEENKOMST

van het LNMB en het NGB

17-19 januari 2012

Congrescentrum De Werelt in Lunteren

De traditionele Lunteren Bijeenkomsten georganiseerd door het LNMB (Landelijk Netwer Mathematische Besliskunde) en het NGB (Nederlands Genootschap Besliskunde) hebben als belangrijkste doel het bevorderen van het contact tussen beginnende en gevorderde onderzoekers. Dit jaar gebeurt dat onder andere rond voordrachten door **Kurt Anstreicher** (The University of Iowa, USA), **Nikhil Bansal** (IBM Research, New York, USA), **Benny Moldovanu** (University of Bonn, Germany) en **Assaf Zeevi** (Columbia University, New York, USA).

De laatste dag is gevuld met een seminar rond de thema's **Supply Chain Regie, Pricing en Optimalisatie in Humanitaire hulp.** Tijdens dat seminar wordt ook de **Ortec Excellence in Advanced Planning Award** toegekend.

Uitgebreide informatie is te vinden op de website <www.lnmb.nl/conferences/2012/>; u kunt zich hier ook aanmelden. Voor meer informatie over de locatie zie <www.congrescentrum.com>.

Geurproef niet meer in gebruik bij strafzaken

Geurt Jongbloed & Frank van der Meulen

In *de Volkskrant* van vrijdag 22 april 2011 was op de voorpagina te lezen:

> De omstreden 'geuridentificatieproef', waarin speurhonden van de politie verdachten van een misdrijf aanwijzen, is van tafel. Hondengeleiders die met de proef sjoemelden om verdachten veroordeeld te krijgen, worden echter niet vervolgd.

Het Volkskrantartikel stelt dat 'uit onderzoek van statistici van de TU Delft blijkt dat hondengeleiders de hand moeten hebben gelicht met de regels'. In deze bijdrage willen we de context en inhoud van ons onderzoek wat breder uiteenzetten.

In wat volgt leggen we eerst uit hoe de geurproef in zijn werk gaat. Vervolgens schetsen we in het kort welke punten van kritiek er in de voorbije jaren zoal naar voren zijn gebracht als het gaat om de bewijskracht van de uitkomst van een geurproef. Deze kritiek, met name geuit door prof. J.E.R. Frijters, deed het College van Procureurs Generaal (PG) tot een onafhankelijk onderzoek naar de geurproef besluiten. Het deelonderzoek dat bij de TU Delft werd uitgezet gaat over de vraag of bepaalde rangschikkingen wel zijn bepaald zoals in het protocol is voorgeschreven. Daarom zullen we de procedure die volgens het protocol gevolgd moet worden nader uitleggen en tot slot onze aanpak en conclusies van de analyse beschrijven.

Procedure van de geurproef

Het uitgangspunt is dat op de plaats delict een voorwerp is gevonden en dat er een verdachte is. Met behulp van een daartoe getrainde hond wil men vaststellen of de verdachte een 'geurovereenkomst' met het voorwerp heeft. Zonder details volledig te willen weergeven, komt de procedure van de geurproef op het volgende neer. Voor de proef leveren zeven personen een geurdrager aan. Dit zijn de verdachte, X, en zes figuranten, A, B, t/m F. De zeven personen houden nagenoeg gelijktijdig ieder twee geurdragers vast; ze houden er één in iedere hand stevig vast gedurende 1 à 2 minuten, en wisselen daarna de buizen om naar de andere hand. Vervolgens worden twee rijen gemaakt. In elke rij worden de zeven geurdragers in willekeurige volgorde gelegd. De hond ruikt vervolgens aan een voorwerp dat door persoon A van geur is voorzien en daarna wordt de hond door de hondengeleider langs de zeven geurdragers in rij 1 gevoerd. Als de hond geur A identificeert, wordt dezelfde proef herhaald bij rij 2. Correcte identificatie bij beide ronden wordt als kwalificatie voor de tweede stap gezien. De eerste stap wordt dus gebruikt om te zien of de hond 'in vorm' is. In de tweede stap wordt de geur van de controlepersoon (A) verwijderd uit beide rijen. Nu geeft de geleider zijn hond lucht van een voorwerp dat waarschijnlijk door de dader op de plek van het misdrijf is achtergelaten, alvorens de gang langs de eerste rij met geurdragers te maken. Als de hond precies bij object X een geurovereenkomst aangeeft, wordt dit als identificatie gerapporteerd. Vervolgens wordt hetzelfde gedaan bij rij 2. Als bij beide rijen correcte identificatie optreedt spreekt men van 'herkenning van de verdachte'.

Punten van kritiek

Vanuit verschillende invalshoeken zijn in het verleden kritische kanttekeningen geplaatst bij de bewijskracht van de uitkomst van een geurproef. Deze kanttekeningen hebben te maken met de filosofie achter de geurproef, de techniek van de geurproef en de vraag of de procedure wel goed gevolgd is. Een kwestie van het eerste type is bijvoorbeeld dat de conclusie die in het proces verbaal wordt opgenomen te sterk is voor wat er feitelijk gebeurt. Ook is er geen experimenteel onderzoek geweest naar de relatie tussen de perceptie van menselijke lichaamsgeur en het keuzegedrag van honden en is het een probleem dat de proef om technische redenen niet herhaald kan worden. Procedurele kwesties zijn bijvoorbeeld de vraag of de hondengeleider daadwerkelijk niet weet waar voorwerp X ligt in de rij (blinde proef). Bij sommige van deze vragen zijn goed opgezette proeven te bedenken die in statistische zin tot een antwoord kunnen leiden. De door het College van PG geformuleerde probleemstelling beperkte zich echter tot het procedurele punt van de randomisering. Worden de posities van de geurdragers wel volgens de daarvoor opgestelde procedure bepaald? Deze vraag kan op basis van de beschikbare gegevens worden beantwoord. Voordat we naar de data-analyse gaan, is het van belang om de procedure op het punt van de randomisering nog wat nader te beschouwen.

De randomisering

Zoals hierboven al aangegeven, vindt de geurproef in twee etappes plaats. Hiertoe moeten de zeven geurdragers aan een positie worden toegewezen in beide rijen. Om de keuze te maken voor deze twee rangschikkingen, zijn 36 zogenaamde uitlegschema's vastgesteld. Voorafgaand aan de proef dient het uitlegschema door loting te worden bepaald. Volgens de procedure moet dit worden gedaan door een (zuivere) dobbelsteen twee keer te werpen. Iedere mogelijke uitkomst van dit experiment is eenduidig aan een van de 36 uitlegschema's gekoppeld. Zo correspondeert een uitkomst (5,1) met een schema XFDCBEA bij de eerste rij en CAXFDBE in de tweede rij.

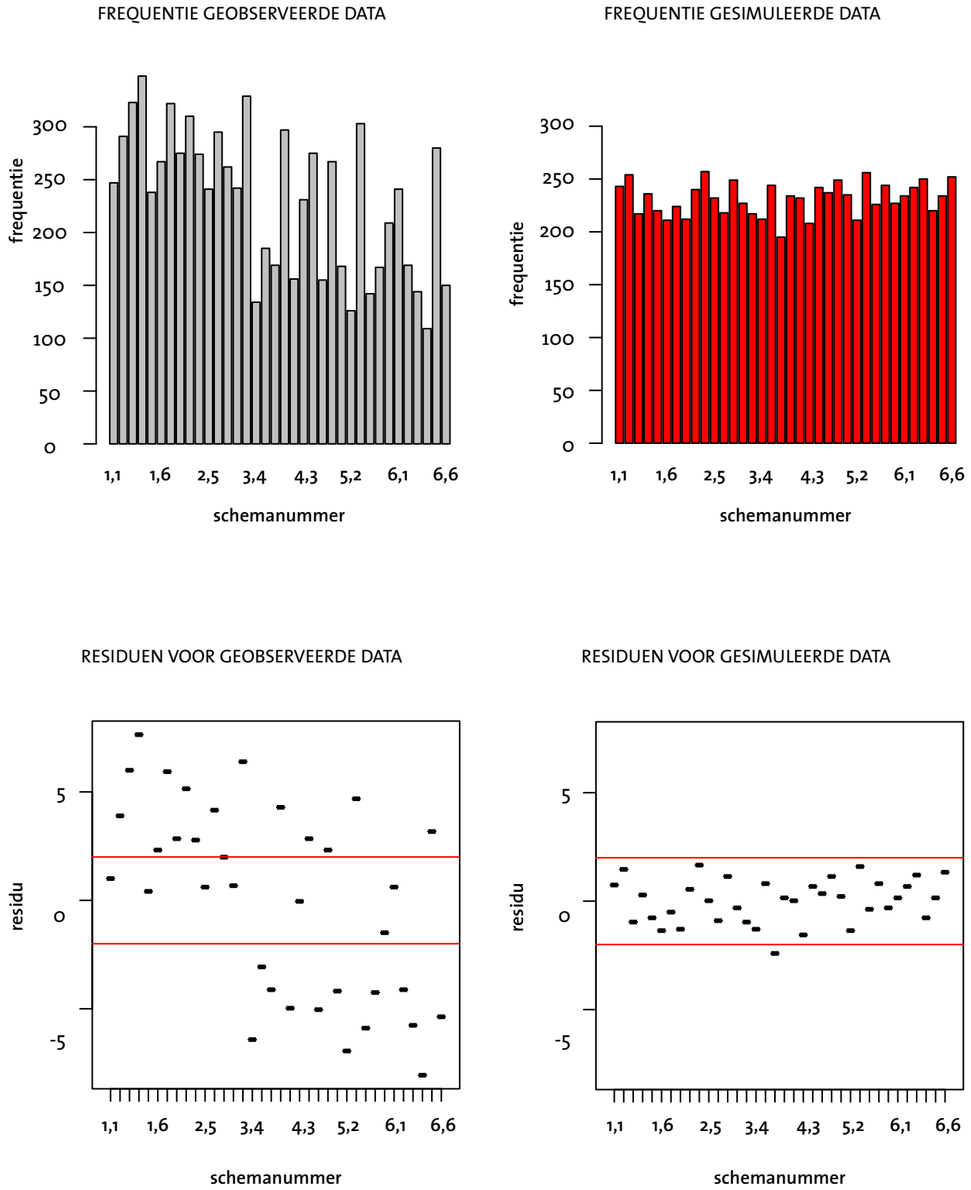
Data en toetsen

Voor het onderzoek kregen we de beschikking over de gegevens van alle 8341 geurproeven die tussen 1999 en 2006 zijn uitgevoerd door de oefengroepen Limburg, Nunspeet, Oost en Rotterdam. Van iedere proef zijn, naast het gekozen uitlegschema, ook de betrokken hondengeleider en de betrokken hond bekend. De uitlegschema's zijn gelabeld door de bijbehorende uitkomsten van de twee worpen: (1,1) tot en met (6,6). Als de procedure van dobbelen met een zuivere dobbelsteen wordt gevolgd, zal de geobserveerde vector van frequenties multinomiaal verdeeld zijn met parameters 8341 en 1/36 (voor iedere cel). Voor iedere hondengeleider geldt dat de bijbehorende vector van frequenties ook multinomiaal verdeeld is, met parameter n (aantal proeven door hem/haar gedaan) en 1/36 voor de celkansen. De toets waarvoor we hebben gekozen is de klassieke Chi-kwadraat toets voor multinomiale kansen. Deze hebben we uitgevoerd voor de gehele dataset, alsmede apart voor iedere helper, althans als de betreffende helper minimaal 180 proeven heeft gedaan. We gaan dus in feite uit van een model waarin de helpers via loting hun uitlegschema kiezen en toetsen de nulhypothese dat de kansen uniform zijn over de 36 uitlegschema's. Op andere afwijkingen van de procedure, zoals het deterministisch kiezen van uitlegschema's, wordt dus niet getoetst. De Chi-kwadraat toetsingsgrootheid gebaseerd op de gehele dataset is gelijk aan 703. Vergelijk dit met het 99,9% kwantiel van de Chi-kwadraat (35)-verdeling, welke gelijk is aan 66,6. Om voor de juristen de uitzonderlijkheid van de gerealiseerde keuzes van schema's nog wat inzichtelijker te maken, zijn frequentiediagrammen gemaakt van zowel de gerealiseerde worpen met de dobbelsteen alsook door ons gegenereerde schema's, volgens het protocol. Afwijkingen kunnen visueel worden geconstateerd. In figuren op deze pagina is dit voor de gehele dataset gedaan.

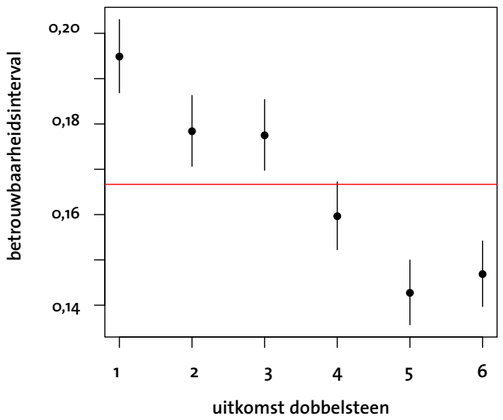
De beschreven procedure kan voor iedere helper afzonderlijk worden uitgevoerd. We hebben ons beperkt tot de 12 helpers die meer dan 180 proeven hebben uitgevoerd. Bij significantieniveau 0,01 en Bonferroni correctie voor meervoudig toetsen, blijkt dat de nulhypothese voor slechts één helper niet verworpen wordt.

Laagst- en hoogstvoorkomende frequentie per helper

Een andere manier die we gebruikt hebben om de uitkomsten van de toetsen te verduidelijken is door de eenvoudige toetsingsgrootheden laagst- en hoogstvoorkomende frequentie te beschouwen. Dit kunnen we voor iedere helper doen. We doen dit aan de hand van helper H. Er zijn gege-



Figuur 1. Linksboven: staafdiagram van de frequenties waarmee de verschillende dobbelsteenuitkomsten in de dataset voorkomen; linksonder: bijbehorend plaatje met residuen; de 2 figuren rechts zijn op dezelfde wijze verkregen, zij het dat de data nu onder de nulhypothese gesimuleerd zijn



Figuur 2. Betrouwbaarheidsintervallen voor kansen op de zes verschillende dobbelsteenuitkomsten

vens van 472 proeven voor deze helper. Bij zuiver dobbelen zouden we verwachten dat ieder van de schema's met frequentie 472/36 ≈ 13,1 voorkomt. Door toevalsvariatie zal dit nooit precies gebeuren. We concentreren ons nu op de hoogste en laagste frequentie. Deze geven immers de grootste afwijkingen van 13,1 naar boven en beneden respectievelijk. Voor helper A zijn dit 26 (bij schemanummer 2,1) en 1 (bij schemanummer 5,2). We vragen ons af hoe waarschijnlijk deze extreme frequenties zijn, indien we zouden werpen met twee zuivere dobbelstenen. Om dit te onderzoeken bootsen we het experiment met 2 zuivere dobbelstenen op de computer na. In ieder experiment simuleren we 472 worpen met 2 zuivere dobbelstenen. Vervolgens noteren we de laagste en hoogste frequentie. Dit gehele experiment herhalen we een groot aantal keren (we hebben gekozen voor 10.000 keer).

In de simulaties wordt slechts 7 keer als laagste frequentie 1 of lager verkregen.

Voor de hoogste frequentie, vinden we in 331 van de 10.000 gevallen een frequentie groter dan of gelijk aan 26.

Bij de uitgevoerde Chi-kwadraat-toets vinden we voor deze helper een veel kleinere p-waarde, namelijk 0,0000333. Dit wordt veroorzaakt doordat deze toets alle afwijkingen van 13,1 meeneemt, en we hier slechts kijken naar de laagste en hoogste frequentie.

Betrouwbaarheidsintervallen

Als we veronderstellen dat bij iedere proef een gelijkwaardige dobbelsteen is gebruikt (dat wil zeggen, met identieke kansen op een één, twee, drie, vier, vijf of zes), dan kunnen we op grond van de beschikbare data betrouwbaarheidsintervallen construeren voor de kansen op ieder van de zes dobbelsteenuitkomsten. Hiervoor zijn meerdere methoden voorgesteld in de literatuur. We maken hier gebruik van de methode zoals voorgesteld in Bailey, welke gebaseerd is op een betrouwbaarheidsinterval voor de succesparameter van een multinomiale verdeling, gebruikmakend van een variantie-stabiliserende transformatie. Een Bonferroni correctie waarborgt het gewenste betrouwbaarheidsniveau, waarvoor we 95% gekozen hebben. In figuur 2 is voor iedere mogelijke dobbelsteenuitkomst (horizontaal), verticaal het verkregen betrouwbaarheidsinterval weergegeven. De rondjes zijn de geobserveerde fracties; de horizontale lijn ter hoogte 1/6 dient als referentie voor een zuivere dobbelsteen. We beschouwen hier de gehele dataset met alle proeven, waarop ook figuur 1 gebaseerd is.

We zien dat de data een duidelijk onzuivere dobbelsteen suggereren. Als er echt gedobbeld is door de helpers, dan zien we wederom dat het onwaarschijnlijk is dat dit gebeurd is met een zuivere dobbelsteen.

Conclusie

De geobserveerde frequenties van de uitlegschema's corresponderen niet met wat je zou mogen verwachten bij werpen met een zuivere dobbelsteen, hetgeen het protocol voorschrijft.

Literatuur

Bailey, B. J. R. (1980). Large sample simultaneous confidence intervals for the multinomial probabilities based on transformations of the cell frequencies. *Technometrics* 22(4), 583–589.

Frijters, J. E. R. (2006). De geuridentificatieproef in het licht van het falsificatiebeginsel. *Nederlands Juristenblad* 17, 945--948.

Frijters, J. E. R. (2008). Dobbelen en positievoorkeuren bij canine geuridentificatieproeven. *Expertise en Recht* 2008-1, 27--34.

Keuringsreglement politiespeurhond menselijke geur. Bijlage 1 bij de regeling politiehonden.

Geurt Jongbloed *is hoogleraar Mathematische Statistiek aan de Technische Universiteit Delft.*
E-mail: <G.Jongbloed@tudelft.nl>

Frank van der Meulen *is universitair docent Statistiek aan de Technische Universiteit Delft.*
E-mail: <f.h.vandermeulen@tudelft.nl>

KNAPPE KOPPEN IN OR BELOOND

Promovendi in OR maken kans op ORTEC Excellence in Advanced Planning Award 2012

Ook in 2012 worden knappe koppen in OR weer beloond voor hun promotieonderzoek. ORTEC en het Nederlandse Genootschap voor Besliskunde (NGB) willen promovendi stimuleren in hun onderzoek met de ORTEC Excellence in Advanced Planning Award 2012. 'Die stimulans blijkt nodig, omdat er soms nog een gat bestaat tussen de theoretische kant van OR en de toepassing van de methoden in de praktijk', aldus Gerrit Timmer, mede-oprichter en CFO van ORTEC. 'Het blijkt dat de innovatieve ideeën die uit de onderzoeken van participanten naar voren komen, veelal ook weer toepasbaar zijn in de praktijk (bij onze klanten). Beter dan op deze manier kan de brug tussen praktijk en theorie niet geslagen worden."

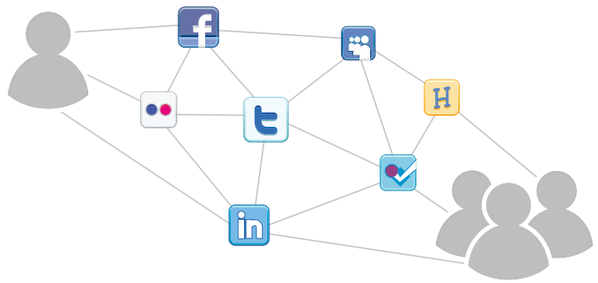
Deelname

Promovendi kunnen deelnemen aan de competitie door het insturen van een document waarin kort (in maximaal 3 tot 4 pagina's) beschreven wordt wat zij met hun OR-onderzoek hebben bereikt. Het document dienen zij voor 30 november 2011 in te sturen naar: secretariaat@ngb-online.nl. Meer informatie over deelname is te vinden op www.ngb-online.nl.

LNMB-NGB Lunteren 17-19 januari 2012

In de week van 19 december zullen de finalisten persoonlijk worden benaderd. Zij zullen worden uitgenodigd een presentatie te houden op de LNMB-NGB Lunteren-conferentie, die van 17 tot en met 19 januari 2012 plaatsvindt. Tijdens de laatste dag van deze conferentie zal tevens de winnaar van de ORTEC Excellence in Advanced Planning Award 2012 bekend worden gemaakt. De winnaar mag een prijs in ontvangst nemen van € 1000,-. Tevens krijgt hij of zij een publicatie van het onderzoek in het blad *STAtOR*.

ANALYSE VAN SOCIALE MEDIA



Het gebruik van sociale media overtreft al enkele jaren in ruime mate het gebruik van klassieke media als e-mail (Nielsen, 2009). Voor velen is een bestaan zonder Facebook, Flickr, Hyves, MySpace, FourSquare, LinkedIn en Twitter dan ook niet meer denkbaar. Door de sporen die elke gebruiker van sociale media achterlaat in de vorm van berichtjes, foto's, profielen en vriendschapsrelaties ontstaat er een enorme digitale schatkamer aan informatie. Bij TNO richten we ons al geruime tijd op de ontwikkeling van technieken voor de automatische analyse van sociale media. In deze bijdrage beschrijven we technieken als sentimentanalyse, netwerkanalyse en datavisualisatie. Deze technieken stellen analisten in staat een indruk te krijgen van de activiteit binnen sociale media.

Erik Boertjes, Almerima Jamakovic & Stephan Raaijmakers

Sociale media bevatten veel en uiteenlopende informatie. Doorgaans zijn sociale media georganiseerd als netwerken. Leden kunnen met andere leden expliciete vriendschapsrelaties aangaan, door elkaar op te nemen in de vriendenlijst. Ook impliciete vriendschapsrelaties zijn mogelijk: personen kunnen hun waardering voor een ander lid uitdrukken zonder dat die persoon in hun vriendenlijst voorkomt.

De netwerkstructuur van sociale media kan formeel worden geanalyseerd, waarbij allerlei netwerkeigenschappen gemeten worden. Deze netwerkeigenschappen bepalen voor een groot deel welke processen zich afspelen op het netwerk. Zo is bijvoorbeeld de snelheid waarmee informatie zich door het sociale netwerk verspreidt vooral afhankelijk van de onderliggende netwerkstructuur. Recente vooruitgangen in socialenetwerkanalyse (Wasserman en Faust (1994); Scott en Carrington (2008)) hebben een verscheidenheid aan praktisch belangrijke netwerkmetrieken opgeleverd die inzetbaar zijn bij onderzoek naar sociale verschijnselen.

Een relevant vraagstuk voor het veiligheidsdomein betreft de mogelijkheid om de autoriteit van personen binnen sociale media af te leiden uit geobserveerd gedrag en reacties daarop. Het is interessant om te kunnen inschatten wie de 'opinieleiders' binnen een bepaald forum zijn: personen met veel invloed en volgelingen. Een hieraan gerelateerde vraag is: hoe komt sociale beïnvloeding via sociale media tot stand? Kunnen we voorspellen wat de impact van bepaalde berichten is, en wat het nut zou zijn van interventies in online discussies?

Het visualiseren (inzichtelijk en doorzoekbaar maken) van de grote hoeveelheden informatie binnen sociale media is eveneens een onderzoeksonderwerp. Ten slotte is *anomaliedetectie* in sociale media een interessant onderwerp: kunnen we plotselinge verschuivingen van patronen waarnemen, zoals veranderingen in sentiment of onderwerp?

We gaan in de volgende secties beknopt in op deze vraagstukken.

Autoriteit in sociale media

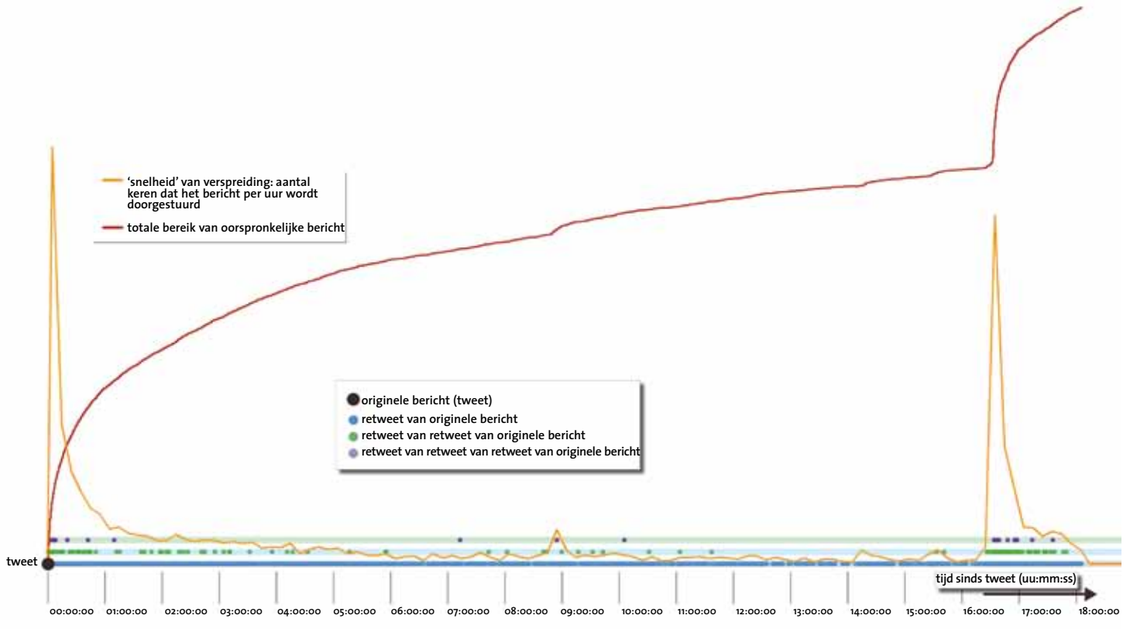
Waar tekst wordt geproduceerd, is sentiment te vinden: een vaak beknopte expressie van gevoelens en meningen. Bekende voorbeelden zijn te vinden op de vele consumentensites, zoals kieskeurig.nl (bv. elektronica) of rottentomatoes.com (filmbesprekingen). Sentiment is tot op zekere hoogte gecorreleerd met autoriteit: wie veel autoriteit heeft binnen een bepaalde gemeenschap oogst overwegend veel bijval in de vorm van bijvoorbeeld positief commentaar. Daarbij speelt ook het aantal connecties in de sociale context een rol: wie veel bijval krijgt van veel verschillende personen heeft vermoedelijk meer impact dan iemand die slechts van enkele digitale vrienden bijval krijgt. Bijval van personen die zelf ook veel impact hebben lijkt ook belangrijker dan bijval komend van laag-autoritaire personen. Sentiment wordt vaak uitgedrukt als *polariteit* op een driepuntsschaal: negatief, neutraal, of positief, maar ook andere schaalverdelingen zijn mogelijk, zoals gradaties van positiviteit op een 5-puntsschaal.

Dit proces kan worden gemodelleerd met een gewogen graaf, met als knopen personen, en als relaties commentaar, waarbij de relaties gewogen worden met de sentimentwaarde van het commentaar. Een sterk positief commentaar kan bijvoorbeeld een hoger 'gewicht' krijgen dan een lauw commentaar. Een dergelijk netwerkgebaseerd model van autoriteit lijkt sterk op *PageRank* (Brin en Page, 1998), de bekende methode van Google om webpagina's te wegen op basis van binnenkomende en uitgaande verwijzingen (links). Door commentaar te beschouwen als binnenkomende links (ontvangen commentaar) of uitgaande links (gegeven commentaar) en ze te 'kleuren' met de sentimentwaarde van het commentaar (positief, negatief) verkrijgen we een met sentiment gewogen versie van PageRank. Deze gewogen graaf leent zich vervolgens voor het uitrekenen van allerlei netwerkmetrieken, zoals de *centrality metric betweenness* (Wasserman en Faust, 1994). Ook methoden om het verspreidingsgedrag van informatie (informatiediffusie) te modelleren kunnen op dit soort grafen worden toegepast (e.g. Yang en Leskovec, 2010).

Hoe kunnen we sentiment automatisch herkennen in teksten? Al meer dan 10 jaar is 'sentimentanalyse' een hot topic binnen de *machine learning*-gemeenschap (zie Pang en Lee, 2008) voor een uitgebreid overzicht). *Machine learning* richt zich op het automatisch laten uitvoeren door computers van analysetaken, doorgaans op basis van door de mens geprepareerd 'trainingsmateriaal'. Speciale 'zelflerende' algoritmen zijn in staat om uit deze, met klassen gelabelde, voorbeelden een model te leren. Aan de hand van zo'n model kunnen computers zelf nieuwe, niet eerder gezien gevallen naar analogie van de trainingsdata labelen of *classificeren*. Voor sentimentanalyse kunnen zelflerende systemen getraind worden op teksten gekoppeld aan polariteitslabels, zoals 'positief', 'neutraal', of 'negatief'. Voorbeelden zijn[1]:

> NEG: Everything in the phantom you have seen many times before and there is nothing new presented here. If you're looking for a fun family movie, go watch the underrated flipper. This is not a good movie.
>
> POS: The movie reminds us of the sacrifices made by our WW II fighting men and women. We must not ever forget them as many gave the ultimate sacrifice, their lives so that we may live in freedom today. For this I thank them and for steve spielberg for making a movie that I will never forget.

We hebben een door PageRank geïnspireerde autoriteitsmeting geïmplementeerd in de vorm van een *content ranking*-toepassing voor Flickr, een sociaal netwerk voor het delen van foto's. Personen kunnen in Flickr, behalve via expliciete relaties, ook impliciet aan elkaar gerelateerd zijn doordat ze elkaars foto's van commentaar voorzien. Van dit commentaar hebben we het sentiment geanaly-



Figuur 1. Verspreiding van een tweet door het twitternetwerk

seerd, waarna we per persoon een ranking hebben berekend. Foto's op Flickr worden door ons systeem geordend op basis van de aldus voorspelde autoriteit van de uploader ervan. De autoriteit van uploaders bepalen we overigens binnen onderwerpsclusters: we clusteren automatisch foto's op basis van de door gebruikers eraan toegekende *tags*.

Visueel inzicht in sociale media

Visualisaties kunnen inzicht geven in de verspreiding van een boodschap door een sociaal netwerk. Zogenoemde *retweets* spelen daarin een cruciale rol. Een retweet is het doorsturen van een bericht (tweet) van iemand anders. Wanneer iemand een bericht op Twitter publiceert, bereikt dat bericht in eerste instantie de 'volgers' van die persoon: de mensen die expliciet hebben aangegeven geïnteresseerd te zijn in wat die persoon te melden heeft. Wanneer een volger het betreffende bericht doorstuurt, bereikt het bericht daarmee niet alleen de volgers van de oorspronkelijke zender, maar ook van degene die het bericht doorstuurt. Ook een retweet kan op zijn beurt weer worden doorgestuurd. Het oorspronkelijke bericht verspreidt zich op die manier door het Twitternetwerk.

De visualisatie in figuur 1 laat zien hoe een origineel bericht door de tijd wordt geretweet[2]. De rode lijn laat het verloop van het totale bereik zien: het totaal aantal Twittergebruikers dat het oorspronkelijke bericht onder ogen krijgt. De oranje lijn toont de snelheid van verspreiding die volgt uit de helling van de rode grafiek. Te zien is dat de grafiek een grote sprong maakt: dat is een moment waarop iemand met een groot aantal volgers het oorspronkelijke bericht doorstuurt (waarmee het totale bereik fors toeneemt).

Visualisaties als deze kunnen helpen bij het in kaart brengen van de dynamiek van informatieverspreiding, naast netwerktopologische en berichtinhoudelijke analyse.

Anomaliedetectie

De detectie van afwijkende patronen in sociale media is een onderwerp dat interessante monitormogelijkheden biedt voor forensische partijen. Ook hier kan sentimentanalyse een interessante bijdrage leveren. Het plotseling omslaan van het sentiment op een bepaald forum rond een bepaald onderwerp kan bijvoorbeeld aanleiding zijn voor nader onderzoek: is er sprake van haatzaaien? Heeft een bepaalde gebeurtenis in de buitenwereld tot een vorm van radicalisering geleid? Wordt er opgeroepen tot rellen?

Clusteringstechnieken zoals *Non-Negative Matrix Factorization* (Lee en Seung, 1999) kunnen een continue stroom van tekstberichten onderverdelen in sentimentclusters (positief, negatief). De clusterdistributie kan worden gemeten door de tijd heen. Als er een plotselinge migratie van berichten optreedt van het ene cluster naar het andere (bijvoorbeeld van positief naar negatief) kunnen we dit visualiseren met een piek in een sentimentcurve: een tijdsreeks die het verloop van het sentiment verdeeld over clusters weergeeft. Toegepast op voetbaltweets waren we op deze manier in staat doelpunten te herkennen in een wedstrijd tussen Feyenoord en FC Utrecht aan de plotselinge erupties van (doorgaans positief) sentiment.

Conclusies

In deze bijdrage hebben wij een aantal vraagstukken op het gebied van socialemedia-analyse beknopt geadresseerd. Netwerkanalyse, sentimentanalyse en visualisatie kunnen worden ingezet om een indruk te geven van wat er leeft binnen bepaalde sociale media. Een weloverwogen koppeling van deze technieken kan het werk van forensisch specialisten ondersteunen. Het blijft uiteindelijk aan de mens om de resultaten van automatische analyse op hun waarde te beoordelen. Hoe slim de computer ook wordt in het nemen van de gevoelstemperatuur van sociale media, de menselijke analist zal altijd de doorslag geven bij het beoordelen van informatie en het eventueel in gang zetten van operationele acties.

Noten

1. *Bron: http://www.cs.cornell.edu/People/pabo/movie-review-data/*
2. *Gebaseerd op door ons verzamelde Twitterberichten over de recente gebeurtenissen in Libië.*

Literatuur

Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer networks and ISDN systems* 30(1-7), 107–117.

Lee, D. D. & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755), 788–791.

Nielsen (2009). *Global faces and networked places, A Nielsen report on social networking's new global footprint.* The Nielsen Company.

Pang, B. & Lee, L. (2008). Foundations and trends in information retrieval 2(1-2), pp. 1–135.

Scott, J. & Carrington, P. (Eds.), (2008). *The Sage handbook of social network analysis*. Sage Publications.

Wasserman, S. & Faust, K.(1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Yang, J. & Leskovec, J. (2010). Modeling Information Diffusion in Implicit Networks. In: *Proceedings ICDM'10*, Sydney, Australia.

Erik Boertjes *studeerde Informatica aan de Universiteit Twente en behaalde een Bachelor graad aan de Academie voor de Beeldende Kunsten in Den Haag. Bij TNO houdt hij zich voornamelijk bezig met het ontwerpen en realiseren van data-visualisaties. E-mail: <erik.boertjes@tno.nl>.*

Stephan Raaijmakers *is machine learning-onderzoeker, en promoveerde in 2009 aan de Universiteit van Tilburg op multinomiale methodes voor tekstanalyse met zelflerende systemen. Bij TNO richt hij zich onder andere op sociale netwerk analyse, waaronder sentimentanalyse en anomaliedetectie. E-mail: <stephan.raaijmakers@tno.nl>.*

Almerima Jamakovic *promoveerde aan de Faculteit Elektrotechniek, Wiskunde en Informatica van de Technische Universiteit Delft op het onderwerp robuustheid van complexe netwerken. Bij TNO doet ze onderzoek naar de performance en robuustheid van ICT systemen en telecomnetwerken. E-mail: <almerima.jamakovic@tno.nl>.*

Codex Hammurabi op een zuil uit het antieke Mesopotamië . Collectie Louvre Parijs

VEILIGHEID IN HET ZIEKENHUIS

DRAAGT OPERATIONS RESEARCH BIJ AAN BETERE ZORG? (DEEL 1)

Joris van de Klundert

Ongeveer tien jaar geleden gaf het invloedrijke Amerikaanse Institute of Medicine twee rapporten: *Crossing the quality chasm* (Committee on Quality of Health Care in America, 2001) en *To err is human* (Kohn e.a., 2001). Het eerste rapport beschrijft de kwaliteit van zorg en benoemt veiligheid als een van de zes dimensies van kwaliteit van zorg. Het tweede rapport behandelt expliciet veiligheid en becijfert dat er in de Verenigde Staten in ziekenhuizen jaarlijks tussen de 44.000 en 98.000 mensen overlijden als gevolg van vermijdbare schade. Deze beide rapporten hebben geneeskundigen en gezondheidswetenschappers in beweging gebracht en hebben de focus van *health services research* wereldwijd verlegd. Aan Health service operations researchers lijkt de boodschap vooralsnog grotendeels voorbij gegaan. Dit artikel laat zien waarom dat zo is en verkent hoe operations research tot een grotere bijdrage aan kwaliteit van zorg kan leiden.

Definitie van 'kwaliteit van zorg'

Voor de goede orde, kwaliteit en veiligheid van zorg zijn geen nieuwlichterij. In de codex van Hammurabi wordt al uitgebreid aandacht besteed aan staaroperaties die ook nu nog wereldwijd tot de meest uitgevoerde operaties behoren. De codex van Hammurabi regelt de veiligheid op juridische wijze door middel van het straffen van operateurs wanneer de operatie lijdt tot vervolgschade. In het ergste geval kon de straf bestaan uit het afhakken van de handen van de operateur.

Alvorens de analyse aan te vangen, is het goed stil te staan bij het ruime begrip kwaliteit van zorg. Wat dienen we te verstaan onder kwaliteit van zorg? Nu we toch in de oudheid zijn, is het passend om het van origine nogal abstracte begrip kwaliteit te verkennen, samen met Plato die er als eerste aandacht aan besteedt. In *De Staat* bespreekt Plato het begrip ποίοτεσ (aard, kwaliteit) in de context van 'het goede'(κάλον): 'Een algemene beschrijving van de eigenschappen die een object, handeling, persoon, of prestatie van welke aard dan ook, dient te hebben om terecht te kunnen worden gekarakteriseerd als hoogwaardig, of als goed te kunnen worden gewaardeerd'. Deze definitie heeft nog niets aan actualiteit verloren, en past prachtig op de complexe aard van zorgverlening. Zorgverlening is immers een dienst en kan daarom worden gezien als object, als handeling en als prestatie, waarvan de kwaliteit wordt gerelateerd aan de persoon of personen die de dienst leveren.

Het Institute of Medicine definieert de kwaliteit van zorg als 'de mate waarin zorgverlening aan individuen en de bevolkingsgroepen tot gewenste gezondheidsuitkomsten leidt' en constateert dat de kwaliteit van zorgverlening door de volgende zes dimensies kan worden gekarakteriseerd (Committee on Quality of Health Care in America, 2001):

1. Doelmatigheid (de beoogde gezondheidseffecten worden bereikt)
2. Veiligheid (afwezigheid van vermijdbare negatieve gezondheidseffecten)
3. Patiëntoriëntatie (normen en waarden van de patiënt staan centraal in de besluitvorming)
4. Efficiëntie (geen verspilling van geld of andere middelen)
5. Gelijkgerechtigdheid (eenieder heeft toegang tot dezelfde zorg)
6. Tijdigheid (levering zonder vertraging – te meer daar vertraging de gezondheid negatief kan beïnvloeden)

Kwaliteitskloof

Het Institute of Medicine spreekt van een *quality chasm*, een kwaliteitskloof, daarmee uitdrukking gevend aan het grote verschil tussen de kwaliteit van de in de Verenigde Staten geleverde zorg, en de kwaliteit van zorg die zij in werkelijkheid wenselijk acht. Met betrekking tot de veiligheid van zorg, wordt het daaruit voortvloeiende appel verder versterkt door het rapport *To err is human* (Kohn e.a., 2001) waaruit duidelijk wordt dat jaarlijkse tienduizenden mensen overlijden als gevolg van onveilige zorg. Daarnaast leidt onveilige zorg tot tijdelijk of blijvend gezondheidsverlies bij een nog veel groter aantal mensen.

De ontwikkelingen die in de Verenigde Staten tot beide voornoemde rapporten hebben geleid, hebben ook de Nederlandse gezondheidszorg bereikt. Daarbij zij echter opgemerkt dat de Nederlandse gezondheidszorg beter lijkt te presteren dan de gezondheidszorg in de VS In ieder geval geldt dit voor de laatste drie van de zes bovenstaande dimensies. We geven een aanzienlijker kleiner deel van ons bruto nationaal product uit aan gezondheidszorg, we kennen een verplichte zorgverzekering die iedereen gelijke toegang tot zorg biedt en de wachttijden

bij vooral de spoedeisende hulp zijn veel lager. Nederland staat met afstand op de eerste plaats in de European Health Consumer Index. Ook in Nederland overlijden echter jaarlijks vele patiënten in ziekenhuizen als gevolg van vermijdbare schade (Wagner, & de Bruijne, 2007). In 2007 is dit aantal becijferd op ruwweg 1500 tot 2000, hetgeen verhoudingsgewijs niet veel verschilt van de aantallen die het IoM noemt ten aanzien van de VS. Dit aantal is bijvoorbeeld ruim hoger dan het aantal dodelijke verkeersslachtoffers. De maatschappelijke bewustwording en urgentie ten aanzien van de veiligheid van zorg groeit, zoals ook blijkt uit de niet aflatende aandacht voor dit onderwerp in de media. Naar aanleiding van de genoemde bevindingen is in Nederland dan ook besloten tot een aantal verbetermaatregelen, waaronder de invoering van veiligheidsmanagementsystemen. In 2010 is het eerdere onderzoek herhaald (Langelaan e.a., 2010) en er is echter geen significante verbetering geconstateerd.

We komen daarmee op een belangrijk punt. De oplossingen voor het probleem 'onveilige zorg' zijn niet eenvoudig. Deze boodschap is ook impliciet aanwezig in de titel van het betreffende IoM rapport *To err is human* (Kohn e.a., 2001). Onveilige zorg is vaak het gevolg van menselijke fouten: mensen maken fouten en zullen ook fouten blijven maken. Oplossingen zijn dus niet gelegen in het nemen van maatregelen die beogen menselijke fouten te elimineren, zoals bijvoorbeeld beoogd lijkt in de codex van Hammurabi. Oplossingen liggen in het ontwikkelen van organisaties/systemen waarbinnen menselijke fouten niet tot onveiligheid leiden. De vraag rijst vervolgens welke oplossingen effectief zijn, en welke bijdrage operations research daaraan kan leveren. Onderstaand gaan we daarop in. Eveneens gaan we in meer algemene zin in op de vraag of en hoe operations research de kwaliteit van de zorg kan verbeteren.

Abstracte modellen

Operations research methoden grijpen voor een belangrijk deel terug op de methode waarvan ook Plato zich bediende, de logica. Operations researchers formuleren via assumpties en axioma's een abstract model van de werkelijkheid, en beschouwen vervolgens de waarheid zoals die geldt voor het model. Voor het model kunnen vervolgens stellingen worden geformuleerd en bewezen. Het vergaren van kennis over deze abstracte wereld is enerzijds waardevol op zich. Anderzijds kan zij ook waardevol zijn door het gebruik van deze kennis in de werkelijkheid, bijvoorbeeld om de kwaliteit van zorgverlening te verbeteren. Zo kan kennis over wachtrijmodellen worden ingezet om de tijdigheid en/of de efficiëntie van zorg te verbeteren.

In zijn vergelijking van de grot laat Plato ons al zien dat kennis over de abstracte ideeënwereld niet altijd goed wordt ontvangen. Hij beschrijft hoe Socrates door de Atheners wordt aangeklaagd voor het bederven van de jeugd, en beschrijft de onsuccesvolle verdediging van Socrates. Socrates beweert dat hij zichzelf impopulair heeft gemaakt door zijn kennis over de abstracte, goede, ideeënwereld naar de rauwe dagelijkse werkelijkheid te brengen, en laat zien hoe de kloof daartussen zo groot is dat machtige personen die deze ideeënwereld niet kennen er door ontstemd zijn geraakt.

Wees gerust, de pointe van mijn betoog zal niet zijn dat operations researchers noodzakelijkerwijs dezelfde niet-patient georiënteerde behandeling en gezondheidsschade wacht als Socrates. De normen en waarden van de operations research gemeenschap hechten echter veel belang aan abstractie en theorie. Dit blijkt bijvoorbeeld uit het prestige en de impactfactoren van de operations research-tijdschriften. Tijdschriften zoals *Interfaces*, dat expliciet veel belang hecht aan verkregen resultaten in de werkelijkheid – aan empirie, kennen een lage impactfactor (*Interfaces* heeft een impactfactor van 0,593) en staat daarmee op een bescheiden 52ste plaats in de ranglijst van OR/MS-tijdschriften.

Evidence based

In de laatste twintig jaar hebben de geneeskunde en de gezondheidswetenschappen een ontwikkeling doorgemaakt die daar loodrecht op staat. Deze ontwikkeling gaat uit van *evidence base*. Zij zoekt bewijskracht, *evidence*, niet in een logisch bewijs dat geldig is binnen een abstract model, maar in (statistisch) significante verbetering van zorgverlening in de praktijk. Met name *evidence-based medicine* (Sackett e.a., 1996), dat nu zo'n 25 jaar als de norm geldt, heeft een grote vlucht genomen. In het Verenigd Koninkrijk en in Nederland geldt bijvoorbeeld dat doelmatigheid van zorg evidence based moet zijn voor zorg in het basispakket, en gelden voor tal van ziektebeelden evidence based-richtlijnen voor behandeling die zorgverleners worden geacht toe te passen. Deze empirische bewijsvoering is gaan prevaleren boven alternatieven zoals natuurwetenschappelijke logische redeneringen, of de individuele ervaring van de arts, juist vanwege de aantoonbaar betere gezondheidsuitkomsten. Aan deze empirische attitude ligt de complexiteit van het menselijk lichaam en de menselijke gezondheid ten grondslag.

Iedere behandeling, iedere interventie kan naast de effecten die op grond van bestaande natuurwetenschappelijke kennis of individuele ervaring verwacht worden, nog tal van andere effecten hebben. Daarom worden heden ten dage vooral robuuste empirische evaluaties van interventies als valide bewijs beschouwd. Als gouden standaard geldt daarbij de double blind randomized controlled trial (RCT). In een randomized controlled trial wordt een patiëntenpopulatie verdeeld in een interventiegroep en een controlegroep. Personen in de interventiegroep ontvangen de te evalueren interventie, en de controlegroep het bestaande alternatief, of geen behandeling. De RCT is dubbel blind wanneer noch de betrokken behandelaars noch de patiënten weten in welke groep de patiënt zit. Om statistisch significante uitkomsten te bereiken waarin gecorrigeerd kan worden voor redelijkerwijs te verwachten ander beïnvloedende factoren, is het van belang dat de populatie van voldoende omvang is. Zogeheten meta analysis en/of systematic reviews, waarin de empirische bewijskracht van verschillende studies, RCTs en andere studies, worden gecombineerd, gelden als nog sterkere *evidence*. In de medische en *health services*-literatuur geldt vervolgens dat juist publicaties en tijdschriften met empirische validiteit gerespecteerd worden en hoge impact hebben, en theoretische bijdragen steeds meer een rol in de marge spelen.

Kan operations research dan toch nog een gewaardeerde rol spelen in health services research? Daarover volgende keer meer in deel 2.

Committee on Quality of Health Care in America (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, D.C.: Institute of Medicine, National Academic Press.

Kohn, L. T., Corrigan, J. M. & Donaldson, M. S. (Eds.), (2001). *To err is human: Building a safer health system. Committee on quality of health care in America*. Washington, D.C.: Institute of Medicine, National Academic Press.

Langelaan, M., Baines, R. J., Siemerking, K. M., Steeg, L. van de, Asscheman, H., Bruijne, M. C. de & Wagner, C. (2010). *Monitor zorggerelateerde schade 2008. Dossieronderzoek in Nederlandse ziekenhuizen*. Utrecht: Nivel; Amsterdam: EMGO.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it isn't. *British Medical Journal*, 312, 71–72.

Wagner, C. & Bruijne, M. de (2007). *Onbedoelde schade in Nederlandse ziekenhuizen*. Utrecht: Nivel.

Joris van de Klundert is hoogleraar Management & Organisatie van Zorgverlening bij het Instituut Beleid & Management van de Gezondheidszorg van de Erasmus Universiteit Rotterdam
E-mail: <vandeklundert@bmg.eur.nl>

Foto: Nederlands Forensisch Instituut

OP ZOEK NAAR HET DNA-SPOOR

De database-controverse

Ronald Meester

Op een plaats delict wordt een DNA-spoor veiliggesteld waarvan men aanneemt dat het afkomstig is van de dader. Dit spoor wordt vervolgens vergeleken met de DNA-profielen die in de Nederlandse database zitten; dat zijn er op dit moment ruim 120.000. De zoektocht in de database kan verschillende resultaten tot gevolg hebben. Allereerst kan het gebeuren dat er helemaal geen match is; er is dan geen enkel profiel in de database dat overeenkomt met het achtergelaten spoor. Daarnaast kan er ook precies één match zijn. Het gevolg van een dergelijke match is dat de donor van het desbetreffende profiel automatisch tot verdachte wordt gepromoveerd. Het kan ook gebeuren dat er meerdere matches zijn, vooral wanneer het aangetroffen spoor van slechte kwaliteit is, of bij een zogenoemd mengspoor, waarin DNA van verschillende bronnen door elkaar is geraakt. Deze mogelijkheid van meerdere matches laten we hier even buiten beschouwing.

Bewijskracht van een match

De meeste mensen denken dat een match onmiddellijk leidt tot de oplossing van het delict, maar dat is zeker niet het geval. Er zijn ruwweg twee obstakels.

1. Het is niet altijd even duidelijk hoe een match geïnterpreteerd moet worden. Wat is eigenlijk de 'bewijskracht' van een match, kunnen we spreken over de kans dat de verdachte inderdaad de dader is, en wat betekent 'kans' in dit verband dan precies?
2. Ook als de afkomst van het DNA-spoor niet wordt betwist, kan er onenigheid bestaan over de vraag hoe het spoor op de plaats van het delict is terechtgekomen. De verdachte zal aan kunnen voeren dat het weliswaar zijn DNA is, maar dat hij niet ter plekke is geweest. Als je je realiseert hoe kwistig wij zijn met het verspreiden van ons DNA op allerlei manieren, dan zal duidelijk zijn dat dit voor de verdediging een sterk punt kan zijn. Het gevolg is dat in zaken waarbij DNA het enige bewijsmateriaal vormt, doorgaans geen veroordeling kan en zal plaatsvinden.

We beperken ons tot de vragen die onder 1. gesteld worden: wat is eigenlijk de bewijskracht van een match en kunnen we nu spreken over een bepaalde kans dat de match inderdaad tot de donor van het spoor leidt? De verhitte discussie over dit punt is op internet goed te volgen wanneer je zoekt naar *the database controversy*. De discussie spitste zich toe op de vergelijking tussen de bewijskracht van een database-match en een match met een willekeurige persoon wiens DNA toevallig bekeken is, bijvoorbeeld als gevolg van een veroordeling van een ongerelateerd delict. Deze laatste situatie wordt wel de cold case genoemd. Er zijn nu ruwweg twee kampen te onderscheiden.

In het eerste kamp wordt beweerd dat de bewijskracht van een database-match beduidend *kleiner* moet zijn dan die van een cold-casematch, omdat er in een grote database een significante kans bestaat op een toevalsmatch, dat wil zeggen een match met een persoon die niets met het misdrijf te maken heeft. Inderdaad: als de database uit 1 miljoen profielen bestaat, en de populatiefrequentie van het spoor is ook 1 op de miljoen, dan is er een kans van ongeveer 0,63 dat in een database van uitsluitend onschuldige personen toch een match optreedt. Volgens proponenten van dit gezichtspunt betekent zo'n match in een database dus niet zo veel.

In het tweede kamp beweert men dat een database-match juist een veel *grotere* bewijskracht heeft dan een cold-casematch. Immers, zo redeneert men, de database-match geeft, naast de informatie dat iemand een match geeft, ook nog eens de extra informatie dat al die andere mensen in de database *niet* matchen. Dat is dus veel meer informatie, en de bewijskracht wordt daarmee navenant groter.

Hypotheses

Beide kampen kunnen rekenen op steun uit gewichtige kringen. Het eerste kamp wordt bijvoorbeeld vertegenwoordigd door de National Research Council in de VS in 1996, de tweede door enkele vooraanstaande Engelse statistici. Maar wie heeft er nu gelijk? Tja, dat blijkt niet zo eenvoudig te liggen. De bewijskracht van bewijsmateriaal wordt doorgaans gekwantificeerd door

een zogenoemde *likelihood ratio* (LR). Men stelt twee hypotheses of scenario's op, H_p en H_d, die verwijzen naar de claim dat de persoon die een match geeft, wel respectievelijk niet de donor van het aangetroffen spoor is. (De notatie is goed te onthouden door je te realiseren dat de 'p' staat voor *prosecutor* en de 'd' voor *defence*.) Men berekent dan de kans – binnen een nauwkeurig geformuleerd wiskundig model – op de match gegeven beide scenario's en beschouwt het quotiënt

$$\frac{\text{de kans op het bewijs gegeven } H_p}{\text{de kans op het bewijs gegeven } H_d}$$

Hoe groter dit quotiënt, hoe groter de bewijskracht. In de cold case bijvoorbeeld bestaat het bewijsmateriaal uit een enkele match, en is de teller gelijk aan 1, en de noemer gelijk aan de kans dat een willekeurig gekozen persoon het gevonden profiel heeft. Als we die kans even p noemen dan is de LR in de cold case dus gelijk aan $1/p$. Bij een unieke database-match is het bewijsmateriaal die ene match, samen met het feit dat alle andere personen in de database niet matchen. Nu blijkt – onder de aanname dat iedereen in de database a priori even waarschijnlijk de donor van het gevonden spoor is – de LR gelijk te zijn aan $(N-1)/p(N-n)$ waarbij N de populatiegrootte en n de grootte van de database voorstelt. Dit laatste getal is groter dan $1/p$ en daarmee lijkt het pleit beslecht in het voordeel van het tweede kamp.

Maar nee, niets is minder waar en wel om de volgende reden. Het is algemeen bekend dat het in statistische zaken niet toegestaan is om een hypothese die je wilt onderzoeken op te stellen aan de hand van data die je hebt, en vervolgens diezelfde data de hypothese te laten bevestigen. En dat is nu precies wat er in de database situatie gebeurt. Immers, de hypothese dat een bepaalde persoon de donor van het spoor is wordt pas opgesteld *nadat* gebleken is dat zijn DNA-profiel overeenkomt met het gevonden spoor. Om deze reden vindt het eerste kamp dat je alleen zogenaamde 'data-onafhankelijke' hypotheses mag toetsen, hypotheses die je op kunt stellen zonder de data gezien te hebben. Meestal wordt dan gesuggereerd om de hypothese 'de database bevat de donor van het spoor' versus 'de database bevat de donor niet' te beschouwen. Voor deze twee hypotheses kunnen we ook weer een likelihood ratio uitrekenen, en deze blijkt gelijk te zijn aan $1/np$, dus veel kleiner dan $1/p$. Volgens dit kamp verkrijg je dus de bewijskracht in het database geval door de bewijskracht van de cold case te delen door de grootte van de database; dit was precies de aanbeveling van de National Research Council in de VS in 1996.

De verwarring wordt nu compleet wanneer je je vervolgens realiseert dat wanneer bijvoorbeeld persoon X als enige een match geeft in een database-zoekactie, de hypotheses 'de database bevat de crimineel' enerzijds en 'X is de crimineel' equivalent zijn! We hebben dus twee equivalente hypotheses (na het zien van de match) die desondanks totaal verschillende bewijskracht hebben. Zowaar een verwarrende situatie, want welke hypothese moet je nou nemen? En welke bewijskracht is nu de juiste?

Schuldig – niet schuldig

Voor het antwoord op deze vraag is het belangrijk je te realiseren dat de gewoonte om alleen likelihood ratios te presenteren nogal gevaarlijk is. Immers, uiteindelijk zijn we niet geïnteresseerd in de likelihood ratio, maar in de kans dat een verdachte schuldig is gegeven het bewijs, dus in

$$\frac{\text{de kans dat verdachte schuldig is gegeven het bewijs}}{\text{de kans dat verdachte niet schuldig is gegeven het bewijs}}$$

Dit laatste quotiënt wordt de 'posterior odds' genoemd, en het verwarren van dit quotiënt met het eerdere quotiënt wordt wel de 'prosecutor's fallacy' genoemd. Een eenvoudige toepassing van de regel van Bayes vertelt ons nu dat

posterior odds = likelihood ratio x prior odds,

waarbij de prior odds gedefinieerd worden als de posterior odds, maar dan vóór het zien van het bewijs. De likelihood ratio verzorgt dus in feite de 'update' van de odds die ontstaat door de DNA-match.

Als er redelijkheid in de wereld is, dan zouden de posterior odds niet mogen afhangen van de specifieke keuze van de hypotheses die je beschouwt, en enig rekenwerk laat zien dat dit inderdaad het geval is: de posterior odds in het database-scenario zijn gelijk aan $1/p(N-n)$, welke van bovenstaande hypotheses je ook kiest. Dit betekent dat de kans dat de match inderdaad de donor van het spoor is, toeneemt naarmate n groter wordt. Het lijkt er dus op dat kamp 2 toch gelijk heeft. Wat er aan de hand is, is dat een andere hypothese weliswaar een hogere bewijskracht kan geven, maar dat dit gecompenseerd wordt door *lagere* prior odds. Als je je dus concentreert op de posterior odds, en niet op de likelihood ratio zelf, dan is een data-afhankelijke hypothese toegestaan: de data zal een hoge bewijskracht geven, maar dat wordt teniet gedaan door lage prior odds. Die lage prior is de prijs die je betaalt voor de zeer specifieke hypothese die je kiest.

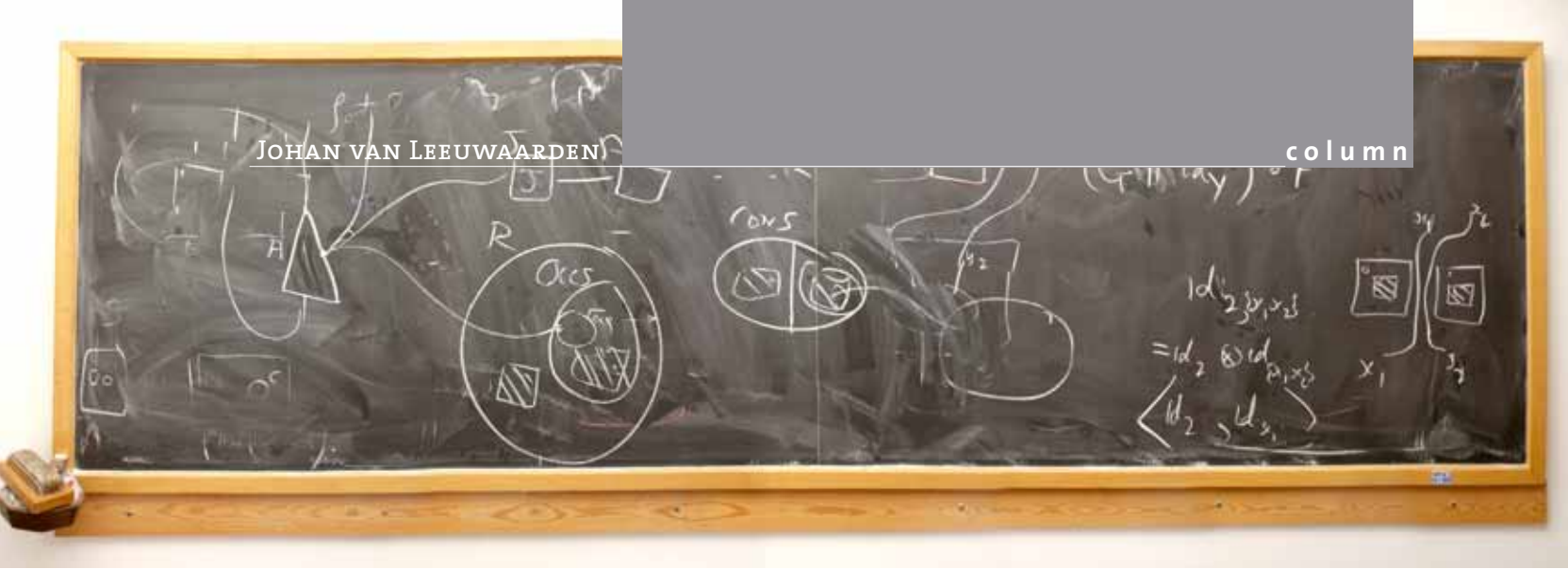
Maakt het dan helemaal niet uit welke hypotheses je kiest? Ja toch wel, maar de reden hiervoor is niet zozeer wiskundig, maar meer juridisch van aard. In de vergelijking

posterior odds = likelihood ratio x prior odds,

kan de forensisch deskundige alleen de likelihood leveren, de prior odds is het domein van de rechter. Dat betekent dat de uiteindelijke posterior odds gezien moeten worden als een gezamenlijke onderneming van rechterlijke macht en deskundige. Er zijn echter situaties, gecompliceerder dan het hierboven beschreven geval, waarin de likelihood ratio zelf afhangt van de prior. Dat is ongewenst, al is het alleen maar omdat het in sommige rechtssystemen niet toegestaan is. In dergelijke gevallen is het dus verstandig – of nodig – om een specifieke keus te maken voor de hypotheses die je wilt gaan toetsen.

In de praktijk zijn de problemen natuurlijk nog veel groter. De prior odds zijn moeilijk formeel te interpreteren; eigenlijk is alleen een subjectieve interpretatie mogelijk. Verder is het in sommige systemen niet toegestaan om de regel van Bayes te gebruiken, of erger nog, is het niet toegestaan om de jury of rechter uitleg te geven over de werking van de bijbehorende kansrekening. In de VS mag de jury niet weten hoe de verdachte geselecteerd is, en het is evident dat dit de relevante kansrekening de facto onmogelijk maakt.

Ronald Meester is hoogleraar Waarschijnlijkheidsrekening aan de Vrije Universiteit Amsterdam.
E-mail: <rmeester@few.vu.nl>
Homepage: <www.few.vu.nl/~rmeester>



STOFFIG ONDERWIJS

Krijtborden hangen sinds het begin van de negentiende eeuw in klaslokalen, maar worden nu ernstig bedreigd. Op grote schaal wordt het krijtbord vervangen door een whiteboard of zelfs een smartboard. Zo zouden deze borden schoner zijn. Maar mensen die whiteboards schoon vinden moeten dezelfde mensen zijn die WC-verfrisser verwarren met schone lucht. Of hun oksels voorzien van verse deodorant in een volle coupé. Mensen die liever viltstiften dan stoepkrijten. Ik daag u uit om twee uur te doceren voor een whiteboard. Je waant je in een chemische fabriek.

Voor wiskundedocenten is het extra pijnlijk. Steeds vaker worden ze verbannen naar klinische zaaltjes met whiteboards. Ze zijn wiskunde gaan studeren om het piepende geluid van het krijtje in de reuzenpasser. Vreselijk vinden ze het, die viltstiften. En stoffig? Daar genoten ze juist van. Het breken van het krijt. De vieze handen na het uitvegen, de pantalon met vlekken. Krijtborden zijn allerminst stoffig, en wiskundedocenten ook niet trouwens.

Maar mijn pleidooi voor het krijtbord strekt verder dan nostalgie of eigenbelang. Een krijtbord valt niet te vervangen. Zo doet de weerstand van het krijtbord mensen beter schrijven. White- en smartboards zijn te glad. Krijt is een heerlijk materiaal. Even toegankelijk als badmintonnen. Als kind ervaar je dit al bij het bekalken van de stoep. Het gevoel een artiest te zijn. Dat gevoel blijft. Ook lesgeven is theater, en het krijtbord is voor dit toneel een onmisbaar rekwisiet. Ingetogen teksten worden afgewisseld met vluchtige tekeningen. Wat is er mooier dan een docent die je af en toe de rug toe keert? Tijd inlast voor overpeinzing en twijfel. Het verhaal gestaag laat ontstaan.

En staat het bord eenmaal vol, dan is het heerlijk vergankelijk, als een bos bloemen of een zandkasteel. Krijt is eenvoudig met water te verwijderen. Chemicaliën komen er niet aan te pas. Wanneer ik de natte spons uitspoel en het gipsachtige water in de goot zie verdwijnen, dan geeft me dat een voldaan gevoel. Klus geklaard. Uit stof zijt gij geboren en tot stof zult gij wederkeren.

Dit pleidooi komt twintig jaar te laat, dat begrijp ik. En ik begrijp ook dat het smartboard, gesteund door voorspelbare Powerpoint-slides, het krijtbord de definitieve doodsteek gaat toebrengen. Maar laat het gezegd zijn.

Johan van Leeuwaarden is werkzaam in de groep Stochastische Besliskunde bij de faculteit Wiskunde en Informatica van de Technische Universiteit Eindhoven. Tevens is hij research fellow bij EURANDOM.
E-mail: <j.s.h.v.leeuwaarden@tue.nl>