

STATATOR

periodiek van de VWS jaargang 11 nummer 1, maart 2010

Nationale veiligheid en de methode
van nationale risicobeoordeling

Hoeveel reizigers zitten er in de trein?
Wedstrijd NS groot succes

Virussen te lijf met regressievergelijkingen

Wanneer is er genoeg 'bewijs'?

Het European Social Survey, een infra-
structuur voor de sociale wetenschappen

STATOR

Jaargang 11, nummer 1, maart 2010

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

Goos Kant (hoofdredacteur), Ana Isabel Barros, Mirjam Moerbeek, Gerrit Stemerding (eindredacteur), Hilde Tobi, Marnix Zoutenbier. Vaste medewerkers: Johan van Leeuwen, Fred Steutel

Kopij en reacties richten aan

Prof. dr. G. Kant (hoofdredacteur), Faculteit der Economische Wetenschappen van de Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg, telefoon 013 - 4668234, mobiel 06-11045089, <G.Kant@uvt.nl>.

Bestuur van de VVS

Voorzitter: prof. dr. R. Gill <voorzitter@vvs-or.nl>
Secretaris: dr. C.G.H. Diks <c.g.h.diks@uva.nl>
Penningmeester: prof. dr. ir. C.A.G.M. van Montfort <kvmontfort@feweb.vu.nl>
Statistische dag: prof. dr. J.J. Meulman <jmeulman@math.leidenuniv.nl>
Namens de Bedrijfssectie (BDS):
prof. dr. R.J.M.M. Does <R.J.M.M.Does@uva.nl>
Namens de Biometrische Sectie (BMS):
prof. dr. A.H. Zwinderman <a.h.zwinderman@amc.uva.nl>
Namens de Economische Sectie (ECS):
dr. P.H.F.M. van Casteren <casteren@fee.uva.nl>
Namens het Ned. Genootschap voor Besliskunde (NGB):
prof. dr. ir. C.P.M. van Hoesel <s.vanhoesel@ke.unimaas.nl>
Namens de Sectie Mathematische Statistiek (SMS):
dr. P.J.C. Spreij <spreij@science.uva.nl>
Namens de Sociaal Wetenschappelijke Sectie (SWS):
prof. dr. J.K. Vermunt <j.k.vermunt@uvt.nl>

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

Marieke Klein, p/a Vrije Universiteit, afdeling Econometrie & Operationele Research, De Boelelaan 1105, 1085 HV Amsterdam, <adverteren.stator@vvs-or.nl>. STATOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos | M. van Hootegem, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

Inhoud

- 3** Redactioneel
- 4** Nationale veiligheid en de methode van nationale risicobeoordeling
Diederik J.D. Wijnmalen
- 8** HOEVEEL REIZIGERS ZITTEN ER IN DE TREIN?
Wedstrijd NS groot succes! — **Jan Hoogenraad, Ramon Lentink & Bert de Vries**
Aanpak LIME — **Janne Brok, Jeroen Buskens & Martijn Slob**
Aanpak VU — **Aart de Vos, & Kai Ming Lee**
Aanpak CQM — **Paul Banens, Monique van den Broek & Marnix Zoutenbier**
- 11** Virussen te lijf met regressievergelijkingen
Nan van Geloven
- 16** Nieuwe master: Statistical Science for the Life and Behavioural Sciences
- 17** Wanneer is er genoeg 'bewijs'?
Ingeborg van der Tweel
- 22** IM Cora Maas
- 23** Er bestaan allochtone insluipsels in het Engels – *column*
Fred Steutel
- 24** Het European Social Survey, een infrastructuur voor de sociale wetenschappen
Ineke Stoop
- 30** IM Gijs de Leve
- 32** Tentamenweek – *column*
Johan van Leeuwen



Winnen stond centraal in de afgelopen weken. Bijvoorbeeld in Vancouver, waar de aanwijzingen van de coach en het blinde geloof van de topsporter in zijn coach dat winnen in de weg stonden. Gewonnen wordt er in deze *STATOR* door een team van de VU Amsterdam. In een door de NS georganiseerde wedstrijd wisten zij het beste, beter dan CQM en LIME, te voorspellen hoeveel reizigers er in treinen zouden zitten. En ook de 'atleten' die in de tentamenzaal van Johan van Leeuwen een voldoende weten te halen kunnen we als winnaars betitelen.

De serie 24 (*Twenty-Four*) is onlangs weer begonnen op TV. In 24 uur moet Jack Bauer voor de zevende keer in zijn leven een terroristische aanslag in Amerika voorkomen. Bauer doet dat voornamelijk met geweld en intimidatie waarbij met grote regelmaat zeer lastige afwegingen moeten worden gemaakt. Moet je als president van Amerika bijvoorbeeld je partner laten vermoorden als dat ertoe leidt dat duizenden andere Amerikanen daarmee gespaard worden? Kwantitatieve methoden worden daarbij niet gebruikt. In deze *STATOR* gebeurt dat wel. Diederik Wellenberg beschrijft een methode om zeer uiteenlopende risico's te vergelijken. Dit helpt de overheid om prioriteiten te stellen.

Virussen hebben maanden het nieuws gedomineerd. Met name de Mexicaanse griep zorgde voor heftige debatten over het nut van inenten en de mate waarin de Gezondheidsraad beïnvloed zou zijn door de industrie. Ook donorbloed kan virussen bevatten. Nan van der Geloven heeft in 2008 de VVS-scriptieprijs gewonnen voor haar onderzoek naar het optimaliseren van processen om die virussen te verwijderen. Daarbij maakte ze gebruik van statistische methoden.

'Wanneer is er genoeg bewijs' is de vraag die bij mij opkwam toen hertellingen het nieuws

domineerden na de gemeenteraadsverkiezingen begin maart. Gebruikelijk is om de laatste telling als waarheid te aanvaarden bij hertellingen. Het bewijs voor de correctheid van de laatste telling moet wel heel sterk zijn om op deze manier te beslissen. Ingeborg van der Tweel gaat in op de vraag 'wanneer is er genoeg bewijs' in het geval wel alle metingen meetellen. Dan wordt vooral de vraag hoeveel er gemeten moet worden en hoe je het aantal metingen kunt reduceren zonder de kwaliteit van de informatie geweld aan te doen.

Direct na de gemeenteraadsverkiezingen werd in de kranten veel gepubliceerd over de winst van de PVV en de scenario's die dat tot gevolg zou kunnen hebben als dit ook landelijk gaat gebeuren. In de kranten werden vergelijkingen gemaakt met de opkomst van andere populistische partijen in het buitenland en de strategieën die de partijen in andere Europese landen hebben gevolgd. Duidelijk werd dat vergelijken tussen Europese landen lastig is. Ineke Stoop beschrijft ontwikkelingen in Europa op het gebied van datacollectie. Het blijkt in algemene zin lastig om Europese landen met elkaar te vergelijken, echter, het lijkt erop dat de kwaliteit van landenvergelijkend onderzoek verbetert. Fred Steutel betoogt dat de buitenlandse inmenging in het wiskundige Engels zeer beperkt is, in tegenstelling tot het gebruik van Engelse woorden in andere landen. Zijn onderzoek leidt tot enkele Duitse, Engelse, en Arabische woorden maar daar blijft het dan ook bij.

Als laatste herdenken we Cora Maas en Gijs de Leve die in de afgelopen periode overleden zijn en beide veel betekend hebben voor de statistiek en de operations research.

Veel leesplezier

DE REDACTIE

Deeltjes van een griepvirus (bruin). Foto: National Geographic Society.

NATIONALE VEILIGHEID

EN DE METHODE VAN NATIONALE RISICOBEOORDELING

DIEDERIK J.D. WIJNMALEN

In mei 2007 lanceerde het kabinet bij monde van minister Ter Horst van Binnenlandse Zaken en Koninkrijksrelaties (BZK) de Strategie Nationale Veiligheid. Die Strategie beoogt om in een rijksbrede analyse de verschillende typen van groot-schalig risico op nationale schaal te definiëren, te beoordelen en te vergelijken. Op basis van een prioritering van die typen risico wordt nagegaan welke capaciteiten moeten worden onderzocht

en wellicht versterkt ter voorkoming van maatschappelijke ontwrichting. Het betreft capaciteiten in de sfeer van voorbereiding, preventie en bestrijding. Dit is een interessant maatschappelijk en beleidsanalytisch probleem waarvan een belangrijk onderdeel met OR technieken is aangepakt. Daarbij is naar een evenwicht tussen wetenschappelijkheid, begrijpelijkheid en toepasbaarheid gestreefd.

Vitale belangen en risicoscenario's

Voordat over beoordeling van risico kan worden gesproken, moet eerst worden gedefinieerd wat er precies wordt bedreigd. Daartoe is een vijftal nationale, vitale belangen gedefinieerd: de territoriale veiligheid (waarbij het grondgebied of het aanzien van Nederland wordt bedreigd), de fysieke veiligheid (waarbij leven en gezondheid en primaire levensbehoeften van de Nederlander op het spel staan), de economische veiligheid (waarbij bijvoorbeeld financiële markten of vermogensposities worden aangetast), de ecologische veiligheid (waarbij natuur en milieu worden bedreigd) en ten slotte de politieke en sociale stabiliteit in ons land (waarbij de instituties van onze democratie en/of zijn normen en waarden worden bedreigd en/of het vertrouwen van de bevolking in de overheid ernstig wordt aangetast met uitingen van angst en woede). Het kabinet heeft voor een *all hazard* benadering gekozen: zowel natuurlijke als opzettelijk veroorzaakte bedreigingen worden beschouwd.

Elke jaarlijkse ronde start met de benoeming van een aantal, uit een 'horizonscan' afgeleide, thema's. Voorbeelden zijn: klimaatverandering, extremisme, energievoorziening, pandemieën. Op elk van deze risicothema's worden scenario's ontwikkeld en toegevoegd aan die van de vorige ronde. Zo'n scenario beschrijft in de vorm van een verhaal een incident en zijn gevolgen, met bijbehorend feitenmateriaal en overige informatie gebaseerd op kennis en inzichten.

De uitdagingen waarvoor men zich gesteld zag, waren dat het bij de bedreiging van de vijf nationale belangen in beginsel om zeer verschillende schadegevolgen kan gaan, dat het bepalen van de

waarschijnlijkheid van een incident vaak wordt bemoeilijkt door ontbrekende data (wegens nieuwe en/of extreme dreigingen, of zich langzaam ontwikkelende processen), en dat om deze redenen ook het prioriteren van incidentscenario's een moeilijke zaak is. Voor deze opgave is een innovatieve Nationale Risicobeoordelingsmethode (NRB) ontwikkeld die op de beslissingsbenadering van de Multicriteria Analyse (MCA) is gebaseerd. Daarbij wordt het begrip 'risico' opgevat als een resultante van de twee componenten 'impact' en 'waarschijnlijkheid'. Elk scenario wordt met behulp van de NRB methode beoordeeld. Dit levert een positionering op zowel impact als waarschijnlijkheid op in een risicodiagram. Daarmee is een vergelijking van de verschillende incidentscenario's mogelijk gemaakt als basis voor een prioritering.

De stappen van de Nationale Risicobeoordelingsmethode (NRB)

Om de gevolgen van een incident te beoordelen, zijn tien *impact criteria* gedefinieerd in de context van de vijf nationale belangen. Elk scenario wordt op elk impactcriterium beoordeeld.

De beoordeling gaat allereerst met behulp van een of meer criteriumspecifieke indicatoren. Milieuvervuiling wordt bijvoorbeeld gemeten in termen van oppervlakte en tijdsduur, terwijl dodelijke slachtoffers in termen van het aantal doden wordt gemeten. Voor minder tastbare gevolgen zijn factoren gedefinieerd die moeten worden afgevinkt. Al deze scores op verschillende meetschalen worden getransformeerd naar een uniforme klassenschaal: van A (beperkt gevolg)

tot E (catastrofaal gevolg). Daarbij is een vertaling van de meetscores naar deze labels gemaakt; bijvoorbeeld 0-10 doden wordt klasse A, terwijl 10.000 doden of meer tot klasse E behoort.

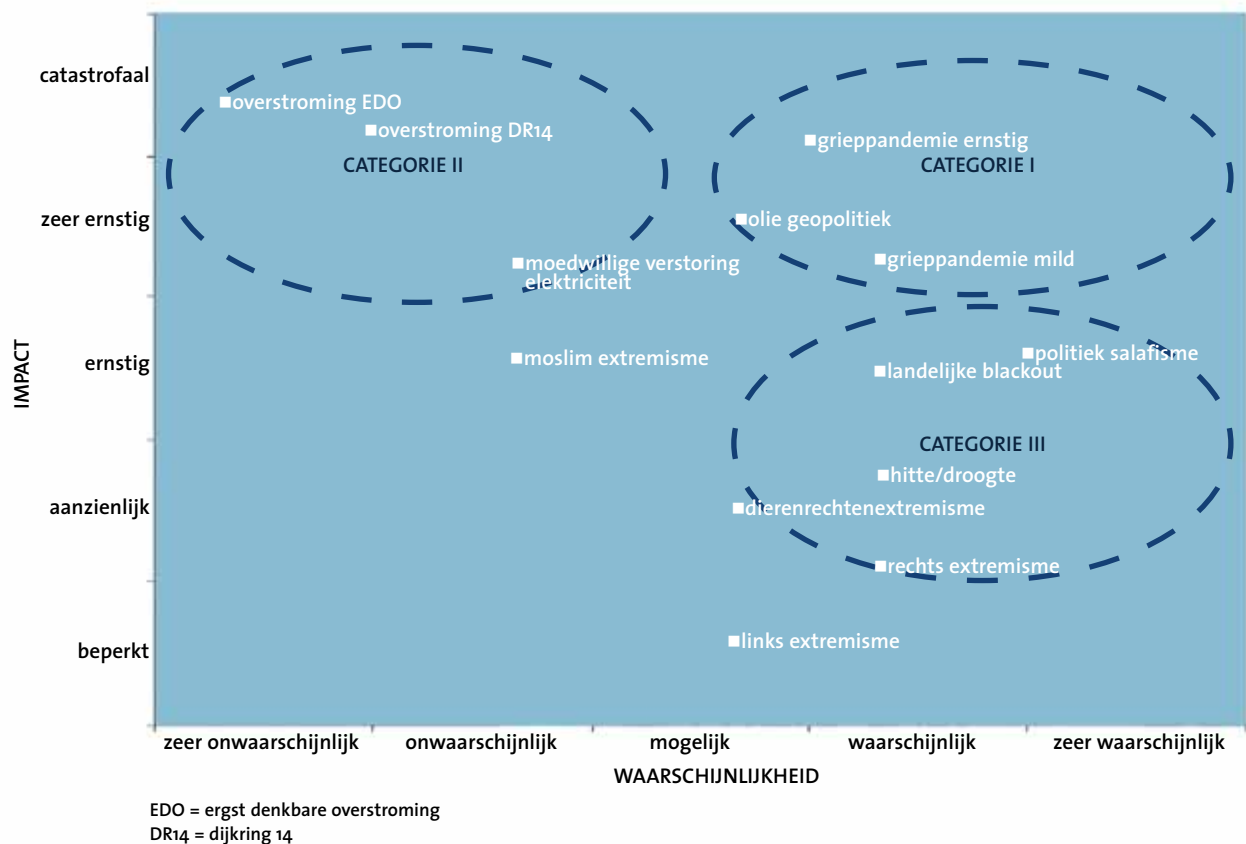
Om een eindscore voor 'gevolg' te berekenen zijn nog drie acties nodig:

- Een weging van de tien impactcriteria, waarvoor een vijftal weegprofielen is gedefinieerd: één met gelijke gewichten en vier profielen die verschillende sociaal-culturele preferenties van Nederlanders representeren. Een gewichtswaarde geeft aan hoe zwaar de score op een criterium meetelt in de eindscore, en in wezen hoe een label op het ene criterium zich verhoudt tot

hetzelfde label op een ander criterium.

- Een kwantificering van de labels A-E bij gebruik van een kwantitatieve aggregatiemethode. Voor de kwantificering worden lineaire en exponentiële functies gebruikt.
- Aggregatie met de gewogen sommethode. Deze is bekend en goed uit te leggen (belangrijke overweging!), maar vereist wel een kwantificering van scores en gewichten zoals al aangegeven.

De berekende eindscore op 'gevolg' wordt genuanceerd door ook onder- en bovengrenzen te berekenen, waarmee enige indicatie van de onzekerheid rond de scoretoekenning zichtbaar wordt



Figuur 1. Risicodiagram. Uit: Nationale Risicobeoordeling Bevindingenrapportage 2008 (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, Programma Nationale Veiligheid, 2008).

gemaakt. Dit is, net als het gebruik van de verschillende weegprofielen en de methodes voor labelkwantificering, onderdeel van een gevoeligheidsanalyse.

De waarschijnlijkheid van een incidentscenario wordt voor de eerstkomende vijf jaar vastgesteld, en eventueel voor een verder in de toekomst gelegen periode van vijf jaar. Er zijn twee redeneerlijnen onderscheiden: een kwalitatieve voor opzettelijk veroorzaakte incidenten waarvoor meestal geen statistische gegevens voorhanden zijn, en een overwegend kwantitatieve voor de overige die leidt tot een met een factor 10 oplopende exponentiële schaal. Het begrip 'kwetsbaarheid' in relatie tot reeds getroffen maatregelen speelt eveneens een rol. Elke redeneerlijn leidt echter tot een waarschijnlijkheidsklasse op de schaal van A (zeer onwaarschijnlijk, of 0-0,05%) tot E (zeer waarschijnlijk, of 50-100%). Evenals de labelschaal voor impact is hij tamelijk ongevoelig voor kleine nuances in de basisinformatie. De rol van experts die hun expertise moeten inbrengen, is bij ontbrekend statistisch materiaal echter cruciaal.

Interpretatie en gebruik van het risicogram

Figuur 1 toont de resultaten van de eerste scenarioverzameling van 2008. Deze scenarioverzameling is inmiddels tot 33 uitgebreid met andere risicotypen, of met dezelfde typen maar dan op andere schaal of omstandigheden.

De onderzochte scenario's kunnen worden geordend in (ovaal of rechthoekig getekende) prioriteitsgroepen op basis van combinaties van hoge en lage waarschijnlijkheid enerzijds en grote en geringe gevolgen anderzijds (op logaritmische as-indelingen!). Uit de nummering van de categorieën II en III blijkt dat aan impact impliciet een hoger gewicht wordt gegeven dan aan waar-

schijnlijkheid. Bij de hoog geprioriteerde scenario's wordt geanalyseerd welke specifieke capaciteiten kunnen bijdragen aan reductie van hoge labelwaarden (op impact en/of waarschijnlijkheid) en hoe het daar nu mee gesteld is. Sommige aldus geïdentificeerde capaciteiten zijn generieker van aard en voor meer risicotypen van belang. De uitkomsten van deze capaciteitanalyse leiden tot beleidsaanbevelingen die, net als de scenariokeuze en prioritering, worden goedgekeurd door de ministerraad.

Slotopmerkingen

De NRB heeft internationaal de aandacht getrokken en is zelfs, samen met de vergelijkbare Engelse aanpak van *national risk assessment*, door de Organisation for Economic Development and Cooperation (OECD) in 2009 internationaal als voorbeeld van een *best practice risk analysis* voorgesteld.

Dat deskundigen een- en andermaal pleiten voor meer geld om het risico in hun eigen domein aan te pakken, is begrijpelijk. Ze baseren dat op een onderbouwing vanuit hun eigen expertisegebied. De overheid heeft nu echter een instrument in handen waarmee verschillende typen risico met elkaar kunnen worden vergeleken, in onderlinge samenhang gezien en op basis van 'kans en gevolg', aangevuld met andere beleidsoverwegingen, naar prioriteit op de agenda kunnen worden gezet. Daarmee worden dankzij OR onderbouwd afgewogen (politieke) keuzes mogelijk.

DIEDERIK J.D. WIJNMALEN is als senior adviseur/onderzoeker werkzaam bij TNO Defensie en Veiligheid. TNO Defensie en Veiligheid levert vernieuwende oplossingen om de algehele veiligheid van de samenleving te bevorderen en is strategisch partner van het Ministerie van Defensie.
E-mail: <diederik.wijnmalen@tno.nl>



HOEVEEL REIZIGERS ZITTEN ER IN DE TREIN?

Wedstrijd NS groot succes!

JAN HOOGENRAAD, RAMON LENTINK & BERT DE VRIES

Op 27 oktober heeft bij de Nederlandse Spoorwegen in Utrecht een symposium plaatsgevonden met als titel 'Hoeveel reizigers zitten er in de trein?'. De aanleiding was een wedstrijd die NS had georganiseerd tussen 3 wiskundige teams. In deze wedstrijd hebben de teams in korte tijd (2 maanden) zo goed mogelijk geschat hoeveel reizigers er per traject en tijdstip in treinen in de Kop van Noord-Holland hebben gezeten. Het ging om treinen die in 2008 hebben gereden. De wedstrijd is gewonnen door econometristen van de Vrije Universiteit Amsterdam.

NS beschikt over een aantal bronnen met gegevens over treinreizigers, maar geen van deze bronnen levert voor iedere trein het exacte aantal. NS wil graag goede voorspellingen gebruiken voor het bepalen van de optimale lengte voor iedere trein, en om de potentie van extra acties voor de verkoop van lege stoelen in te schatten. Als NS de beschikking heeft over goede modellen om

te kunnen voorspellen hoeveel reizigers er in iedere trein zitten wordt hiermee de kans op een zitplaats voor de reiziger vergroot terwijl ook het materieel efficiënter wordt ingezet.

Als eerste stap heeft NS gegevens uit allerlei bronnen die zij zelf verzamelt samengebracht met informatie over weer, vakanties, evenementen enzovoorts. NS heeft zelf gegevens over de dienstregeling, verschillende methoden om reizigers te schatten, als ook van uitgevallen treinen en vertragingen. Sommige van deze bronnen zijn te gebruiken om voorspellingen te doen. Andere gegevens, zoals informatie over vertragingen, zijn alleen achteraf bekend. Voor deze wedstrijd mochten de teams alle gegevens gebruiken die zij relevant achtten om tot goede interpolaties te komen, dus ook informatie die achteraf pas bekend is.

Vervolgens is aan een aantal partijen gevraagd om mee te doen aan een wedstrijd om aan te tonen hoe goed een door NS achtergehouden testset van meetgegevens geschat kan worden. Hiervoor is zijn gegevens van telploegen, die voor een klein aantal treinen beschikbaar zijn maar in dat geval wel het meest betrouwbaar zijn, als referentie gebruikt. Dit geeft NS al een indicatie

Aanpak LIME

JANNE BROK, JEROEN BUSKENS & MARTIJN SLOB

LIME (Laboratory for Industrial Mathematics Eindhoven, Technische Universiteit Eindhoven) is een afdeling binnen de faculteit Wiskunde en Informatica die zich uitsluitend bezighoudt met het uitvoeren van projecten voor het bedrijfsleven. LIME telt op dit moment ongeveer 10 medewerkers, en is organisatorisch onderverdeeld in de groepen Computational Engineering en Data Analysis.

Wij hebben een model ontwikkeld dat de reizigersbehoefte bepaalt op basis van de tellingen en de bijbehorende dienstregeling. We hebben ervoor gekozen om het overstapgedrag van reizigers expliciet te modelleren. Dit is gebeurd met strategische informatie over reizigersgedrag die door NS al gevalideerd was. De LIME aanpak resulteert

hierdoor in consistentere schattingen van dezelfde trein omdat 'zijn verleden' wordt meegenomen.

Aangezien de dienstregeling expliciet in ons model gebruikt wordt, is ons model ook in staat om voor een nieuwe dienstregeling een voorspelling te doen, gegeven de reizigersbehoefte.

Omdat we alleen over verkoopgegevens van losse kaartjes beschikken hebben we aannamen moeten maken over het reisgedrag van abonneementhouders. Toevoegen van gegevens over hen zou waarschijnlijk tot verbeteringen leiden. De introductie van de OV-chipkaart zou bijvoorbeeld dergelijke informatie kunnen opleveren.

Een andere mogelijkheid om ons model te verbeteren zit in het meenemen van de capaciteit van de treinen. We nemen nu aan dat iedereen altijd in de trein past. In werkelijkheid klopt dit niet altijd.

We zouden in ons model ook dynamische informatie, bijvoorbeeld over uitgevallen of extra treinen, kunnen gebruiken om voorspellingen te verbeteren.

van hoe precies er voorspeld kan worden met de gegevens die NS nu reeds verzamelt. De kwaliteit van een voorspelling met een model zal niet beter zijn dan de kwaliteit van een interpolatie. Voor NS betekent dat dus dat de resultaten geïnterpreteerd kunnen worden als een ondergrens voor de voorspelkwaliteit.

De geselecteerde teams waren econometristen van de VU, CQM te Eindhoven, en LIME van de Technische Universiteit Eindhoven. Op het symposium presenteerden zij hun methoden: de VU en CQM hadden uiteindelijk loglineaire regressies gemaakt, LIME had gekozen om het instappen, overstappen, en uitstappen van reizigers te beschrijven op basis van de gereden dienstregeling. Aan het einde van het symposium werden de resultaten bekend gemaakt. Het model van de VU kwam het dichtst bij de weggelaten waarden. De spreiding van de resultaten bleek voldoende

klein, ten opzichte van de lengte van de kortste trein, om voor NS bruikbaar te zijn. NS heeft daarnaast ook geleerd welke van de door haar verzamelde bronnen het meest waardevol zijn voor voorspellingen.

Het symposium werd gehouden in het Trefpunt van de Nederlandse Spoorwegen, een mooie locatie waar levensechte modellen staan van alles wat je op en rond het spoor tegenkomt. De helft van de aanwezigen kwam uit de spoorsector en de andere helft bestond uit wiskundigen (met wat overlap tussen beide). De verschillende inzichten werden in de pauzes uitgewisseld, waarbij juist de mix van mensen uit de spoorsector en wiskundigen tot interessante discussies leidde.

Kortom, voor zowel NS als de deelnemende teams als de deelnemers aan het symposium een interessante ochtend in een aangename en inspirerende omgeving.

Aanpak VU

AART DE VOS & KAI MING LEE

Wij hebben (uiteindelijk) ervoor gekozen om de logaritme van de tellingen te modelleren met lineaire regressie : eerstegraads *splines* per dag, stationseffecten, wat speciale effecten en maand-effecten,

Onze ambities waren veel groter. Ons doel was eigenlijk om een model voor 'reizen' te ontwikkelen maar de beschikbare gegevens waren hier niet rijk genoeg voor. Zelfs een model voor in- en uitstappen per station (een mooi doel omdat bij de treintellingen steeds maar een of twee wagons worden geteld tussen de stations) wilde niet goed lukken.

Het gebruikte *fixed effects model* zou je ver-

der willen uitbreiden met een verklarend model voor de diverse patronen. Dan krijg je een *random effect model*, of hiërarchisch model zoals Bayesianen dat noemen. Maar interacties tussen treinen zijn complex, en een echt goed model maken is een opgave die meer tijd vergt. En alleen een heel goed model maakt uit als je veel data hebt (30.000 tellingen). Het is pas nodig als je de effecten van veranderingen in dienstregelingen wilt voorspellen (die waren er in dit geval niet).

De bruikbaarheid van simpele *splines* heeft ons verbaasd, we hadden Kalman Filters willen gebruiken (die je kunt zien als een model achter *splines*).

Kortom: het is mooi dat je met een relatief simpel model als nu gebruikt toch redelijk robuuste en bruikbare resultaten kunt behalen, maar het resultaat is nog niet iets waar we trots op zijn.

Aanpak CQM

PAUL BANENS, MONIQUE VAN DEN BROEK & MARNIX ZOUTENBIER

CQM heeft ervoor gekozen om de logaritme van de tellingen te modelleren met een lineair regressie model. Hierin hebben we aparte modellen geschat voor 'kaartjes' en 'abonnementen' omdat blijkt dat de tellingen van deze twee kaartsoorten verschillende patronen doorlopen in de tijd. Zo is bijvoorbeeld het effect van vakantie verschillend.

De uiteindelijke factoren in ons model hebben we geselecteerd door iteratief naar de patronen in de gegevens te kijken, een model daarop te baseren en dat te schatten en te toetsen, en

de patronen in de residuen te interpreteren, het model aan te passen etc.

De variatie in reizigersaantallen is groot. Zo is de variatie in reizigersaantallen in één trein op één baanvak binnen een dag al bijzonder groot, laat staan tussen de verschillende dagen in de week, over de verschillende perioden in het jaar, en dan ook nog voor verschillende treinen en verschillende baanvakken. Dit maakte de keuze voor de verschillende factoren en vooral de combinaties van factoren (interacties) tot een uitdagend vraagstuk.

Naast onze schattingen hebben we ook andere aanbevelingen gedaan om de voorspelkwaliteit te verbeteren. Wij denken bijvoorbeeld dat een substantiële verbetering van de resultaten bereikt kan worden door de telmethode aan te passen.

VIRUSSEN TE LIJF MET REGRESSIEVERGELIJKINGEN

NAN VAN GELOVEN

Als je een bloedtransfusie moet ondergaan, wil je natuurlijk dat het veilig gebeurt. Ondanks een uitgebreide screening en selectieprocedure van bloeddonoren, kan donorbloed virussen bevatten. Bij Sanquin (de bloedbank) wordt onderzoek gedaan naar methoden om virussen te verwijderen uit bloedproducten. Doordat het moeilijk is om virusaantallen te meten, komen daar dure proefopstellingen aan te pas. Een wiskundige modellering van deze proeven maakt beter onderzoek naar virussen mogelijk.

Virussen verwijderen

Donorbloed wordt niet alleen gebruikt voor directe transfusies. Bloed, en vooral het bloedplasma, wordt ook verwerkt in andere medische producten. Albumine bijvoorbeeld, een middel

gebruikt bij de behandeling van ernstige brandwonden, wordt gemaakt uit plasma. Dergelijke producten hebben in het verleden geleid tot de verspreiding van het hepatitis A, B en C virus en AIDS.

Tegenwoordig moeten bloedproducten stappen ondergaan die mogelijk aanwezige virussen verwijderen of inactiveren. Een voorbeeld van zo'n reductiestap is pasteurisatie: door plasma langere tijd op 60 °C te verwarmen, worden virussen uitgeschakeld. Naast pasteurisatie kunnen ook filtratie-, centrifuge- of bijvoorbeeld belichtingsstappen ingezet worden tegen virussen. Sanquin doet onderzoek naar de optimale omstandigheden van dergelijke processen. Bij welke pH-waarde is de pasteurisatie het meest effectief? Welke afmetingen moeten de gaatjes van de (nano-)filtratieplaat hebben? Met hoeveel g-kracht stel ik een viruscentrifuge in?

Om dergelijke vragen te beantwoorden, test Sanquin de reductiestappen onder verschillende omstandigheden. In het laboratorium worden bloed- of plasmasamples opzettelijk besmet met virus. Na het uitvoeren van een reductiestap wordt gekeken hoeveel virus er nog aanwezig is. De hoeveelheid verwijderde virus, uitgedrukt in een log reductie factor (LRF), is een maat voor de effectiviteit van de desbetreffende stap.

Opzet experiment

In Tabel 1 staat een voorbeeld van een experiment zoals dat gebruikt is om de optimale omstandigheden van pasteurisatie te vinden. In dit experiment wordt er gezocht naar de invloed van temperatuur, pH en sacharose op het virusreducerend vermogen van de pasteurisatie. Vanuit een statistisch oogpunt kijkend naar de tabel, vallen er een aantal zaken op.

Ten eerste worden niet alle mogelijke combinaties van omstandigheden getest. Om het aantal (dure) testen te beperken, heeft Sanquin gekozen voor een zogeheten *fractional factorial design*. Hierbij wordt op een slimme manier een beperkt aantal combinaties van hoog en laag instellingen gekozen, terwijl toch het effect van iedere factor apart te onderzoeken is. De eerste vier pasteurisaties horen bij een 2^{3-1} ontwerp waarbij de effecten van drie procesfactoren in slechts vier combinaties getest worden. Pasteurisaties 5 en 6 zijn gedaan onder 'standaard' instellingen en dienen vooral om vertrouwen te krijgen in de herhaalbaarheid van het experiment. Het totale aantal van zes meetpunten is niet veel voor het beoordelen van het effect van drie procesfactoren.

Ten tweede zijn enkele van de LRF's gecensureerd. Doordat er na pasteurisatie 3 en 4 niets meer van de vooraf aanwezige virussen teruggevonden werd, is de reductie ervan bestempeld als minstens de beginhoeveelheid virus. Immers, als

Experiment	Temperatuur (°C)	pH (-)	Sacharose (%)	Virusreductie (LRF ± SE)
1	58,5	6,6	62	3,4 ± 0,2
2	58,5	7,4	58	4,0 ± 0,2
3	62,5	6,6	58	≥ 5,1
4	62,5	7,4	62	≥ 4,9
5	60,5	7,0	60	4,0 ± 0,2
6	60,5	7,0	60	4,0 ± 0,2

Tabel 1. Experiment om de optimale omstandigheden van een pasteurisatiestap te bepalen.

er met meer virus gestart was, dan had de pasteurisatie wellicht nog meer virus verwijderd.

Ten derde worden de LRF's gemeten met een bepaalde meetfout, te zien aan de vermelde *standard errors* (SE's). Virussen kunnen niet direct gemeten of geteld worden. De hoeveelheid virussen aanwezig voor en na de pasteurisaties wordt geschat met behulp van een verdunningsreeks. Het virusmateriaal wordt daarbij herhaaldelijk verdund en iedere verdunning wordt toegevoegd aan meerdere oplossingen van levende cellen. Nadat dit materiaal een aantal weken op kweek gezet is, wordt gescoord bij welke oplossingen het virus ook echt actief is geworden. Het aantal maal dat het virus actief wordt per verdunning, geeft een schatting van de hoeveelheid aanwezige virus, met bijbehorende meetfout.

De uitdaging is nu om met deze beperkte hoeveelheid, deels gecensureerde en met meetfout gerapporteerde LRF's een antwoord vinden op de vraag hoe de drie procesfactoren de virusreductie beïnvloeden.

Twéé modellen in één

Wie goed naar de bovenstaande experimentele opzet kijkt, ziet dat het twee fases bevat. Eerst wordt de virusreductie (LRF) per pasteurisatie geschat op basis van verdunningsreeksen en vervolgens willen we met deze LRF's een uitspraak doen over het effect van de procesfactoren. Samenvoegen van deze twee fases heeft als voordeel dat we de analyse kunnen uitvoeren op meer metingen omdat in iedere verdunningsreeks 12 verschillende concentraties in 8 celoplossingen getest zijn. Dit vergt echter wel een speciale manier van analyseren.

Het *most probable number* algoritme (Cochran 1950) waarmee aan de hand van een verdunningsreeks een hoeveelheid virus geschat wordt, kan uitgebreid worden om de methode ook geschikt te maken voor het schatten van de effecten van de procesfactoren. Simpelweg door in de (likelijkheid-)formule de 'hoeveelheid virus' te vervangen door een voorspelling op basis van

PROCESFACTOR	EEN-FASE MODEL		TWEE-FASE MODEL	
	EFFECT	P-WAARDE	EFFECT	P-WAARDE
Temperatuur	0,80	<0,0001	0,90	0,0069
pH	-0,00	0,8589	-0,10	0,7227
Saccharose	-0,27	<0,0001	-0,40	0,2441

Tabel 2. Resultaten van het één -fase model afgezet tegen een twee-fase model. Het een-fase model gebruikt direct de ruwe data van de verdunningsreeksen. Het twee-fase model is gefit op de tussentijdse LRF schattingen.

de gebruikte niveaus van temperatuur, pH en sacharose. Het schatten van de procesfactoren op basis van dit uitgebreide model heeft geen last van gecensureerde datapunten. Ook niet van tussentijdse meetpunten. Uitgaande van een vaste beginhoeveelheid virus in iedere run, heeft de methode niet 6 LRF's, maar $6 \times 12 \times 8 = 576$ onderliggende datapunten voor de regressie.

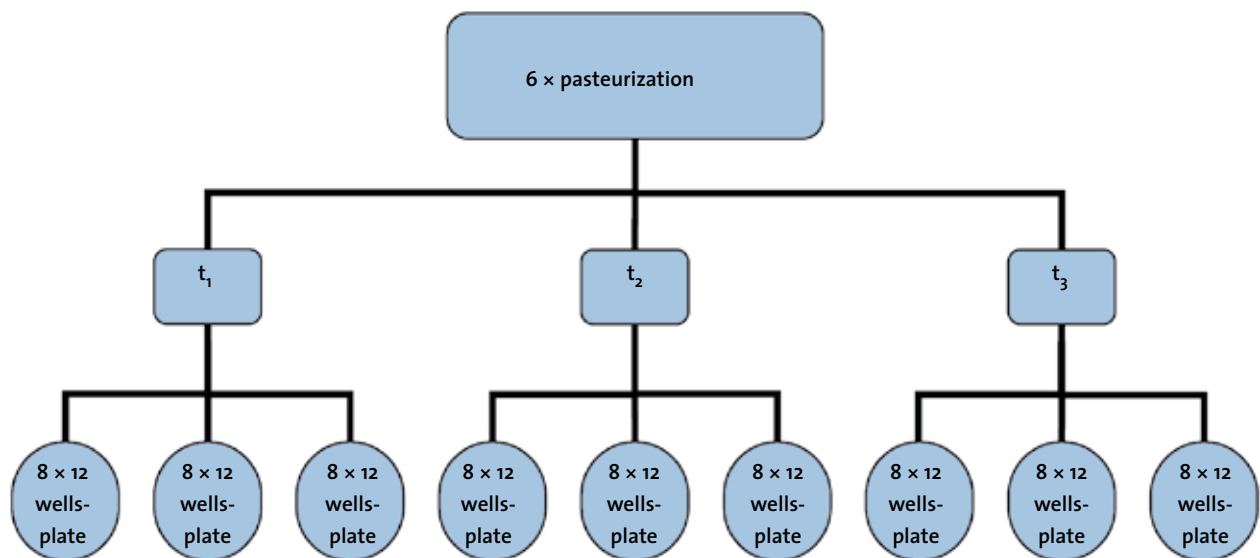
Klinkt niet slecht, maar krijgen we op deze manier een goed antwoord onze vraag? In Tabel 2 staan de effectschattingen die het model aan de procesfactoren geeft. Ernaast ter vergelijking een twee-fase benadering: een lineair regressiemodel op de 6 LRF's waarbij de censurering voor het gemak genegeerd is.

Variatie meenemen

In Tabel 2 zien we dat het één-fase model vergelijkbare effecten schat als een basic twee-fase model, maar met hogere significantie. De overeenkomst van de effectschattingen geeft ons vertrouwen in

het model. Het nieuwe model heeft duidelijk voordelen. Het herbergt op een natuurlijke manier de onzekerheid van de verdunningsreeksen in zich. Het hoeft zich niet druk te maken om SE's en gecensureerdheid. Maar hoe zit het met die significanties? Zou het model de informatie in de data niet wat overdrijven? Ieder datapunt uit een verdunningsreeks hoort namelijk niet bij een nieuwe onafhankelijke pasteurisatie. Het is goed mogelijk dat data uit dezelfde verdunningsreeks een bepaalde correlatie vertoont waar ons regressiemodel tot nu toe geen rekening mee houdt.

Om een idee te krijgen van de grootte van deze afhankelijkheden en bijbehorende variatie, is een nieuw validatie-experiment opgezet. Door een aantal herhalingen te doen van een pasteurisatie onder gelijke procesomstandigheden konden we de variatie behorend bij sec een pasteurisatie (en dus niet veroorzaakt door de procesfactoren) ontwaren. Bovendien waren vooral de laboranten erg benieuwd naar waar variatie aan toe te wijzen is. Wordt er netjes gepipetteerd, gedragen virussen zich ieder experiment vergelijkbaar?



Figuur 1: Opzet van het validatie experiment. Iedere sample wordt hierin drie maal getest in een verdunningsreeks.

In de validatie-proef werd onder gelijke procesomstandigheden zes maal een pasteurisatie uitgevoerd, waarbij op drie tijdstipmomenten een sample uit het verwarmde vat genomen werd. Om los naar de variatie van het pipetteren te kunnen kijken, werd bovendien de virushoeveelheid in een sample getest met behulp van drie verdunningsreeksen (Figuur 1).

Uit het feit dat ons één-fase regressie model deze 6 ‘identieke’ pasteurisaties niet als vergelijkbaar bestempelde, concludeerden we dat het model nog niet met alle aanwezige variatie rekening hield. Op twee plekken hebben we het model uitgebreid. Ten eerste hebben we de volumeon-nauwkeurigheden toegevoegd. Hiervoor konden we de specificaties gebruiken van de pipetteer-

materialen die in het lab gebruikt werden. Het uitgebreide model houdt er rekening mee dat er bij het pipetteren van een sample een kleine volumefwijking kan ontstaan en evenzo iedere keer dat het materiaal verdund wordt.

Het bleek dat het meenemen van deze technische variatie maar een klein deel van de totale variatie verklaarde. De overgebleven variatie, die veroorzaakt kan worden door het deels onvoorspelbare gedrag van de biologische materialen, maar ook door het niet precies op hetzelfde tijdstip nemen van een sample, hebben we meege-nomen als ‘biologische variatie’. We zijn er daarbij van uit gegaan dat er bij het nemen van iedere sample een kleine afwijking van het gemiddelde gedrag van virussen aanwezig is. Zodoende zal



Biological / time error:

$$t \rightarrow t + \delta$$

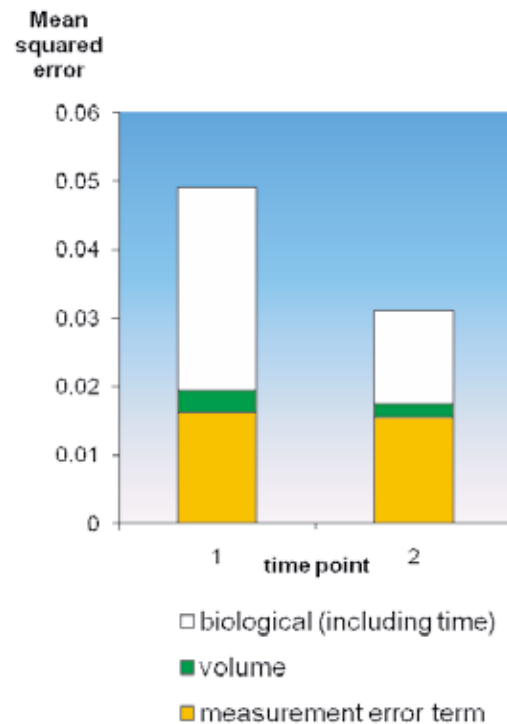


Volume error:

$$v_i \rightarrow (v_0 + \tau_{il}) \prod_{d=1}^i (f + \epsilon_d)^{-1}$$



Measurement error
(inherent to the dilution assays)



Figuur 2. Resultaten van het validatie experiment: de meetfout (behorend bij het schatten van de hoeveelheid virus uit een verdunningsreeks) en de biologische fout dragen het meest bij aan de totale variatie. De volumefouten behorend bij de pipetteerstappen hebben veel minder invloed.

ook het model er rekening mee houden dat alle verdunningsreeks data afkomstig uit deze sample een gezamenlijke afwijking heeft. Figuur 2 laat zien hoe groot de toegevoegde variaties zijn op het eerste en tweede tijdspunt.

Het uitgebreide model vindt de 6 identieke test pasteurisaties uit het validatie-experiment wel vergelijkbaar. Bij het beoordelen van het oorspronkelijke pasteurisatie experiment is het nu ook voorzichtiger in het aanwijzen van significante effecten (temperatuur $p < 0,0001$, Ph $p = 0,8982$, sacharose $p = 0,0035$). Uiteraard veranderen de effectgroottes in dit met *random errors* uitgebreide model niet.

Conclusie

Met het resulterende model kan het gedrag van virussen geanalyseerd worden rekening houdend met de aanwezige variatie. Tussentijdse onzekere en mogelijk gecensureerde metingen vormen daarbij geen obstakel. Hoewel biologische materialen zich lastig laten vangen, is het goed om de aanwezige variatie zo veel mogelijk mee te nemen in een analysemodel.

LITERATUUR

Cochran, W.G. (1950). Estimation of bacterial densities by means of the 'most probable number'. *Biometrics*, 5, pp. 105-116.

Geloven, N, van, Cator, E.A., Lopuhaä, H.P., & Janssen, M.P. (2009). Regression analysis on serial dilution data from virus validation robustness studies. *Statistica Neerlandica*, 63(3), pp. 245-257.

NAN VAN GELOVEN studeerde Technische Wiskunde aan de TU Delft en won voor haar afstudeerscriptie 'Design and analysis of virus validation robustness studies' in 2008 de VVS-scriptieprijs. Op dit moment werkt zij als klinisch statisticus bij het Academisch Medisch Centrum te Amsterdam.

E-mail: <n.vangeloven@amc.nl>.

NIEUWE MASTER

Statistical Science for the Life and Behavioural Sciences

De Universiteit Leiden heeft een nieuwe master-specialisatie toegepaste statistiek: Statistical Science for the Life and Behavioural Sciences. De tweejarige, interdisciplinaire specialisatie is in september 2009 van start gaat en het resultaat van samenwerking tussen het Mathematisch Instituut en de Faculteit der Sociale wetenschappen van de Universiteit Leiden, de afdeling Medische statistiek en bio-informatica van het LUMC, het departement Wiskunde van de Vrije Universiteit in Amsterdam, de afdeling Biostatistiek van het Erasmus MC en de afdeling Biometrie/Toegepaste Statistiek van de Landbouwwuniversiteit Wageningen. Binnen de specialisatie kunnen studenten kiezen tussen twee richtingen: *life sciences* en gedragswetenschappen. De opleiding mikt op Leidse studenten maar ook studenten van de andere deelnemende universiteiten.

Prof.dr. Richard Gill, werkzaam bij het Mathematisch Instituut, voorzitter van de VVS en een van de initiatiefnemers, zegt dat gestart is met een groepje van ongeveer twaalf studenten. Dat zijn niet alleen doorstromers vanuit de bachelor: 'Er is ook belangstelling vanuit de industrie en andere sectoren. Daar zitten onderzoekers, soms met een PhD op zak, die zich selfmade hebben ontwikkeld tot bijvoorbeeld biostatistici. Er zijn er die deze opleiding dolgraag willen volgen.' Gill zegt dat Nederlands, anders dan andere landen, statistiek helaas niet als apart wetenschapsgebied onderscheidt.

Er is veel vraag naar toegepaste statistici. Afgestudeerden kunnen in allerlei sectoren aan de slag: in academische ziekenhuizen, de industrie (de farmaceutische, landbouw- voedsel- en olie-industrie), onderzoeksinstituten, financiële instellingen, rekenbureaus van de overheid, onderwijskundige diensten en marketingbureaus.

Informatie: <www.math.leidenuniv.nl/statscience/>



WANNEER IS ER GENOEG ‘BEWIJS’?

INGEBORG VAN DER TWEEL

Het getuigt van *Good Statistical Practice* om bij de opzet van een wetenschappelijk onderzoek na te denken over de steekproefomvang (Moerbeek 2009). In het bijzonder in klinisch onderzoek is er een ethische noodzaak. Te weinig patiënten kan te weinig bewijskracht opleveren. Te veel patiënten is onethisch en kan betekenen dat een aantal van hen mogelijk minder goed behandeld wordt. Aan de hand van een literatuurvoorbeeld wil ik ingaan op de mogelijkheden van sequentiële analyse om ethisch en efficiënt (klinisch) wetenschappelijk onderzoek uit te voeren.

Onderzoekers wilden weten of orale corticosteroïden vroegtijdige verslechtering konden voorkomen bij patiënten met aids-gerelateerde longontsteking (Montaner *et al* 1990). Bij behandeling met een placebo was de kans op verslechtering gelijk aan 0,4. Men hoopte deze kans met orale corticosteroïden te kunnen verminderen tot 0,1. In een gerandomiseerd klinisch onderzoek wordt de schatting van het benodigde aantal patiënten gebaseerd op het relevant geachte verschil tussen de kans op verslechtering bij behandeling met placebo en die bij behandeling met orale cortico-

steroiden. Het geschatte aantal hangt verder af van de type I en type II fout van de statistische toets. De kans op een type I fout is de kans op een vals-positieve conclusie (men concludeert op basis van de gevonden resultaten dat er een verschil is, terwijl dat verschil er in werkelijkheid niet is); de kans op een type II fout is de kans op een vals-negatieve conclusie (men concludeert dat er geen verschil is, terwijl dat verschil er in werkelijkheid wel is).

'Vaste' versus variabele groepsgrootte

Voor genoemd onderzoek met veronderstelde fracties van 0,4 (placebo) en 0,1 (orale corticosteroiden), en kansen op een type I fout (éénzijdig) en een type II fout van 0,05 zijn in totaal ten minste 70 patiënten nodig, 35 in iedere groep. We noemen dit de 'vaste' groepsgrootte.

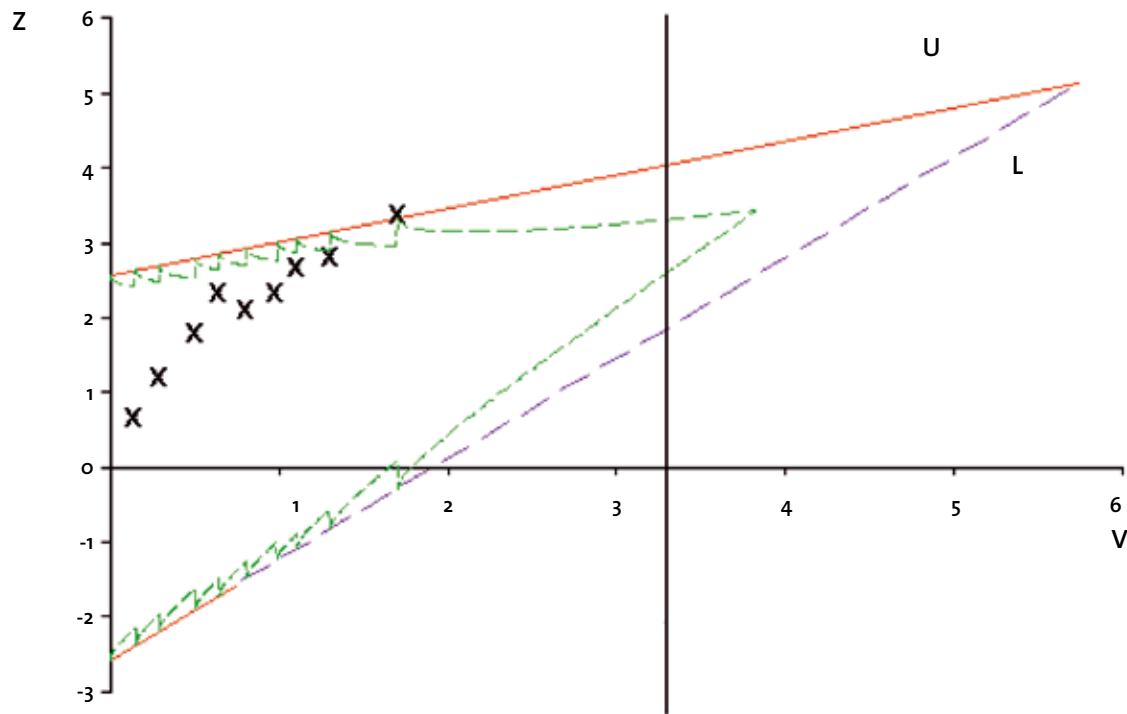
Deze patiënten worden echter niet allemaal tegelijk in het onderzoek betrokken, maar meestal in de loop van een tijdsperiode geïnccludeerd. Hun gegevens komen dan ook meestal sequentieel in de tijd beschikbaar. Het ligt voor de hand dat de statistische analyse van deze gegevens ook sequentieel wordt uitgevoerd. Wat nu als tijdens de sequentiële analyse blijkt dat het verschil in voortijdige verslechtering tussen de twee behandelingen (veel) groter is dan de verwachte 30%? Wanneer hebben we voldoende 'bewijs' om de studie te stoppen en alle patiënten te behandelen met orale corticosteroiden?

Als na iedere nieuwe patiënt of groepje patiënten een statistische analyse op de cumulatieve gegevens wordt uitgevoerd, loopt de kans op een type I fout al snel op. Bij 1 analyse is deze kans 0,05, bij 2 analyses 0,08, bij 5 analyses 0,14 en bij 10 analyses al 0,19. De kans dat we na 10 analyses op de cumulatieve data ten onrechte concluderen

dat er een verschil is tussen orale corticosteroiden en placebo is dan bijna 0,20. Om dit ongewenste effect te voorkomen kan men van te voren beslissingsgrenzen definiëren. Deze grenzen worden, net als bij de schatting van de groepsgrootte hierboven, bepaald door het relevante verschil in effect en de type I en type II fouten (zie Figuur 1). Voor deze analyse is het verschil in fractie verslechtering omgerekend naar een *odds ratio* en gelijk aan 0,167 ($= 0,1 \times 0,6 / (0,9 \times 0,4)$).

Iedere nieuwe groep patiënten waarvan de gegevens beschikbaar komen, levert een nieuw punt op in de grafiek. De V-coördinaat van dit punt is een functie van het cumulatieve aantal patiënten (n) in de analyse; de Z-coördinaat is het cumulatieve effect (een maat voor het verschil tussen placebo en orale corticosteroiden) in de beschikbare uitkomsten. Z is gelijk aan het verschil tussen het aantal *waargenomen* uitkomsten (vroegtijdige verslechtering) in de placebogroep en het aantal *verwachte* uitkomsten onder de nulhypothese (*observed - expected*). Voor grote n en gelijke verdeling van de patiënten over de twee armen van de studie is V ongeveer gelijk aan $n/4 \times p(1-p)$ met p de proportie patiënten met vroegtijdige verslechtering in de cumulatieve groep van n patiënten. Z^2/V is de bekende chi-kwadraat toetsingsgrootte.

Zo ontstaat een 'pad' van punten met een zekere trend. Als dit 'pad' de bovengrens U overschrijdt, dan kunnen we concluderen dat het verwachte verschil in fractie verslechtering minstens 0,3 is. We verwerpen de nulhypothese 'er is geen verschil' ten gunste van de alternatieve hypothese 'het verschil is minstens 0,3'. Overschrijdt het 'pad' van punten de ondergrens L, dan is de kans dat de nulhypothese met deze data ooit verworpen zal gaan worden heel klein geworden. De verticale lijn in Figuur 1 correspondeert met het 'vaste' aantal van 70 patiënten. De driehoeksvorm van de



Figuur 1. Cumulatieve analyse van de data uit de studie van Montaner et al (1990). Z is het verschil tussen het aantal waargenomen uitkomsten in de controlegroep en het aantal verwachte uitkomsten onder de nulhypothese; V is een functie van het cumulatieve aantal patiënten in de analyse (zie verder de tekst). Het 'pad' van X overschrijdt de bovengrens U, zodat de conclusie is dat orale corticosteroiden leiden tot significant minder vroegtijdige verslechtering dan placebo. De onderbroken, kerstboomachtige binnengrenzen vormen een continuïteitscorrectie.

test heeft tot gevolg dat het proces met zekerheid zal stoppen: uiteindelijk wordt één van de grenzen overschreden.

In bovengenoemd onderzoek werden na het optreden van iedere nieuwe uitkomst (vroegtijdige verslechtering) de cumulatieve data geanalyseerd. Nadat de resultaten van 37 patiënten met 9 uitkomsten beschikbaar waren kruiste het 'pad' van 9 (Z,V)-punten de bovengrens U. Op dat moment waren 18 patiënten behandeld met orale corticosteroiden, 1 daarvan vertoonde vroegtijdige verslechtering; van de 19 patiënten met een

placebobehandeling vertoonden 8 vroegtijdige verslechtering. De waarde voor Z was $8 - 9 \times 19/37 = 3,378$; de waarde voor V was $37/4 \times 9/37 \times 28/37 = 1,703$. Dit betekende een winst van 33 patiënten ten opzichte van de 70 benodigde patiënten voor deze studie, ongeveer 47%. De schatting voor de *odds ratio* is 0,15 met 90%-betrouwbaarheidsinterval (0,04; 0,57). Deze punt- en intervalschatters zijn gecorrigeerd voor het herhaald toetsen van de data. (NB Het 'pad' van punten was door mij niet te reconstrueren. Figuur 1 is een weergave van hoe de studie verlopen zou kunnen zijn. De

	MEDIANE AANTAL PATIËNTEN	GEMIDDELDE AANTAL PATIËNTEN	90-PERCENTIEL VAN HET AANTAL PATIËNTEN
nulhypothese: 'er is geen verschil'	28	31	50
alternatieve hypothese: 'het verschil is minstens 0,3'	36	39	64
tussenliggende waarde voor het verschil	46	44	71

Tabel 1. Geschatte aantal benodigde patiënten voor een sequentiële analyse van de data van Montaner et al (1990).

Z- en V-coördinaat bij het overschrijden van de bovengrens zijn uiteraard dezelfde als in de publicatie van Montaner et al, 1990.) Met een sequentiële analyse kan dus winst behaald worden in het aantal patiënten dat in een studie betrokken wordt.

Als twee behandelingen uiteindelijk een vergelijkbaar effect blijken te hebben op de uitkomsten van patiënten wordt de ondergrens L overschreden. Ook in die situatie kan winst behaald worden in het aantal in te sluiten patiënten.

Toepassingen van sequentiële analyse

Sequentiële analyse werd ontwikkeld in de Tweede Wereldoorlog en kort daarna beschreven voor industrieel onderzoek (Wald, 1947). In klinisch onderzoek wordt sequentiële analyse regelmatig toegepast en niet alleen voor het aantonen van superioriteit van een nieuwe behandeling. Sequentiële analyse kan ook worden ingezet om de veiligheid van een (nieuwe) behandeling te

monitoren, bijvoorbeeld het optreden van sterfte in de kindercirurgie (De Leval et al, 1994). Als deze sterfte een van te voren vastgestelde waarde overschrijdt, kan met behulp van monitoring tijdig worden gewaarschuwd. Hiermee wordt voorkomen dat pas aan het eind van een studie geconstateerd wordt dat de sterfte hoger was dan verwacht of dan acceptabel gevonden wordt.

In epidemiologisch onderzoek wordt soms gebruik gemaakt van biologisch materiaal (denk aan bloed-, urine- of weefselmonsters) om onderzoekshypothesen te toetsen. Voor een onderzoek naar een mogelijk verband tussen een seleniumtekort in het lichaam en het later optreden van kanker hadden epidemiologen de beschikking over teennagels, opgeslagen in een biologische bank, van vrouwen die ooit voor borstkanker in de regio Utrecht gescreend waren. Seleniumconcentraties in de teennagels van vrouwen die later kanker gekregen hadden werden vergeleken met seleniumconcentraties van controlevrouwen. Een monster kan meestal maar één keer voor een statistische toets gebruikt worden. De onderzoekers zijn

daarom genoodzaakt dergelijke monsters zuinig en efficiënt te gebruiken, met name de monsters van de kankerpatiënten. Sequentiële analyses kunnen in observationeel onderzoek op efficiënte wijze interessante onderzoekshypothesen scheiden van minder interessante (Kaaks et al, 1994; Van der Tweel et al, 1996).

Het zal duidelijk zijn dat ook wetenschappelijk onderzoek met proefdieren en met kinderen om ethische redenen zo efficiënt mogelijk uitgevoerd dient te worden (Russell et al, 1992; Van der Lee et al, 2008).

Benodigde aantallen

Een mogelijk nadeel van een sequentiële analyse is dat niet van tevoren het benodigde aantal patiënten vastgesteld kan worden. Wel kan het gemiddelde of mediane aantal geschat worden evenals bijvoorbeeld het 90-percentiel van het benodigde aantal.

In Tabel 1 staan de gemiddelden, medianen en 90-percentielen van het totale benodigde aantal patiënten voor bovengenoemd voorbeeld, aangenomen dat de nulhypothese waar is, de alternatieve hypothese waar is en voor een tussenliggende waarde voor het werkelijke verschil.

Gemiddeld genomen vraagt een sequentiële analyse minder in te sluiten patiënten dan een studie met een 'vaste' grootte. Er is echter geen sprake van een 'gratis lunch': soms vraagt een sequentiële analyse meer patiënten dan de 'vaste' groeps grootte om uiteindelijk tot een beslissing te komen. Dit kan optreden als het werkelijke verschil kleiner is dan geformuleerd onder de alternatieve hypothese. In de figuur overschrijdt het 'pad' van punten tussen de onder- en de bovengrens in dat geval de verticale lijn die met de 'vaste' groeps grootte correspondeert.

In dit artikel heb ik de sequentiële analyse zoals ontwikkeld door John Whitehead beschre-

ven (Whitehead, 1997). Er zijn meer manieren om een studie (groeps)sequentieel te analyseren. De methode van Whitehead is flexibel en omvat de meeste andere mogelijke analyses.

Samengevat kan gesteld worden dat sequentiële analyse om ethische of economische redenen een efficiënt alternatief kan zijn voor een studie met een 'vaste' grootte.

LITERATUUR

De Leval, M.R. et al (1994). Analysis of a cluster of surgical failures. Application to a series of neonatal arterial switch operations. *The Journal of Thoracic and Cardiovascular Surgery*, 107, pp. 914-924.

Kaaks, R. et al (1994). Efficient use of biological banks for biochemical epidemiology: exploratory hypothesis testing by means of a sequential *t*-test. *Epidemiology*, 5, pp. 429-438.

Moerbeek, M. (2009). Het waarom en hoe van statistische power analyses. *STATOR*, 3, pp. 17-21

Montaner, J.S.G. et al (1990). Corticosteroids prevent early deterioration in patients with moderately severe *Pneumocystis carinii* pneumonia and the acquired immunodeficiency syndrome (AIDS). *Annals of internal medicine*, 113, pp. 114-120.

Russell, W.M.S. & Burch, R.L. (1992). *The principles of humane experimental technique*. (special edition). London: Methuen & Co, Ltd, London, pp. 110-114.

Van der Lee, J.H. et al. (2008). Efficient ways exist to obtain the optimal sample size in clinical trials in rare diseases. *Journal of Clinical Epidemiology*, 61, pp. 324-330.

Van der Tweel, I. et al (1996). Comparison of one-sample two-sided sequential *t*-tests for application in epidemiological studies. *Statistics in Medicine*, 15, pp. 2781-2795.

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Whitehead, J. (1997). *The design and analysis of sequential clinical trials*. (second edition). Chichester: Wiley.

INGEBORG VAN DER TWEEL is universitair hoofddocent Biostatistiek bij het Julius Centrum voor Gezondheidswetenschappen en Eerstelijns Geneeskunde van het Universitair Medisch Centrum Utrecht. Zij is in 2004 gepromoveerd op toepassingen van sequentiële analyse in epidemiologisch onderzoek. Zij adviseert onderzoekers bij het gebruik van sequentiële analyse in klinisch, epidemiologisch en genetisch onderzoek.

E-mail < i.vandertweel@umcutrecht.nl >

In Memoriam

Cora Maas (1964-2010)



Totaal onverwachts overleed op 8 februari onze collega dr. Cora Maas. Zij werd slechts 45 jaar oud. Cora studeerde onderwijskunde aan de Universiteit Utrecht en promoveerde in 1992 op een proefschrift getiteld *Probleemleerlingen in het basisonderwijs*. Promotores waren Tom Snijders en Wim Meijnen; copromotor was Henk de Vos. Als promovenda was zij verbonden aan de onderzoeksschool Interuniversity Center for Social Science Theory and Methodology (ICS).

Voor het analyseren van de data van haar proefschrift maakte zij gebruik van multilevel analyse, een statistische techniek die op dat moment nog in de kinderschoenen stond. Cora was daarmee één van de pioniers in Nederland en haar liefde voor multilevel analyse bleef zich uiten in haar werk als universitair docent aan het departement Methoden en Technieken van de Universiteit Utrecht, waar zij na haar promotie ging werken. Zo publiceerde zij samen met Joop Hox diverse statistische artikelen en begeleidde zij samen met Joop Hox en Mirjam Moerbeek het promotietraject *Robustness issues for cluster randomized trials* van Elly Korendijk. Daarnaast was zij co-auteur en adviseur bij veel inhoudelijke artikelen waarin multilevel analyse werd gebruikt en was zij betrokken als statistisch adviseur en copromotor bij proefschriften van sociaal-wetenschappelijk onderzoekers. Samen met Jeroen Vermunt was zij gastredacteur van een speciaal nummer van *Statistica Neerlandica* over multilevel analyse.

Cora was actief in het uitdragen van multilevel

analyse binnen de VVS en andere organisaties. Zo bestuurde zij samen met Tom Snijders de Multilevel Onderzoeksgroep (MULOG) van de Nederlandse Organisatie Voor Sociaal-Wetenschappelijk Methodologisch Onderzoek (NOSMO). De MULOG was vooral in de jaren negentig actief. Vanuit deze onderzoeksgroep organiseerde zij enkele keren een multilevel sessie op de Statistische Dag. Daarnaast was zij mede organisator van the International Amsterdam Conference on Multilevel Modelling.

Velen zullen zich Cora herinneren als een bevoegen docente. Zij gaf cursussen Methoden en Technieken op alle niveaus aan studenten binnen de Faculteit Sociale Wetenschappen van de Universiteit Utrecht. Zij had de gave om ingewikkelde stof op een begrijpelijke manier aan studenten uit te kunnen leggen. Ze kreeg daarom keer op keer hoge evaluaties voor haar onderwijs en werd zelfs onderscheiden als docent van het jaar binnen de opleiding psychologie. Ze was ook zeer betrokken bij de junior collega's die veel van haar hebben geleerd. Naast het reguliere onderwijs gaf zij ook cursussen multilevel analyse aan Summer Schools in Ljubljana en Essex.

Wij herinneren ons Cora als een sociale, betrouwbare en hardwerkende collega en missen haar aanwezigheid op de afdeling. Wij willen haar familie en vrienden sterkte wensen in het dragen van dit verlies. Cora, we zullen je missen.

Collega's Departement Methoden en Technieken, Universiteit Utrecht

ER BESTAAN ALLOCHTONE INSLUIPSELS IN HET ENGELS

FRED STEUTEL

In de vorige *STATOR* (2009, 4) deed ik mijn beklag over de alleenheerschap van het Engels, in de statistiek en daarbuiten. Mijn eerste column van dit jaar is wat korter: over de ‘besmetting’ van het Engels met andere talen valt minder te berichten.

Engelstaligen kennen niet veel Frans, en de paar woorden die ze tegenkomen, kunnen ze niet goed uitspreken. Ik kijk en luister wel eens naar *The Weakest Link* van de BBC. Anne Robinson is heel bijdehand, maar een Frans woord dat ter sprake kwam was moeilijk te duiden: ‘lahnge-ray’. Bedoeld was ‘lingerie’ – is ook geen Britse specialiteit. Bij een muziekprogramma hoorde ik iemand een lied aankondigen met de woorden ‘say lahvee, say lahmore’. Bedoeld was ‘c’est la vie, c’est l’amour’. De in de statistiek veel voorkomende naam ‘Poisson’ wordt veelal als ‘Pojsòon’ uitgesproken. Nu we dit weten kunnen we op zoek naar buitenlandse woorden in Engeltalige statistiek; er zijn er toch een paar.

Een veel gebruikte afkorting in de geavanceerde theorie van de stochastische processen is ‘càdlàg’, uitgesproken als ‘kedleg’. Over de precieze uitleg van deze letters bestaat verschil van opvatting. De eerste drie explicaties zijn van niet-Franse auteurs:

continue à droite, limitée à gauche,
continue à droite, limite à gauche
continu à droite, limites à gauche

In een Frans artikel vind ik, neem ik aan, de juiste uitleg:

[fonctions] continues à droite, [ayant des]
limites à gauche.

Functies dus, die rechts continu zijn en linker limie-

ten hebben. Dit is nog niet helemaal alle Frans dat te vinden is, in Engelstalige artikelen. In de literatuur over stochastische processen komt ook het woord ‘balayage’ voor – ik weet niet hoe men dat uitspreekt. Het betekent ‘vegen’, het vegen van de maat van een domein D naar de rand van D .

En er is nog meer. We hebben nog de hybride uitdrukking ‘Porte-Manteau Theorem’. Dat is niet, zoals sommigen denken, ‘de stelling van Porte-Manteau’, maar de ‘kapstokstelling’, een stelling waaraan van alles opgehangen kan worden.

Ook uit het Duits zijn een paar woorden in de Engelstalige kansrekening – en de wiskunde in het algemeen – doorgedrongen. We kennen allemaal de woorden ‘eigenvalue’ en ‘eigenvector’, nog aan te vullen met ‘eigenfunction’ en ‘eigenspace’. In de functietheorie is sprake van ‘schlicht’ functions; Turing hield zich bezig met het ‘Entscheidungsproblem’. Verdere Duitse woorden, niet helemaal beperkt tot de wiskunde: ‘Festschrift’, ‘Ansatz’, ‘Angst’ en ‘verhext’, waarvan de vertalingen voor zichzelf spreken.

Ik ken geen Nederlandse termen in de kansrekening, of het zouden woorden als ‘Stieltjes integral’ moeten zijn. De ‘Lorenz curve’ is niet van Nederlandse oorsprong. Uit het Amerikaanse dagelijks leven ken ik twee woorden van Duits-Nederlandse oorsprong: het gelukssymbool ‘Distelfink’ en het woord voor een vorm van kleiduiven schieten: ‘skeet’.

Met de uit het Arabisch afkomstige woorden zoals ‘algebra’ en ‘algorithm’ zijn we wel aan het eind van de buitenlandse inmenging in het wiskundige Engels.

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



HET EUROPEAN SOCIAL SURVEY

een infrastructuur voor de sociale wetenschappen

INEKE STOOP

Aan het eind van de vorige eeuw kwam een internationale groep sociale wetenschappers tot de conclusie dat goed landenvergelijkend sociaal onderzoek in Europa niet mogelijk was. Gegevens over waarden, houdingen en meningen – essentieel om de samenleving en veranderingen hierin te begrijpen – ontbraken, hadden slechts betrekking op een beperkt aantal landen, schoten tekort in kwaliteit of waren in ieder land op een verschillende manier verzameld. Met steun van de European Science

Foundation ontwikkelden deze wetenschappers een blauwdruk voor het European Social Survey (ESS). Inmiddels wordt in 30 landen het veldwerk voor de vijfde ronde van het ESS voorbereid, met financiële steun van de Europese Commissie en nationale wetenschapsorganisaties. Het ESS heeft in de afgelopen tien jaar een belangrijke aanzet gegeven tot verbetering van de kwaliteit van landenvergelijkend onderzoek en heeft meer dan 25 duizend geregistreerde gebruikers.

Inhoud en deelnemende landen

Het ESS is opgezet om veranderingen in waarden, houdingen, meningen en daaraan gerelateerd gedrag te meten in veranderende samenlevingen. Iedere twee jaar komt een breed scala aan onderwerpen aan de orde (zie Tabel 1). Een kernvragenlijst met onderwerpen die elke ronde terugkomen wordt aangevuld met roterende modules met

wisselende onderwerpen.

De roterende modules worden geselecteerd na een open competitie. Multinationale teams van academici kunnen voorstellen doen voor theoretisch onderbouwde onderwerpen die in alle deelnemende landen een zinvol deel van de vragenlijst kunnen zijn. De Scientific Advisory Board van de ESS selecteert twee voorstellen die vervolgens tot een module in de vragenlijst worden uit-

KERNVRAGENLIJST	ROTTERENDE MODULES
Trust in institutions	ESS 1 (2002/2003)
Political engagement	Immigration
Socio-political values	Citizen involvement and democracy
Social capital, social trust	ESS 2 (2004/2005)
Moral and social values	Family, work and well-being
Social exclusion	Economic morality
Human values	Health and care-seeking
National, religious, ethnic identities	ESS 3 (2006/2007)
Well-being and security	Indicators of quality of life
Demographic composition	Perceptions of life course
Education and occupation	ESS 4(2008/2009)
Financial circumstances	Attitudes to welfare
Household circumstances	Experience and expressions of ageism

Tabel 1. Onderwerpen in het European Social Survey.

gewerkt. Het is de bedoeling dat een deel van de roterende modules na verloop van tijd, mogelijk in verkorte vorm, in de vragenlijst terugkeert om veranderingen te kunnen meten. Op de ESS website <www.europeansocialsurvey.org> staat hoe de vragenlijsten tot stand zijn gekomen en welke vragen precies worden gesteld.

Het aantal deelnemende landen in het ESS neemt toe met de tijd (zie Tabel 2). Uit de tabel blijkt dat de landen ook heterogener worden: het aandeel niet EC-landen is in de vierde ronde bijvoorbeeld groter dan in de eerste. Dat heeft ook gevolgen voor de vragen die men kan stellen en de vergelijkbaarheid van concepten en methoden.

Methoden en vergelijkbaarheid

Het ESS hanteert waar mogelijk een model van inputharmonisatie: met dat model wordt het onderzoek in alle landen op dezelfde manier uitgevoerd en worden in alle landen dezelfde vragen gesteld. De vragenlijst wordt afgenomen in een face-to-face interview van ongeveer een uur, waarvan de helft besteed wordt aan de kernvragenlijst en de helft aan de roterende modules. Inputharmonisatie blijkt niet altijd uitvoerbaar of effectief. Vragen naar partijvoorkeur en opleiding worden bijvoorbeeld in alle landen anders gesteld. De antwoorden worden uiteindelijk wel vertaald naar één begrip.

Een ander terrein waar inputharmonisatie niet uitvoerbaar is, is de steekproeftrekking. In alle landen wordt een random steekproef getrokken en alleen de geselecteerde persoon mag meedoen. Dat betekent bijvoorbeeld dat een vader de vragenlijst niet voor zijn 16-jarige zoon mag invullen, als deze zelf geen zin heeft. De steekproefkaders verschillen echter per land, wat betekent dat in sommige landen een steekproef van individuen wordt getrokken uit de bevolkingsregistratie, in andere landen

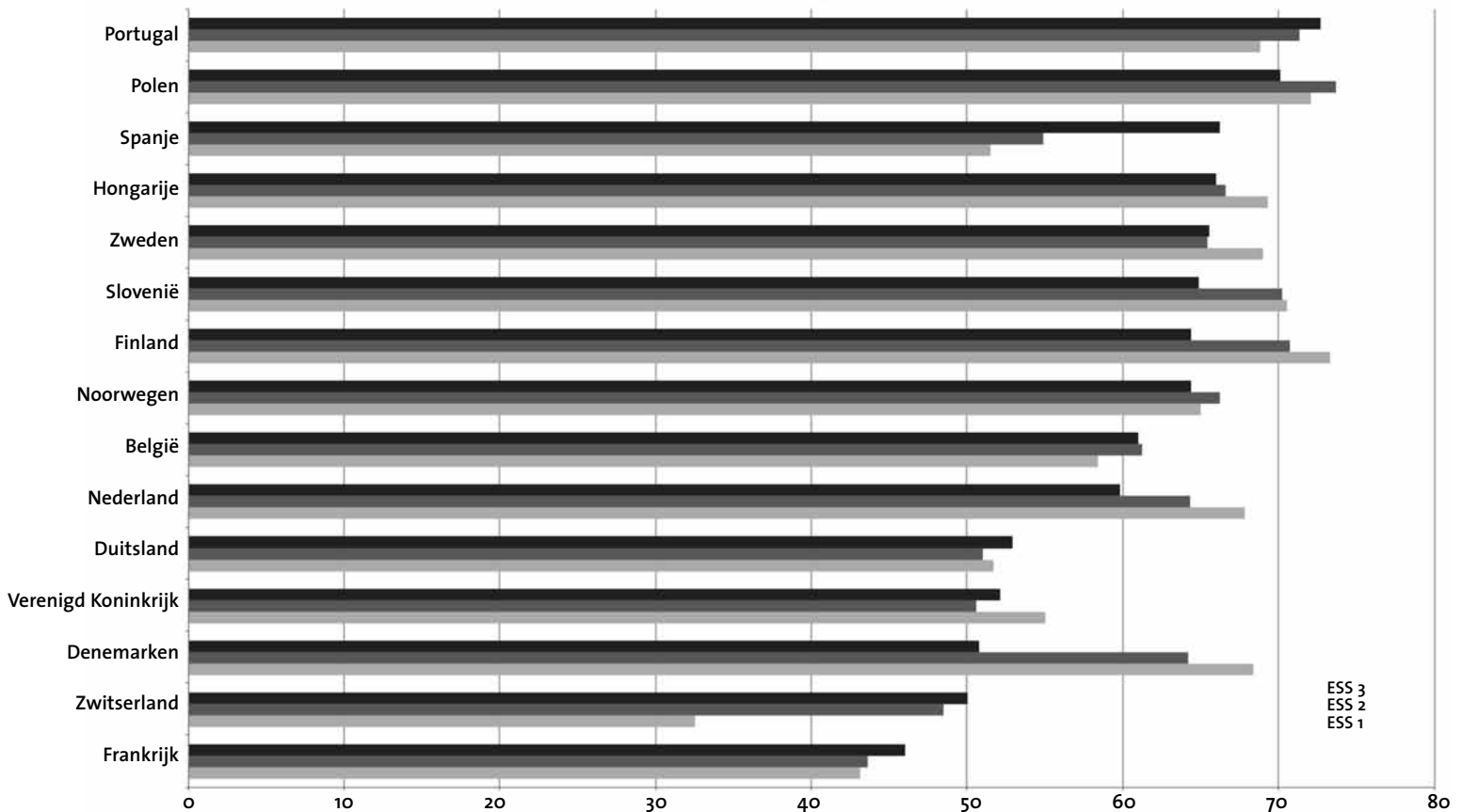
	ESS 1	ESS 2	ESS 3	ESS 4
België	•	•	•	•
Bulgarije			•	•
Cyprus			•	•
Denemarken	•	•	•	•
Duitsland	•	•	•	•
Estland		•	•	•
Finland	•	•	•	•
Frankrijk	•	•	•	•
Griekenland	•	•		•
Hongarije	•	•	•	•
Ierland	•	•	•	•
IJsland		•		
Israël	•			•
Italië	•	•		
Kroatië				•
Letland			•	•
Litouwen				•
Luxemburg	•	•		
Nederland	•	•	•	•
Noorwegen	•	•	•	•
Oekraïne		•	•	•
Oostenrijk	•	•	•	•
Polen	•	•	•	•
Portugal	•	•	•	•
Roemenië			•	•
Rusland			•	•
Slovenië	•	•	•	•
Slowakije		•	•	•
Spanje	•	•	•	•
Tsjechië	•	•		•
Turkije		•		•
Verenigd Koninkrijk	•	•	•	•
Zweden	•	•	•	•
Zwitserland	•	•	•	•
TOTAAL	22	26	25	31

Tabel 2. Deelnemers aan het European Social Survey.

een steekproef van huishoudens en in weer andere (waaronder Nederland) een steekproef van adressen. Centraal bij de steekproeftrekking staat het begrip 'effectieve steekproefgrootte'. Dat betekent dat het uiteindelijke aantal te interviewen personen door ongelijke trekkingskansen en mogelijke clustereffecten veel groter kan zijn dan de 1.500 personen die zouden meedoen in het geval van simple random sampling van individuen.

Ook op het terrein van nonrespons wordt een optimale balans gezocht tussen harmonisatie, uitvoerbaarheid en effectiviteit. Het veldwerk van het ESS wordt betaald door de nationale wetenschapsorganisaties (in Nederland NWO),

die in het eigen land de beste veldwerkorganisatie zoeken. Soms is dat een statistisch bureau, soms een universiteit en soms een commerciële organisatie. De ervaring van deze veldwerkorganisaties met random steekproeven en face-to-face interviews verschilt. In alle gevallen wordt geprobeerd een hoge respons te halen: door het zenden van een aanschrijfbrief naar de potentiële respondenten, het geven van een incentive (een klein cadeautje) aan de respondenten, het veelvuldig langsgaan op verschillende tijdstippen en dagen van de week bij mensen die niet thuis worden getroffen en het overhalen om toch mee te doen van mensen die aanvankelijk nee zeggen. Het streven is een respons van 70%,



Figuur 1. Responspercentage European Social Survey, 1^e drie ronden.

maar dat blijkt in lang niet alle landen haalbaar. In Figuur 1 staat een overzicht van de respons in de landen die aan de eerste drie ronden hebben meegedaan. Uit dit overzicht blijkt dat er grote verschillen zijn tussen landen en iets minder grote in de tijd. In twee landen zijn de verschillen tussen de eerste en de derde ronde overigens wel heel groot. In Zwitserland nam de respons fors toe dankzij zeer grote inspanningen en hoge beloningen voor de respondenten. De daling van de respons in Denemarken in de derde ronde was te wijten aan het feit dat de steekproef getrokken werd uit de bevolkingsregistratie en dat in de derde ronde een aanzienlijk aantal geregistreerden had aangegeven dat ze niet voor enquêtes benaderd wilden worden. Volgens de strenge richtlijnen van het ESS is dat nonrespons.

Niet alleen de responsgeneigdheid verschilt tussen landen. Er zijn natuurlijk ook verschillen in taal, cultuur en instituties. Soms zijn simpele begrippen lastig te vertalen. Zo is 'disagree' in sommige landen alleen te vertalen met 'not agree', wat een bipolaire schaal (van 'disagree' naar 'agree') verandert in een unipolaire schaal (van 'not agree' naar 'agree'). En dan hebben we het niet eens over de middencategorie: 'neither not agree nor agree'. Naast vertaalproblemen kent landenvergelijkend onderzoek ook conceptuele problemen: wat doe je als de institutie 'huisarts' in een aantal landen niet bestaat en je onderzoek doet naar medisch hulpzoekgedrag? Dergelijke problemen zijn niet altijd op te lossen. Bij het ESS wordt echter de betekenis van concepten zoveel mogelijk vooraf doorgesproken en staat in de Engelstalige 'moedervragenlijst' een toelichting over wat we willen meten (met xxx bedoelen we ...). Verder wordt gewerkt met een strikte vertaalprocedure met twee vertalers die hun vertalingen vergelijken en doorspreken

in de aanwezigheid van een derde, waarna het beste alternatief wordt gekozen maar ook de discussie wordt vastgelegd. De vertalingen van landen met dezelfde talen (bijvoorbeeld Duits in Duitsland, Oostenrijk en Zwitserland) worden ook vergeleken om onnodige verschillen te vermijden. Desondanks gaat er nog steeds wel eens iets mis, zoals de keer dat 'wealthy' werd vertaald met het equivalent van 'healthy'.

Continuïteit en toekomst

Het ESS lijkt voor een belangrijk deel zijn doel, verbetering van de kwaliteit van landenvergelijkend onderzoek, bereikt te hebben. Na vier ronden dataverzameling zijn er nu meer dan 25 duizend geregistreerde gebruikers. Het aantal publicaties over het ESS, zowel inhoudelijke als methodologische, groeit snel en ook nemen andere surveys, nationaal en internationaal, de aanpak van het ESS of delen daarvan over. In 2007 is het ESS door een internationale commissie beoordeeld (Bethlehem et al, 2008): 'The panel unanimously finds that the importance of ESS, its demonstrated success in initial launch, and its clear signals of impact justify fully continuous funding at levels necessary to achieve its vision and maintain its quality.' Daarmee stipt het review panel echter ook een van de problemen aan: de financiering.

Goed survey-onderzoek is niet goedkoop, en voor het onderhouden en vernieuwen van een goede onderzoeksinfrastructuur in de sociale wetenschappen (wat het ESS nu is) is veel geld nodig. Het ESS is de afgelopen jaren gefinancierd door nationale wetenschapsorganisaties, verschillende kaderprogramma's van de EU en de European Science Foundation. Meestal moest in ieder land voor elke ronde opnieuw om geld

worden gevraagd, en dat geld kwam er niet altijd, of niet in voldoende mate, of niet op tijd. Gelukkig staat het ESS nu internationaal en in Nederland op de roadmap voor grootschalige onderzoeksinfrastructuren. Hierdoor ziet de toekomst er iets zonniger uit, maar de financiële toekomst blijft onzeker.

Continuïteit is echter niet alleen een financieel probleem. Omdat het ESS is opgezet om ontwikkelingen door de tijd te meten, is het van belang dat vraagstellingen en methoden gelijk blijven. Trends in de tijd zijn immers moeilijk te meten als steeds iets anders wordt gevraagd, of als een vraag steeds op een andere manier wordt gesteld. Dat is een lastig probleem, om verschillende redenen. Ten eerste komen uit de actualiteit verschillende onderwerpen naar voren die eigenlijk in de vragenlijst opgenomen zouden moeten worden (denk aan de klimaatverandering of de kredietcrisis). Ten tweede kunnen vragen verouderen. In de vorige eeuw hadden we nog niet kunnen denken dat internet bij mediagebruik zo'n grote rol zou spelen. Ten derde kunnen ook vraagmethoden verouderen. Het is bijvoorbeeld niet zeker dat in de Scandinavische landen face-to-face surveys zullen blijven bestaan. Als dan alleen het ESS op deze manier wordt uitgevoerd, beschikt men in die landen niet meer over ervaren interviewers. Dat zou een reden kunnen zijn om op een andere vraagmethode over te gaan, maar dat maakt de resultaten weer minder vergelijkbaar. En ten slotte hebben we de afgelopen jaren veel geleerd over steekproef-designs, nonrespons, het testen van vragenlijsten en allerlei andere onderwerpen. Deze verworven kennis zouden we graag in de praktijk brengen, maar als we te veel veranderen of verbeteren komt de vergelijkbaarheid in de tijd weer in het geding.

Het ESS is uitgegroeid van een droom van een groep wetenschappers tot een grootscha-

lige infrastructuur. Dat zal op korte termijn overigens grote gevolgen hebben voor de organisatiestructuur. Inmiddels zijn de data van de eerste vier rondes te downloaden vanaf de ESS-website <www.europeansocialsurvey.org>. De voorbereidingen voor de vijfde ronde zijn nu in volle gang. Op de website is behalve de data ook veel te vinden over alle praktische en methodologische aspecten van de dataverzameling. Als je het nog niet gedaan hebt, schroom niet en ga gauw eens kijken. Hoe meer gebruikers hoe beter.

LITERATUUR

- Bethlehem, J., Medrano, J., Groves, R., Gundelach, P. & Norris, P. (2008). *Report of the Review Panel for the European Social Survey*. Strasbourg Cedex: European Science Foundation.
- Commissie Nationale Roadmap Grootschalige Onderzoeksfaciliteiten (2008). *Nederlandse Roadmap Grootschalige Onderzoeksfaciliteiten*. Amsterdam: Auteur.
- European Strategy Forum on Research Infrastructures (ESFRI) (2006). *European Roadmap for Research Infrastructures, Report 2006*. Luxembourg: Office for Official Publications of the European Communities.
- Jowell, R., Kaase, M., Fitzgerald, R. & Eva, G. (2007). The European Social Survey as a measurement model. In: R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (eds.) *Measuring Attitudes Cross-Nationally. Lessons from the European Social Survey*, pp. 1-31. London: Sage.
- Stoop, I.A.L. (2006). Landenvergelijkende surveys: Vergelijking tussen landen en tussen surveys. In: A.E. Bronner et al (eds.). *Ontwikkelingen in het marktonderzoek. Jaarboek 2006 Marktonderzoek Associatie*, pp. 95-107. Haarlem: Spaar en Hout.
- Stoop, I., Billiet, J., Koch, A. & Fitzgerald, R. (forthcoming). *Improving Survey response : Lessons learned from the European Social Survey*. Wiley.

INEKE STOOP is Afdelingshoofd van de Stafafdeling Informatievoorziening & Automatisering van het Sociaal Cultureel Planbureau. Ze studeerde methoden en technieken van onderzoek in de psychologie aan de Universiteit Leiden en promoveerde in Utrecht.
E-mail: <i.stoop@scp.nl>.

IN MEMORIAM GIJS DE LEVE (1926 - 2009)

Gijs de Leve werd op 15 augustus 1926 geboren in Amsterdam maar verhuisde op tweejarige leeftijd met zijn ouders naar Den Haag, waar hij zijn jeugd doorbracht. Na onderwijs op de HBS gevolgd te hebben, waar hij geen eindexamen deed vanwege de oorlogsomstandigheden, ging hij in 1945 wiskunde en natuurkunde studeren aan de Universiteit van Leiden na eerst overwogen te hebben scheikunde te gaan doen. Kandidaats heeft Gijs gedaan in december 1948, op dezelfde dag als zijn studiegenoot en goede vriend Koos Kriens, met wie hij later vele jaren zou samenwerken op het Mathematisch Centrum (MC) in Amsterdam. Na zijn doctoraal wiskunde in Leiden gedaan te hebben, trad Gijs op 1 oktober 1954 in dienst van het MC.

In de begintijd van het MC werd een nieuwe medewerker min of meer aan zijn lot overgelaten en moest hij zijn eigen weg vinden. Kenmerkend is de anekdote dat Gijs op zijn eerste werkdag door de directeur werd aangezien voor de nieuwe magazijnmeester, die ook die dag in dienst trad. Een anekdote waar Gijs smakelijk over kon vertellen. In zijn eerste jaren op het MC is Gijs samen met Koos Kriens intensief bezig geweest om cursussen *Operations Research* op te zetten, cursussen waaruit de beroemde en veelgebruikte *Leergang Besliskunde* met zijn deeltjes 1 t/m 7c ontstaan is. In 1964 is Gijs gepromoveerd bij Jan Hemelrijk op het proefschrift *Generalized Markovian Decision Processes*. In dit proefschrift vol originele ideeën werd een stochastische beslissingstheorie ontworpen voor bestuurbare dynamische systemen met een algemene toestandsruimte en een continue interventiemogelijkheid. Zijn theorie was te abstract en te veelomvattend om ingang te

vinden in de toegepaste wetenschappelijke literatuur. Echter, de kerngedachten van het werk van Gijs hebben hun weerklank gevonden in het veelvuldig geciteerde boek *Stochastic Modelling and Analysis* van zijn eerste promovendus Henk Tijms.

Op de afdeling Mathematische Besliskunde van het MC is in de jaren zeventig en tachtig onder leiding van Gijs een internationaal georiënteerde groep van onderzoekers gevormd, die zowel op het gebied van de Markov beslissingstheorie als de combinatorische optimalisering wereldfaam heeft verworven. Menigeen van deze onderzoekers was een student van Gijs aan de Universiteit van Amsterdam, waar hij op 1 januari 1965 tot deeltijdhoogleraar was benoemd. Met zijn creativiteit en sprankelende ideeën heeft Gijs zijn medewerkers op het onderzoekspad begeleid, waarbij hij zichzelf niet op de voorgrond stelde maar het publiceren aan anderen overliet. Twaalf dissertaties met een breed spectrum van onderwerpen zijn onder zijn supervisie tot stand gekomen en vijf van zijn promovendi zijn later hoogleraar geworden, met zelf weer vele promovendi die indirect schatplichtig zijn aan Gijs. Zonder overdrijving kan Gijs de Leve de 'Godfather' van de besliskunde in Nederland genoemd worden.

Niet alleen bracht Gijs de Leve vele studenten ertoe in de besliskunde verder te gaan, hij heeft ook veel betekend voor de popularisering van de besliskunde bij bedrijfsleven en overheid in Nederland. Een initiatief dat niet onvermeld mag blijven, betreft de zogenaamde werkweken die hij begin jaren zeventig op het MC introduceerde om mensen uit de praktijk in aanraking te brengen met de nieuwste ontwikkelingen in de besliskun-





de. Gijs had een scherp oog voor toekomstige en gewenste ontwikkelingen in het vak. Gijs is niet alleen de vader van de besliskunde in Nederland, hij is ook de architect van de opmerkelijke eenheid en eensgezindheid onder de Nederlandse besliskundigen. In mei 1975 legde hij in zijn oude Simca een handelsreizigersroute af langs zijn collega-besliskundigen in het land. Tijdens deze tocht, waarop hij vergezeld werd door zijn toenmalige medewerker Jan Karel Lenstra, legde Gijs de basis van de adviescommissie van de afdeling Mathematische Besliskunde van het MC, van een Newsletter, van de bijeenkomsten van besliskundigen in Lunteren en van een Landelijke Werkgemeenschap Mathematische Besliskunde. Kortom, een samenwerking werd tot stand gebracht waardoor het in 1987 vanzelfsprekend was dat de besliskundigen één van de eerste groepen waren met een landelijk netwerk voor promovendi met gezamenlijke colleges, een netwerk dat tot op de dag van vandaag heel succesvol is. De innemende persoonlijkheid van Gijs, met altijd een oprechte en warme belangstelling voor

anderen, heeft veel bijgedragen aan het eenwordingsproces in de besliskunde in Nederland.

Gijs heeft zijn werk op het MC lang gecombineerd met een hoogleraarschap aan de UvA. Op 1 september 1983, de dag dat het Mathematisch Centrum werd omgedoopt tot Centrum voor Wiskunde en Informatica, trad Gijs terug als chef van de afdeling Mathematische Besliskunde en werd hij adviseur van de directie van het instituut. Zijn werk had hij toen al grotendeels verlegd naar de UvA, waar hij met zijn hoffelijkheid en diplomatieke gaven alom gerespecteerd werd aan de roerige Interfaculteit voor Actuarieel en Econometrie. Als decaan zorgde hij tevens voor vernieuwingen en was hij één van de grondleggers van een zelfstandig en op het bedrijfsleven gericht instituut voor toegepaste statistiek.

Hoewel Gijs na zijn emeritaat in 1991 contact bleef houden met de universitaire wereld en hij nog vijfmaal de tweejaarlijkse, naar hem vernoemde Gijs de Leve prijs voor het beste proefschrift op het gebied van de besliskunde, heeft mogen uitreiken, heeft Gijs zich in deze periode voornamelijk gericht op zijn familieleven met zijn vrouw Annechien, kinderen en kleinkinderen. Zeven jaar na zijn vrouw Annechien is Gijs in zijn geliefde omgeving van Ratum op 19 november 2009 overleden na een noodlottige val op zijn boerderij. Wij zullen ons Gijs de Leve als een origineel en creatief wetenschapper maar bovenal als een beminnelijk en warm mens blijven herinneren.

JAN KAREL LENSTRA
HENK TIJMS



Tentamenweek

JOHAN VAN LEEUWAARDEN

Verdieping tien doet dienst als tentamenzaal. In normale weken is de lift een nare ruimte, waar normale mensen zich opgelaten voelen. Tentamenweken zijn anders. De lift is voller dan anders, qua gewicht en emoties. Gesperde pupillen zoeken elkaar. De eenzaamheid van het studeren is vergeten zodra de liftschacht zich opent. Met maximale bezetting suist de stalen kooi door de beladen ether. Een kleine ruimte vol lotgenoten, mensen met een missie. Een enkeling mompelt iets, maar er is geen ruimte voor liftpraatjes. Onvaste gelaatstrekken verraden de toestand van mijn medelifters. Niemand weet wat er boven wacht.

Ik stap als enige uit op verdieping negen, loop naar mijn kamer, pak de grote bruine envelop en neem de wenteltrap. Surveilleren. Driehonderd tafels, gerangschikt als een Romeins legioen, in paraatheid gebracht voor de op handen zijnde slag. De zenuwen gieren door de ruimte. De heroïek van de deelnemer is onomstreden. Voor mij het legertje studenten dat mijn tentamen zal trotseren. Twaalf weken lang heb ik ze voorbereid op dat wat komen gaat, maar er zullen slachtoffers vallen. Een goed tentamen maakt immers slachtoffers, biedt een reële kans op overleven, en laat ruimte voor een zwaarbevochten triomf.

Ik open de verzegelde envelop en begin uit te delen. Vol overgave storten de deelnemers zich op het tweezijdige velletje. De meesten nemen de vijf vragen vluchtig door en beseffen meteen dat het kantje boord gaat worden. De lage tonen van de radiatoren gaan op in het zuchten van de menigte. Er zijn maar weinig momenten in het leven waar-

op de mens precies weet wat hem te doen staat, maar dit is er zo een. Het tentamen de strijd aanzeggen. Het tentamen dat de weg verspert tussen het nu en een onbezorgde toekomst. Verwijtende blikken. Ik ben het immers die hen dit aandoet. Driftig wordt er geschreven op de kladvellen. Ook de calculators worden aangewend, terwijl ik weet dat er niets te rekenen valt. De ramen beslaan.

Na een klein uur begint het gapen, een teken van verslapping. Spanning wordt berusting en het onbestemde gevoel maakt plaats voor honger. Boterhammen. Zelf neem ik er ook nog een. Wie moet presteren kent geen honger. Sissende blikjes energiedrank doorbreken de stilte. Het pleit is beslecht en de rust keert weer. Nog twee uur zal het tentamen voortduren, maar de grootsheid van het moment is vervlogen.

Voormalige atleten kunnen trainer worden, maar zo gelukkig als toen worden ze nooit meer. Aan de zijlijn staan. Voor voormalige studenten die docent worden, geldt dat evenzeer. Kent u het nog, dat gevoel dat het nu moet gebeuren, nu of nooit? Twee keer per jaar word ik eraan herinnerd, als ik weer een tentamen in elkaar sleutel. Dan wil ik ook direct door naar het strijdtoneel op vloer tien. Laat me nog een keer een tentamen doen, liefst een met onredelijke vragen, die ik met veel moeite uit het vuur kan slepen.

JOHAN VAN LEEUWAARDEN is werkzaam in de groep Stochastische Besliskunde bij de faculteit Wiskunde en Informatica van de TU Eindhoven. Tevens is hij research fellow bij EURANDOM.

E-mail: <j.s.h.v.leeuwaarden@tue.nl>.