

STATOR

THEMA ETHIEK

Openheid is noodzakelijk

Statistiek, onbehagen en de menselijke maat

Medisch-ethische vraagstukken van de verloskunde

Statistici mogen niet liegen; beroepsethiek onder het vergrootglas

Ethische vragen in operations research

Ethiek en OR

De wereld der waarden: statistiek, evolutie en ethiek

Ethiek in de OR-business

Ethiek met sambal – column

Informed consent: noodzakelijk kwaad?

Randomised response: over het hoe en waarom van privacybescherming

Onafhankelijke statistiek

De menselijke kant van Statistiek

Statistiek, data monitoring en ethiek

Ook ik heb gezondigd – column

Ethisch optimaliseren van het ruilen van nieren

Wat is krediet in het casino waard? Een ethische vraag

INHOUD

- 3** De focus op ethiek
- 4** Openheid is noodzakelijk
Richard Gill
- 7** Statistiek, onbehagen en de menselijke maat
Richard Starmans
- 10** Medisch-ethische vraagstukken van de verloskunde
Ingrid Brussé
- 14** Statistici mogen niet liegen; beroepsethiek onder het vergrootglas
Willem de Vries
- 17** Ethische vragen in operations research; persoonlijke overpeinzingen
Jack P. C. Kleijnen
- 21** Ethiek en OR
Alexander Rinnooy Kan
- 22** De wereld der waarden: statistiek, evolutie en ethiek
Richard Starmans
- 26** Ethiek in de OR-business
Gerrit T. Timmer
- 29** Ethiek met sambal – column
Fred Steutel
- 30** Informed consent: noodzakelijk kwaad?
Gerben Bekker & Hilde Tobi
- 33** Randomised response: over het hoe en waarom van privacybescherming
Ardo van den Hout & Marije Altorf
- 37** Onafhankelijke statistiek
Kees Zeelenberg
- 40** De menselijke kant van Statistiek
Jelte M. Wicherts
- 44** Statistiek, data monitoring en ethiek
Kit C. B. Roes
- 51** Ook ik heb gezondigd – column
Gerrit Stermerdink
- 52** Ethisch optimaliseren van het ruilen van nieren
Kristiaan Glorie, Joris van de Klundert & Albert Wagelmans
- 57** Wat is krediet in het casino waard? Een ethische vraag
Henk Tijms
- 59** Young Statisticians
- 60** In Memoriam Rien van der Leeden (1956-2012)
Willem Heiser
- 61** In Memoriam Wijbrandt van Schuur (1946-2012)
Roel Popping
- 62** Agenda
- 62** Colofon
- 63** Oproep om kandidaten te nomineren voor de VvS+OR Thesis Award 2012
Peter Grunwald & Mark van der Loo

DE FOCUS OP ETHIEK



Dit themanummer heeft Ethiek als onderwerp. Veel lezers zullen daarbij als eerste aan iets on-ethisch denken: wetenschappelijke fraude. En inderdaad, enkele van de artikelen in dit nummer gaan over fraude en hoe dat te ontdekken en er mee om te gaan. Maar het is belangrijk om toch ook aan dat ándere aspect te denken, het ethisch handelen.

Een voorbeeld dat vrijwel iedereen kent is de gewetensstrijd van veel van de fysici die in de Tweede Wereldoorlog de atoombom ontwikkelden. Mag men zijn wetenschappelijke kennis wel inzetten om een wapen te ontwikkelen waarmee honderdduizenden slachtoffers gemaakt kunnen worden? Maar wat was het alternatief? Toen de ontwikkeling startte waren Duitsland en Japan bezig grote delen van Europa en Azië te veroveren en men hoopte daar met de bom een einde aan te kunnen maken. Daarnaast bestond de vrees dat Duitsland óók bezig was een atoomwapen te ontwikkelen. Het is illustratief dat Albert Einstein, die bekend stond als een pacifist, in een brief aan de Amerikaanse president heeft aangedrongen op de ontwikkeling van de bom. Maar na de oorlog was hij weer een fervent tegenstander van het gebruik daarvan!

Het thema Ethiek van deze *STAtOR* is ongeveer een jaar geleden door de redactie gekozen. Ethiek, en professionele ethiek zoals moet blijken uit integer professioneel handelen; het heeft aan actualiteit sindsdien niet ingeboet. Integendeel. In het afgelopen jaar zijn op verschillende instituten ethische commissies en vertrouwenspersonen voor wetenschappelijke integriteit ingesteld of afgestoft. En eind september heeft de commissie Schuyt haar rapport *Zorgvuldig en integer – omgaan met wetenschappelijke onderzoeksgegevens* aan de KNAW en een groot aantal belangstellenden gepresenteerd. U heeft erover kunnen lezen in vrijwel alle dagbladen. Bij de presentatie van het rapport werd onder meer aandacht gevraagd voor de verantwoorde-

lijkheid van opleiders en co-auteurs (zoals de begeleiders van promovendi), het gevaar van de *lone wolf* met autoriteit (casus Stapel), het belang van openheid van data waar dat kan, en het belang van weerstand bieden tegen de *publish or perish*-druk.

Dat de dagelijkse praktijk van onderzoekers op gespannen voet kan staan met de normen van wetenschappelijk onderzoek blijkt indirect en wellicht onbedoeld uit de bijdrage van Richard Gill. De enorm hoge eisen die veel werkgevers aan productiviteit stellen hebben ook hun gevolgen. Zo zijn er bibliometrische ontwikkelingen die co-auteurs een gewicht toekennen, gebaseerd op auteursvolgorde opdat de academische output, die uw promotie in Tenure Track bepaalt, nóg nauwkeuriger becijferd kan worden. Dat integriteit in de praktijk moeilijk zal blijven, blijkt ook duidelijk uit de bijdrage van Jelte Wicherts.

Maar deze *STAtOR* wil u ook uitdagen met enkele filosofische bijdragen (zoals de bijdragen van Richard Starmans), een interessante bijdrage van Alexander Rinnooy Kan als gelegenheidscolumnist en columns met extra lading van onze meer reguliere columnisten.

Verder geven we u een kijkje in de ethische problematiek waar statistici en gebruikers van statistiek mee te maken hebben in de geneeskunde. Daarnaast kunnen we lezen hoe toepassingen van de OR tot ethische problemen kunnen leiden: verschillende auteurs zijn bereid gevonden om te vertellen hoe zij met ethiek in hun OR praktijk om zijn gegaan.

We doen in dit nummer meer dan alleen problemen signaleren: er worden voorbeelden gegeven van relevante gedragscodes, een strategie om problemen te minimaliseren (randomised response) of – altijd fijn – vóór te zijn (informed consent).

Wij wensen u veel leerzaam leesplezier.
de STAtOR-redactie

OPENHEID IS NOODZAKELIJK

Ethiek, in het bijzonder wetenschappelijke en professionele ethiek, is weer in de mode. Daar dienen we de financiële crisis voor te danken, die een periode afsloot waarin de maatschappij hebzucht leek te waarden en ethiek een achterhaald begrip was. Laten we dankbaar gebruik van maken van deze hernieuwde aandacht. In deze inleiding wil ik relaas doen van enkele recente persoonlijke ervaringen. Dankzij Lucia de Berk weten Nederlandse journalisten tegenwoordig goed professionele statistici te vinden; ik ben meermalen benaderd toen de zaken Smeesters en Geraerts zich aan het ontvouwen waren. Ik vind dat het onze – leden van de VvS+OR – gemeenschappelijke verantwoordelijkheid is om bij dit soort voorvallen de media met advies en informatie bij te staan. Het bestuur van de VvS+OR, waarvan ik tot voor kort voorzitter mocht zijn, is dan ook erg blij met het initiatief van de STAtOR-redactie om het jaarlijkse themanummer aan Ethiek te wijden. De zeer vele bijdragen die zijn ontvangen onderstrepen het belang van dit onderwerp.

De Smeesters affaire barstte in juni 2012 los met persberichten van de Erasmus Universiteit. Het begon met het artikel '*The effect of color (red versus blue) on assimilation versus contrast in prime-to-behavior effects*' van Dirk Smeesters (Erasmus Universiteit) en Jia Liu (Universiteit Groningen), gepubliceerd in mei 2011 in *Journal of Experimental Social Psychology*, 47(3), 653–656. Rond september 2011 werd *fraud buster* Uri Simonsohn (Wharton University) door een kennis geattendeerd op merkwaardige patronen in de cijfers in dit artikel. Simonsohn was bezig een methodiek te ontwikkelen waarmee data-massage soms opgespoord kan worden en liet zijn methode los op het onderhavige artikel.

Simonsohns idee is eenvoudig: artikelen in de sociale psychologie bevatten vaak tabellen met gemiddeldes en standaarddeviaties van variabelen, uitgesplitst over verschillende klassen van respondenten. Tegelijkertijd valt in het artikel te lezen dat sommige van deze klasse-indelingen er niet toe doen; ze kunnen gepoold worden. Dit betekent dat de spreiding tussen de klassegemiddelden ons ook informatie geeft over de binnen-klasse variatie. In een klassieke ANOVA gebruiken we dit verschijnsel om de hypothese van gelijke verwachtingswaardes te toetsen. Een hele kleine p -waarde van de F -toets vertelt ons dat de hypothese van gelijke verwachtingswaardes onhoudbaar is: de tussen-groepen variatie is veel groter dan zij hoort te zijn op grond van de binnen-groepen variatie, in het geval er geen werkelijk verschil is tussen de groepen. Maar wat als de p -waarde van de bijbehorende F -toets juist heel erg dicht bij 1 is? Dit kan een andere alarmbel laten rinkelen.

Simonsohn vroeg om het oorspronkelijke data bestand behorend bij het onderzoek en kreeg dat van Smeesters. Hij gebruikte de niet-parametrische *bootstrap*-methode om zijn eerdere vermoeden verder te onderbouwen. Zijn conclusie was duidelijk: deze data

zijn te mooi om waar te kunnen zijn. Zo lijken bijvoorbeeld de groepsgemiddeldes veel te veel op elkaar. Hij kwam ook andere merkwaardige patronen in de data tegen en kwam tot de conclusie dat de gegevens niet alleen behoorlijk gemasseerd waren maar zelfs wellicht gefabriceerd. Op een bepaald moment hebben zowel Smeesters als Simonsohn de commissie wetenschappelijke integriteit van Erasmus op de hoogte gesteld van Simonsohns verdenking van fraude.

Al in september 2011, kort na de eerste contacten tussen Simonsohn en Smeesters, crashte volgens Smeesters zijn harde schijf, waardoor alle databestanden van alle onderzoeken die hij in het verleden gedaan had verloren gingen. Tijdens een eerdere verhuizing waren ook, volgens Smeesters, alle originele papieren versies van de bestanden kwijtgeraakt. De commissie wetenschappelijke integriteit van Erasmus benaderde alle co-auteurs van Smeesters, maar het bleek dat geen van allen ooit de data hadden ingezien van de artikelen waarvan ze co-auteur waren. De commissie was genoodzaakt – bij het uitvoeren van Simonsohns toets op de gegevens van andere artikelen – om de parametrische bootstrap te gebruiken (gebaseerd op veronderstelling van normaliteit).

Dit allemaal leidde tot verdenkingen van ongeoorloofde data-massage in meerdere publicaties van Smeesters. In juni 2012 heeft de commissie zijn bevindingen gepubliceerd en werd Smeesters' ontslag aanvraag geaccepteerd. Hierdoor verloor hij niet alleen zijn baan bij de Erasmus Universiteit, maar werd ook zijn reputatie als wetenschapper publiekelijk vernietigd. De basis van deze veroordeling ligt in statistische analyses van data-bestanden behorend bij publicaties van Smeesters, maar tot op heden zijn zowel details van de analyses als de data zelf geheim. Zelfs de statistici onder de leden van de Commissie Wetenschappelijke Integriteit beantwoorden geen vragen van vakgenoten, ongetwijfeld op last van hun autoriteiten. Zelf ben ik

ervan overtuigd dat men een methodiek heeft gebruikt om *error rates* onder controle te houden bij het toetsen van zeer vele hypothesen, namelijk de FDR-methode die absoluut ongeschikt is in deze situatie. De FDR veronderstelt namelijk vooraf, dat een bepaalde percentage van de zeer vele te toetsen nul-hypothesen niet waar is. In deze situatie houdt dit in, kort door de bocht, dat we vooraf veronderstellen dat Smeesters in een bepaald percentage van zijn artikelen gefraudeerd heeft! Tot op heden is het niet mogelijk geweest om via Simonsohn, Smeesters, Liu, of de Erasmus Universiteit het data-bestand behorend bij *the colour red* in te zien. (medeauteur Liu zegt het nooit te hebben gehad).

De kern van de wetenschappelijke methode is reproduceerbaarheid. In onderzoek waar statistiek wordt gebruikt, is de statistische analyse van de gegevens een centraal onderdeel van het gehele onderzoek. Alle bewerkingsstappen moeten geregistreerd worden en de oorspronkelijke bestanden moeten zo mogelijk behouden worden, zodat ten alle tijde het gehele proces van verwerking en analyse herhaald kan worden. Geïnteresseerde onderzoekers moeten zoveel mogelijk brongegevens en protocollen (scripts) kunnen inzien. In het bijzonder is er iets ernstig mis in een wetenschapsgebied waar het kennelijk gewoonte is dat medeauteurs van wetenschappelijke publicaties volstrekt onbekend zijn met de data en de data-bewerkingen die geleid hebben tot het artikel waar hun naam boven staat.

Hierbij aansluitend vind ik dat de eisen van reproduceerbaarheid, en openheid over mogelijke punten van kritiek, ook moeten gelden voor het werk van een commissie die mogelijke fraude onderzoekt. Aangezien de Erasmus Universiteit de stap nam om hun bevindingen openbaar te maken ligt hier een verantwoordelijkheid voor hen.

De Geraerts affaire

Recent is een nieuw affaire ontstaan rondom het artikel '*Linking thought suppression and recovered memories*

of childhood sexual abuse' van Elke Geraerts, Richard J. McNally, Marko Jelicic, Harald Merckelbach, en Linsey Raymaekers, in 2008 verschenen in *Memory*, 16(1), 22–28. In het artikel worden data gekoppeld die in Harvard en in Maastricht zijn verzameld. De senior-auteurs Merckelbach en McNally (Maastricht en Harvard) herkennen niet de oorspronkelijke gegevens verzameld in hun twee instituten in het totaal bestand, en verdenken de junior-onderzoeker Geraerts (tevens hoofdauteur van het artikel – toen werkzaam in Maastricht, nu in Rotterdam) van grootschalige niet-correcte manipulatie van de gegevens.

Ook hier zijn geen volledige data-sets en geen volledige scripts met data-bewerkingen beschikbaar. Kennelijk hebben de meeste auteurs van het *Memory*-artikel dan ook geen inzage gehad in de volledige (gecombineerde) gegevens en de bewerkingen die Geraerts op ze heeft uitgevoerd, totdat ze jaren later bedenkingen kregen en in conflict raakten.

Tot mijn verbazing was in de gecombineerde dataset van Geraerts geen indicatie meer te vinden over welke respondent uit Harvard of uit Maastricht afkomstig was. Ook wordt in het artikel geen vergelijking tussen deze twee groepen gemaakt. Merkwaardig. Hoe zit het met mogelijk culturele verschillen?

Conclusie

Ik kan alleen maar herhalen: de kern van iedere wetenschappelijke methode is reproduceerbaarheid. In onderzoek waar statistiek gebruikt wordt, is de statistische analyse van de gegevens een centrale onderdeel van het gehele onderzoek. Laten we veel leren voor de toekomst van de in deze inleiding beschreven blunders.

RICHARD GILL is hoogleraar Statistiek aan de Universiteit Leiden. Hij geldt als een van de meest invloedrijke statistici van Nederland. Richard was van 2008 tot 2011 voorzitter van de VvS+OR.
E-mail: <gill@math.leidenuniv.nl>

STATISTIEK, ONBEHAGEN EN DE MENSELIJKE MAAT

RICHARD STARMANS

De wijsgerige ethiek is bij uitstek de filosofische discipline die zich bezighoudt met de concepten van goed en kwaad, de studie van waarden en normen, soms ook enigszins archaisch, maar eloquent aangeduid als de leer van het zedelijk handelen van de mens. Thema's als het morele oordeel, vrijheid en verantwoordelijkheid, deugden en plichten, schuld en boete, het geweten als (ken)bron van de morele ervaring, ethische dilemma's en valkuilen raken ook het werkterrein van andere wijsgerige disciplines, zoals de wijsgerige antropologie, sociale filosofie, kennisleer en in ruimere zin de wetenschapsfilosofie.

Vanzelfsprekend vormt dit alles niet het exclusieve domein van de wijsbegeerte. Alle vormen van menselijk handelen kennen een morele dimensie en qualitate qua krijgen theologen, geestelijk leiders, politici, juristen, artsen, wetenschappers, ondernemers en feitelijk alle burgers of *Vernunftwesen* evenzeer met voornoemde thema's te maken. De vele invalshoeken en begrippen die hiermee verbonden zijn vormen en vormen dikwijls een opmaat tot conceptuele verwarring. Zonder precies een boedelscheiding of verkaveling te willen bewerkstelligen hebben filosofen pogingen gedaan deze te verminderen. De drieslag moraal, ethiek en meta-ethiek kan daarbij van nut zijn. Moraal is dan het geheel aan overtuigingen, uitingen

en gedragingen van mensen inzake goed en kwaad. De term ethiek wordt afwisselend gedefinieerd als de studie van de moraal, dan wel als een uitgewerkt moreel kader, een theorie, bij voorbeeld een christelijke ethiek, een plichtethiek of het utilitarisme. Deze ambiguïteit wordt van oudsher weerspiegeld in een principiële tweedeling binnen de ethiek: beschrijvende versus normatieve ethiek. Men kan de ethiek zuiver beschrijvend benaderen en op de stoel van de historicus of antropoloog plaats nemen zonder daarbij zelf ethische standpunten in te nemen. Doel is dan te trachten waardehiërarchieën en motieven en gevolgen van het morele handelen in verschillende culturen of tijdsgewrichten te ontrafelen, in kaart te brengen en te vergelijken. Men kan ook een normatieve benadering voorstaan, die vaak in de moraalfilosofie op de voorgrond treedt. Zo'n ethiek vooronderstelt een mens- en wereldbeeld en behelst een moreel appèl. Centraal staat het ontwikkelen van een theoretisch kader van waaruit men tracht de moraal te rechtvaardigen of anderszins te funderen, doorgaans door te zoeken naar onwrikbare grondslagen of 'eerste principes' waartoe deze is te herleiden. Voorbeelden zijn een ethiek gebaseerd op leven conform de natuur (de Stoa), identificatie van deugden als leidraad voor het menselijk handelen (Aristoteles, MacIntyre) of een

deontologische, op plicht gebaseerde ethiek (Kants categorische imperatief). Canoniek is uiteraard ook een theonome ethiek, gebaseerd op religieuze uitgangspunten, sacrale teksten of openbaringen en een door God gegeven universele moraal (Thomas van Aquino) of een heteronome ethiek, zoals het utilitarisme of sociale contracttheorieën. Complementair aan dit alles kan men een verklarende ethiek nastreven, waarbij een verklaring wordt gezocht voor de morele vermogens of het ontstaan van de bestaande moraal. Dikwijls staan daarbij genetische verklaringen centraal, gericht op de wordingsgeschiedenis van een verschijnsel. Beroemde voorbeelden hiervan zijn onder meer Nietzsches *Zur Genealogie der Moral* en geheel andersoortig, een evolutionaire ethiek. Zo trachten bij voorbeeld sociobiologen in navolging van Darwin de moraal te verklaren als een (bij)product van de evolutie. In een poging de verschillende perspectieven te overstijgen, onderscheiden we tot slot de meta-ethiek, die feitelijk voortkomt uit de analytische traditie. Deze is gericht op conceptuele analyses en methodologische aspecten van de begrippen en noties die in de descriptieve en normatieve ethiek worden gehanteerd en geformuleerd. Heeft een ethische propositie een waarheidswaarde? Is het een vorm van kennis? Hoe kan die worden gevalideerd? Is de kloof tussen 'is' en 'ought' overbrugbaar? Hoe is het geweten te rechtvaardigen? Is er een universele moraal? Is er vooruitgang in de moraal?

Tegen deze summier geschetste achtergrond wordt de ethische reflectie op specifieke kennisdomeinen, probleemgebieden of beroepspraktijken en de daar vigerende gedragsregels dikwijls beschouwd als een proeve van toegepaste of praktische ethiek. Dat geldt ook voor de wetenschap en dus evenzeer voor statistiek en OR. Het postuleren van deze rechttoe rechtaan dependentie lijkt voor de hand liggend, maar is ontoereikend. In het Westerse denken neemt de statistiek en de daarmee verbonden conceptie van de werkelijkheid een weliswaar dominante, maar tevens problematische positie in, die bovendien middels een moeizaam en slechts ten dele voltooid emancipatieproces tot

stand is gebracht. Dat maakt de relatie tussen ethiek en statistiek veeleer reciproom met wederzijdse uitdagingen en opdrachten.

Zo behoeft het weinig betoog dat de relatie ethiek-statistiek vaak nogal pejoratief wordt geduid. Uitingen van een breed gedragen kritiek en onbehagen zijn manifest en talrijk. Darrell Huffs klassieker *How to lie with statistics* kende zoveel bijval en navolging dat een bescheiden nieuw genre ontstond vol caveats en waarschuwingen. Het aan Benjamin Disraeli toegeschreven, maar ongetwijfeld apocriefe citaat '*There are three kinds of lies: lies, damned lies, and statistics*' vormt evenzeer regelmatig een aanzet tot zwaarmoedige overpeinzingen. Minder vluchtige illustraties van het onbehagen vinden we volop in de media. Statistiek wordt regelmatig gelieerd aan bedrog en fraude, aan belangenverstremgeling of twijfelachtige onafhankelijkheid of aan onwil dan wel onkunde bij het toepassen van statistische methoden, vaak met grote maatschappelijke gevolgen. Frauderende hoogleraren vervalsen data of gebruiken deze selectief uit eigenbelang of ter ondersteuning van hun ideologie. Statistische fouten in de rechtszaal baren niet minder opzien. Twijfelachtige verkiezingspolls worden ingezet als strategisch wapen ter beïnvloeding van het electoraat. Geneesmiddelen waarvan de werking ten onrechte als bewezen wordt aangemerkt, worden grif voorgeschreven. Econometrische voorspellingen gebaseerd op foute modelaannamen en bizarre extrapolaties worden door voor- en tegenstanders van een beleidmaatregel naar eigen inzicht aangewend. De lijst kan eenvoudig worden aangevuld. Ironisch genoeg lijkt in dit publicitair geweld een ontegenzeggelijk ethisch thema als de belangen van proefpersonen en burgers (*informed consent, privacy*) zelfs relatief bescheiden aandacht te krijgen.

Al deze misstanden vallen met geen van vernoemde normatieve kaders te rijmen, maar kunnen door prudentie, een goede gedragscode en dito beroepscommissie, gecombineerd met betere werving en selectie worden beheerst of ondervangen. Een meer wezenlijke kritiek betreft ethische problemen

die intrinsiek zijn aan de statistiek. In hoeverre bevat het probabilistische denken, dat in toenemende mate wetenschap en maatschappij beheerst, uitgangspunten en concepten die op gespannen voet staan met het vigerende mens- en wereldbeeld, de daarmee verbonden waardehiërarchieën en de daarin verankerde ethische kaders en intuïties?

We moeten ons hier beperken tot een tweetal observaties. Allereerst kent, zoals gebruikelijk, ook deze problematiek een genealogie. Dat wordt duidelijk als we te rade gaan bij Ian Hacking, die in *The Taming of Chance* (1990) de opkomst van de moderne statistiek in de 19de eeuw belicht vanuit een historisch-filosofisch perspectief. De auteur laat zien dat uiteenlopende en vooraanstaande intellectuelen, zoals de schrijver Fjodor M. Dostojevski, de filosofen Comte en Nietzsche en de bioloog William Bateson om verschillende redenen bezwaren koesterden jegens de statistiek. Die stond toen nog in de kinderschoenen; Adolphe Quetelets *l'homme moyen* werd een curieuze metafysische abstractie en menig statisticus viel ten prooi aan satirici. Sommigen betoogden dat door de statistiek de mens zijn individualiteit en menselijkheid verliest (Dostojevski), anderen ontzegden, redenerend vanuit een deterministisch wereldbeeld, de probabilistische wetten een zelfstandige status als volwaardige kennis of basis voor oordelen en handelen (Comte). Weer anderen hekelden de biometrische 'oppervlakigheid' en de dreigende teloorgang van de door veel biologen noodzakelijk geachte notie van causaliteit (Bateson). Kans en toeval waren allerminst getemd of aan banden gelegd en de basis voor het huidige onbehagen werd toen gelegd.

De tweede opmerking sluit daarbij aan. Door de probabilistische revolutie verloor het wereldbeeld veel van haar aanschouwelijkheid, de concrete objecten van de wetenschap werden vervangen door (parameters van) kansverdelingen. Data, variatie in data, datagenererende mechanismen coderen en bouwen een werkelijkheid, veeleer dan een (vermeende) fysische werkelijkheid af te beelden. Daarmee werd de statistiek deelachtig aan een belangrijk proces in de

Westerse ideeëngeschiedenis, het groeiende spanningsveld tussen wetenschappelijk wereldbeeld en de intuïtieve concepten en categorieën waarmee de mens de werkelijkheid en de contingenties van zijn bestaan duidt. Paradoxaal genoeg leidde de succesvolle ontwikkeling in de statistiek tot meer ongemakken met de menselijke maat: interpretatie van het kansbegrip, schattingstheorie, significantie testen, hypothese toetsen, betrouwbaarheidsintervallen, etc. Het Bayesianisme kan deels worden beschouwd als een poging tot verzoening.

In de kennisleer is de geschetste transitie al min of meer voltrokken, zij het verre van rimpelloos. Klassieke noties rond waarheid, kennis, betrouwbaarheid, onzekerheid en causaliteit worden thans in probabilistische termen geduid. Bij de moraal is de menselijke maat evenwel nadrukkelijker in het geding. Als de mens zijn uniciteit verliest en verwordt tot een abstractie, een inwisselbare representant van een equivalentieklasse, waarvan de essentiële kwaliteiten worden gereduceerd tot een gewogen som van kwantitatieve variabelen, is dat problematisch. Als ethische categorieën, intuïties en oordelen (acceptabel/onacceptabel, gewenst/ongewenst, rechtvaardig/onrechtvaardig, veiligheid, risico, betrouwbaarheid) en bovenal de taal waarin deze zijn geformuleerd op gespannen voet staan met de probabilistische conceptie van de werkelijkheid en statistische gevolgtrekkingen inzake (wenselijkheid van) menselijk handelen, is dat een heuse aporie. Dat de mens geen *intuitive statistician* is, zoals Kahneman en Tversky rond 1970 experimenteel beargumenteerden, doet daaraan niet af. Dat geldt evenzeer voor bij voorbeeld de juridische bezwaren van L. J. Cohen tegen wat hij de Pascaliaanse traditie noemt. Het tijdperk van Big Data biedt dan ook vele uitdagingen voor de moeizame, maar onverbiddelijke relatie tussen statistiek en ethiek.

RICHARD STARMANS is verbonden aan de Faculteit Bèta-wetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica.
E-mail: <starmans@cs.uu.nl>

MEDISCH—ETHISCHE VRAAGSTUKKEN VAN DE VERLOSKUNDE

Medische ethiek is het vakgebied waarin wordt nagedacht over een zo goed mogelijke uitvoering van de geneeskunde, en waarbij ethische vragen worden gesteld.

Artsen komen dagelijks in aanraking met ethische vraagstukken. Medische ethiek is daarom ook ruim vertegenwoordigd in de geneeskunde opleidingen. Door het

curriculum voor geneeskundestudenten heen, wordt er training gegeven in communicatievaardigheden en attitudevorming, mede gestalte gegeven en onderwezen door ethici.

Al omstreeks 400 jaar voor Christus gaf de Eed van Hippocrates richtlijnen voor het gedrag dat van artsen wordt verwacht. Met toename van de medische kennis en kunde, zowel op diagnostisch als therapeutisch vlak, treden er vaker dilemma's op; is alles wat mogelijk is ook gewenst?

INGRID BRUSSÉ

In de praktijk van het specialisme Verloskunde en Gynaecologie kom je dagelijks in aanraking met medisch ethische vraagstukken. Mag je bijvoorbeeld vrouwen verplicht steriliseren waarvan eerdere kinderen uit huis geplaatst zijn wegens mishandeling of verwaarlozing binnen het gezin? In hoeverre moeten wij als artsen meewerken aan het zwanger maken van vrouwen met geassisteerde voortplantingstechnieken als zij ernstige lichamelijke of geestelijke ziekten hebben? Of als haar (lichamelijke) conditie een verhoogde kans geeft op prematuriteit en morbiditeit van haar nageslacht?

Op het gebied van verloskunde en prenatale diagnostiek zijn er ook velerlei vraagstukken en dilemma's actueel binnen onderwerpen als screening naar aangeboren afwijkingen en prematuriteit.

Eerste trimester-screening naar het syndroom van Down

Leidt het screenen in de zwangerschap niet tot het overmatig medicaliseren van de zwangerschap? Kunnen statistische algoritmes helpen om zwangerschapsuitkomsten te voorspellen? Een zwangerschap is tenslotte een natuurlijk proces? Moeten wij daar wel inbreuk op willen maken?

De discussie rondom het toestaan van het screenen in het eerste trimester van de gehele zwangere populatie op een verhoogd risico op het syndroom van Down (trisomie 21) door middel van

een combinatietest* is nog niet lang geleden. In ongeveer diezelfde tijd (2007) werd de discussie gevoerd of je wel iedere zwangere aan moet bieden om rondom 20 weken zwangerschapsduur de foetus echoscopisch met het Structureel Echoscopisch Onderzoek (SEO) na te kijken op structurele afwijkingen (zoals open rug, hartafwijkingen, open lip en navelbreuk). Het is ondoenlijk om te overzien in wat voor emotionele roller-coaster een zwangere terecht kan komen als er iets afwijkends wordt gevonden. Er kunnen behalve aandoeningen die goed behandelbaar zijn, ook niet met het leven verenigbare aandoeningen worden vastgesteld. Daarnaast zijn er ook beelden zonder eenduidige diagnose of met een onzekere of slechte prognose. Beelden waarbij niet te voorspellen is of het kind levensvatbaar zal zijn of een menswaardig bestaan zal kunnen leiden.

* Bij de combinatietest worden verschillende algoritmes berekend op basis van titers vrij β -hCG (Beta-humaan ChorionGonadotrofine, zwangerschapshormoon) en PAPP-A (pregnancy-associated plasma protein-A, een eiwit), samen met nekplooidikte, de kop-stuitlengte van de foetus, zwangerschap tot stand gekomen met IVF, de duur van de zwangerschap, de leeftijd, gewicht en rookgedrag van de moeder.

Het definiëren van een menswaardig bestaan is al een opgave. Wat de één nog menswaardig vindt, vindt de ander niet. De ene ouder vindt een kind met kans op doofheid en cognitieve problematiek wegens een congenitale cytomegalovirusinfectie absoluut geen reden om een zwangerschap te onderbreken, de ander zal de zwangerschap willen beëindigen omdat het virus de foetus heeft geïnfecteerd.

Eerste trimester-screening naar andere syndromen

Kort geleden was er discussie over het toestaan van screening met de combinatietest naar twee andere chromosoomafwijkingen, namelijk het syndroom van Edwards (trisomie 13) en Patau (trisomie 18). Bij deze beelden hebben de kinderen ernstige aangeboren afwijkingen die gepaard gaan met mentale retardatie. De levensverwachting van deze kinderen is ernstig verkort. Er treden relatief vaak miskramen op, slechts een klein percentage van deze kinderen (ongeveer 5%) wordt levend geboren. Van de levend geboren kinderen met het syndroom van Patau sterft 80% in het eerste levensjaar. Ongeveer 5% van de levendgeborenen wordt ouder dan zes maanden.

In het geval van chromosoomafwijkingen zijn er later in de zwangerschap met echoscopisch onderzoek ernstige afwijkingen zichtbaar en is aanvullend diagnostisch onderzoek geïndiceerd. Dan wordt een vruchtwaterpunctie aangeboden, waarbij het chromosoompatroon van de foetus bepaald kan worden. Verder biedt een verhoogd risico bij screening in het eerste trimester de mogelijkheid om, wanneer het ouderpaar afziet van de mogelijkheid tot invasieve diagnostiek, in het tweede trimester beter en gericht te kijken bij het SEO, waardoor de ontdekking van het syndroom van Edwards en Patau vergroot wordt.

Een aantal argumenten vóór het toestaan van de eerste trimester-screening zijn in overweging genomen. Iedereen heeft het recht op weten. Dit is één van de belangrijkste ethische principes.

Zwangerschapsafbreking

Na gebruik te hebben gemaakt van de eerste trimester-testen heb je de mogelijkheid om een zwangerschap van een kind met het syndroom van Edwards en Patau in een vroeg stadium te beëindigen door een abortus provocatus (zwangerschapsafbreking). Na tweede trimester-zwangerschapsafbrekingen treedt er meer verdriet en post-traumatische stress symptomen op dan na eerste trimester-zwangerschapsafbrekingen (Korenromp 2005). Bovendien is een zwangerschap een emotionele en lichamelijke opgave en hoe langer de zwangerschap bestaan heeft, hoe meer impact dat heeft gehad op het dagelijkse bestaan van de zwangere en haar partner. Wanneer al vroeg in de zwangerschap bekend is dat foetus is aangedaan met het syndroom van Edwards of Patau kan voorkómen worden dat de zwangere door middel van een keizersnede bevalt. Immers, het uitvoeren van de keizersnede is voor een kind met bijna geen levenskansen niet zo zinvol. Een keizersnede is een ingrijpende medische ingreep waarbij de moeder risico's voor haar eigen gezondheid ondervindt.

Bij het afbreken van een zwangerschap met name bij een syndroom van Patau, vermijd je risico's op zwangerschapscomplicaties zoals een pre-eclampsie (in de volksmond een zwangerschapsvergiftiging).

Tegenargumenten van screening

Er kan ook een aantal tegenargumenten voor het aanbieden van screening worden genoemd. Allereerst medicaliseren we met screening de zwangerschap. Als de testuitslag een verhoogd risico aangeeft zal invasief onderzoek worden aangeboden. Het verrichten van een invasieve test als een vruchtwaterpunctie of vlokentest voor een zwangerschap die in het merendeel van de gevallen spontaan zal eindigen in een miskraam. Ongeveer 96% van de zwangerschappen met het syndroom van Patau en 94% van de zwangerschappen met het syndroom van Edwards zullen

vóór de geboorte tot het overlijden van de foetus in de baarmoeder leiden. Het percentage overleving bij levendgeborenen is eveneens klein. Met het SEO, waarbij er naar structurele afwijkingen bij de foetus wordt gezocht, zullen de meeste zwangerschappen met het syndroom van Edwards en Patau worden opgespoord, waardoor het screenen hierop in het eerste trimester overbodig is. Invoering van screening voor de eerste trimester-combinatietest ook voor het syndroom van Edwards en Patau heeft geleid tot een uitgebreidere counseling voor de test zal worden verricht. Er dient duidelijke uitleg gegeven te worden over de betekenis van de mogelijke uitslagen en over de optie van afzien van een invasieve test totdat eventuele afwijkingen op de echo te zien zijn. Uitgebreidere counseling kost meer tijd. De zwangere krijgt extra informatie te verwerken, wat tot onnodige onrust kan leiden.

De argumenten vóór en tegen het screenen op deze afwijkingen zijn serieus overwogen en uiteindelijk heeft de minister van VWS (ministerie van Volksgezondheid, Welzijn en Sport) toestemming gegeven voor het aanbieden van de screening aan alle zwangeren. De effecten die de invoering van de testen (combinatietest op syndroom van Down, Edwards en Patau, maar ook het SEO) heeft gehad worden uitgebreid bestudeerd evenals de testeigenschappen en de detectierates. Maar er is ongetwijfeld geen weg meer terug.

Grenzen van levensvatbaarheid

Een heel ander, maar niet minder belangrijk of minder interessant probleem en dilemma in de verloskunde is het volgende, namelijk het afwegen van het belang van een moeder versus het belang van het ongeboren kind. Een vroege ernstige pre-eclampsie is een ingrijpende en potentieel gevaarlijke complicatie voor een zwangere vrouw. Wanneer deze complicatie optreedt ruim voor de termijn van levensvatbaarheid van het ongeboren kind, zal er met de zwangere en

haar partner gesproken moeten worden over het beëindigen van deze zwangerschap voor het welzijn van de moeder. Wanneer deze complicatie optreedt rondom de termijn van levensvatbaarheid, zijnde rondom een zwangerschapsduur van 24 weken, dan zal de beslissing om de zwangerschap te beëindigen inhouden dat er een zeer ernstig prematuur kind geboren zal worden. Deze prematuur heeft een grote kans op korte en lange termijncomplicaties en overlijden. Het is niet goed van te voren in te schatten hoeveel prematurenproblematiek het kind zal krijgen, maar in ieder geval wordt met het kind het volgende medische probleem geboren. De kans dat het kind overleeft, wordt met toename van de zwangerschapsduur bij geboorte steeds groter en de kans op sterfte en lichamelijke en geestelijke handicaps neemt af. Indien er zich grote complicaties bij het kind voordoen op de intensive care voor neonaten bestaan er ook wel weer mogelijkheden om te stoppen met de behandeling, alhoewel de impact hiervan niet mis te verstaan is.

Ten slotte

De afgelopen jaren is de medische kennis door onderzoek waarbij vaak complexe statistische analyses zijn uitgevoerd toegenomen. Hierdoor zijn echter ethische dilemma's ontstaan die voor zowel (toekomstige) ouders als behandelaars moeilijk te overzien zijn.

LITERATUUR

Korenromp, M. J., Christiaens, G. C., Van den Bout, J., Mulder, E. J., Hunfeld, J. A., Bilardo, C. M., Offermans, J. P. & Visser, G. H. (2005). Long-term psychological consequences of pregnancy termination for fetal abnormality: a cross-sectional study. *Prenatal Diagnosis*, 25(3), 253–260.

INGRID BRUSSÉ is gynaecoloog-perinatoloog en werkzaam op de subafdeling Verloskunde en Prenatale Geneeskunde van de vakgroep Verloskunde en Gynaecologie van het Erasmus MC in Rotterdam. Tevens is zij betrokken bij het medisch onderwijs op de Erasmus Universiteit.
E-mail: <i.brusse@erasmusmc.nl>

STATISTICI MOGEN NIET LIEGEN

beroepsethiek onder het vergrootglas

WILLEM DE VRIES

Recent is beroepsethiek veel in het nieuws geweest. Neem Tulleken, neem het VU Ziekenhuis: gezamenlijk (*bien étonnés*) afgebrand door Youp van 't Hek in zijn column in de *NRC* van 25 februari 2012. Jannetje Koelewijn had voor diezelfde *NRC* een artikel geschreven over de medische conditie van prins Friso. De informatie had Koelewijn verkregen via haar man, neurochirurg Kees Tulleken. Jannetje en Kees waren toevallig in Innsbruck en kwamen daar in gesprek met een behandelend arts van de prins. Jannetje publiceerde vervolgens haar bevindingen. Er was veel kritiek op het handelen van zowel Tulleken als de krant. De neurochirurg zou zijn medisch beroepsgeheim hebben geschonden en *NRC* de privacy van Friso. De hoofdredacteur van *NRC* ging door het stof.

Kort later raakte het VU Medisch Centrum (VUmc) in opspraak. Dat had toestemming gegeven aan het televisiebedrijf Eyeworks om via 35 verborgen camera's patiënten op de afdeling spoedeisende hulp te bespieden. Eyeworks-baas Reinout Oerlemans en de directie van VUmc hielden aanvankelijk vol dat het een keurig 'project' was, maar nadat specialisten op het gebied van medische ethiek er schande van spraken in de media en patiënten een rechtszaak tegen de VU begonnen krabbelde men terug. Het was weliswaar

een goed 'project', zei VUmc-directeur Elmer Mulder, maar het was helaas niet correct uitgevoerd. RTL zond één door Eyeworks gemaakte aflevering van het programma uit en daarna werd het 'project' afgeblazen.

VUmc blijkt van 'projecten' te houden. Schrijver Ronald Giphart betreurde de ophef over het project Schrijvers op de Afdeling van het VUmc. Volgens Giphart had niemand in het ziekenhuis problemen met zijn aanwezigheid. 'Ik loop daar niet stiekem of undercover rond.' Bij het project lopen auteurs mee in het ziekenhuis ter inspiratie voor een roman. Na de Eyeworks-affaire besloot het VUmc om het project te stoppen. Giphart liep mee op de afdeling cardiologie. Het VUmc maakte vooraf afspraken over geheimhouding met de schrijvers. Patiënten en bezoekers mochten niet herkenbaar zijn in een eventuele roman. De schrijver kreeg geen inzage in de medische dossiers van patiënten. Giphart droeg bij het meelopen een witte jas, met een badge met 'stagiair' erop.

Ethiek is niet zwart-wit

Dat ethiek soms evenwichtskunst is, bleek uit een andere candid-camera-zaak. De journalisten Jelle Visser en

Jan Ponsen interviewden in 2009 oorlogsmisdadiger Heinrich Boere met een verborgen camera. Ze werden verdacht van het schenden 'van de vertrouwelijkheid van het woord'. In 2010 sleepte Boere de twee verslaggevers voor de Raad voor de Journalistiek. Die gaf de journalisten gelijk. Daarna kwam de zaak voor de Duitse rechter. Die oordeelde dat de aantasting van de privacy van Broere niet opwoog tegen het belang van vrije nieuwsgaring: de verborgen camera is een middel dat zorgvuldig moet worden ingezet, maar kan cruciaal zijn om 'misstanden aan de kaak te stellen'. Zou misschien, als Eyeworks het oogmerk had gehad aan te tonen dat het VUmc een zootje was, Oerlemans nu een held zijn in plaats van een gemankeerde voyeur?

Vanwege het privacy-aspect zijn de bovengenoemde affaires uitgebreid in het nieuws geweest. Googelen van Tulleken, Eyeworks of Boere levert duizenden hits op. Veel minder aandacht kregen enkele recente affaires op het gebied van statistische ethiek. Het is zelfs lastig om er op internet informatie over te vinden, hoewel het om ernstige misstanden gaat.

Zoals voor andere beroepsgroepen bestaan er ook beroepscode's voor statistici. Naast twee internationale codes zijn er verschillende nationale, bijvoorbeeld die van de Royal Statistical Society en de American Statistical Association. De meest universele code is de Declaration on Professional Ethics (DPE) van het International Statistical Institute (ISI). De nieuwste versie werd in juli 2010 vastgesteld door de ISI Council in Reykjavik. Tevens werd een Advisory Board on Ethics ingesteld om het Executive Committee en de Council te adviseren over ethische kwesties.

Een tweede belangrijke code op het gebied van de statistiek zijn de Fundamental Principles of Official Statistics (FPOS), aangenomen door de United Nations Statistical Commission in 1994. De twee genoemde codes zijn verschillend van aard. De DPE richt zich tot individuele statistici, terwijl de FPOS is gericht aan regeringen. Een ander verschil is dat de DPE tamelijk gedetailleerd is, terwijl de FPOS slechts tien artikelen telt en daarom ook de Tien Geboden van de overheidsstatistiek wordt genoemd. Maar de

geschiedenis heeft uitgewezen dat goed en kwaad niet in Tien Geboden te vangen zijn. Mozes kwam met een overzichtelijke boodschap van de berg af. Daarna hebben duizenden moralisten samen bibliotheken volgeschreven over wat de boodschap precies betekende. Op kleine schaal is iets dergelijks gebeurd met de FPOS. De Europese Unie heeft bijvoorbeeld een eigen, uitgebreide versie van de FPOS gemaakt: de European Statistics Code of Practice. Het IMF heeft de FPOS verwerkt in zijn Data Quality Assessment Framework (DQAF).

De Argentijnse kwestie

Het is niet altijd duidelijk of statistici van de overheid zich aan alle regels houden. Dat is ook bijna niet te controleren. Nationale Rekeningen zijn bijvoorbeeld een zó ingewikkeld statistisch product dat het zelfs voor ingewijden erg moeilijk is om te beoordelen of het System of National Accounts (de internationale standaard) naar letter en geest is gevolgd. De EU heeft een heel inspectie-systeem in het leven geroepen om de Nationale Rekeningen van lidstaten te controleren. Nationale Rekeningen leveren onder meer het Bruto Nationaal Inkomen (BNI) op, een van de bronnen van de zogenaamde 'eigen middelen' van de EU.

De 'zaak Argentinië' is eigenlijk eenvoudig. De overheid probeerde inflatiecijfers lager voor te stellen dan ze waren. De internationale economische pers kreeg daar uiteraard lucht van. Toen sloeg de Argentijnse regering door. Statistici die niet wilden meewerken aan het vervalsen van de cijfers werden ontslagen, bedreigd en beboet. De American Statistical Association (ASA) nam het initiatief om deze duistere praktijken aan de kaak te stellen. ASA schreef op 11 augustus 2011 een brief (begeleid door tal van bewijsstukken) aan Frank La Rue, *special UN Rapporteur on the promotion and protection of the right to freedom of opinion and expression*. 'This letter is written to draw your attention to the continuing and escalating attacks by the Government of Argentina directed against a group

of statisticians and allied professionals [...] These attacks, beginning in February 2011, first took the form of levying of confiscatory fines (500,000 Argentine pesos, about US\$125,000) based on a law that originated during the military dictatorship of Reynaldo Bignone (1982-1983), and more recently have escalated to multiple fines and, reportedly, to threats of imprisonment under the criminal law.'

De aanvallen waren in feite al in 2007 begonnen. De Argentijnse inflatie was in die tijd erg hoog en in antwoord daarop 'reorganiseerde' de regering het werk van het Argentijnse CBS (INDEC), ontsloeg medewerkers die zich niet wilden schikken, en begon onwaarschijnlijk lage inflatiecijfers te publiceren. Gebruikers schakelden vervolgens consultancy-bedrijven en onderzoeksinstellingen in om aan betere cijfers te komen. Ook die bedrijven en instellingen kwamen onder vuur van de regering te liggen. ASA wijst erop dat Argentinië hiermee zondigt tegen internationale verdragen die het heeft geratificeerd. De algemene geloofwaardigheid van INDEC is aangetast en dus worden ook de andere statistieken van het bureau niet langer als betrouwbaar gezien – al zijn ze dat misschien wel. 'Statistics - A matter of trust', zo luidt de titel van een 'consultation document' van de Britse regering uit 1998.

Griekse methodes

In de Europese Unie is statistiek sinds het Verdrag van Maastricht (1993) steeds belangrijker geworden. Dat Griekenland met de cijfers knoeide wisten of vermoedden insiders al lang. Net zoals de Griekse belastingdienst en andere overheidsdiensten slecht functioneren, was dat met de statistiek ook zo. Nadat was besloten Hellenic Statistical Authority (Elstat) – het Griekse CBS – te verzelfstandigen ging het beter.

Door de recente crisis dreigt echter weer achteruitgang. Ook de Griekse overheid wordt nu beschuldigd van intimidatie van statistici die de waarheid

willen vertellen. De president van het ISI, Jae C. Lee (Zuid-Korea) heeft daarom (5 januari 2012) een brief gestuurd aan Andreas Georgiou, President van Elstat, om hem te steunen. Lee schrijft onder meer: '*It is with great concern that we have followed the attacks on the Hellenic Statistical Authority and the threats that you as President of Elstat, your deputy and the head of National Accounts are facing.*' Hij wijst op de ISI Declaration of Professional Ethics, waarvan *pursuing objectivity* een hoeksteen is.

Een afschrift van de brief is gestuurd aan de Premier van Griekenland en aan de Voorzitter van het Griekse Parlement. De afloop is op het moment van schrijven onbekend.

Tot slot

Statistische ethiek kan ingewikkeld zijn. Ik ben bij veel nationale statistische bureaus over de vloer geweest, de laatste jaren vooral in ontwikkelingslanden, en het is vaak geen kunst de vinger te leggen op zondes tegen de *Fundamental Principles*. Ik heb het dan niet over ernstige zaken als in Argentinië en Griekenland, maar over pekelzondes.

Er kunnen omstandigheden zijn waarin 'het landsbelang' boven 'de waarheid' of 'de ethiek' gaat. Naar verluidt heeft het Nederlandse CBS tijdens de oliecrisis van 1973 bijvoorbeeld bepaalde cijfers over buitenlandse handel onderdrukt. Veel landen publiceren geen of incomplete cijfers over de buitenlandse handel in strategische goederen. En in de Verenigde Staten zijn na Pearl Harbour door het Census Bureau volkstellinggegevens beschikbaar gesteld om Japanse Amerikanen efficiënt te kunnen interneren. Dit is lang ontkend, maar in 2007 definitief bewezen.

WILLEM DE VRIES (1942) was plaatsvervangend directeur-generaal van het CBS, Deputy Director van de United Nations Statistics Division (New York), en President van de International Association of Official Statistics (sectie van ISI). Hij is lid van de Advisory Board on Ethics van het ISI. E-mail: <wfmdevries@yahoo.com>

Deze bijdrage begint met persoonlijke ervaringen met OR-toepassingen die ethische vragen kunnen oproepen, zoals nucleair afval, waterbeheer, duurzame ecologie, militaire tactiek en dierenwelzijn. Al deze toepassingen gebruiken simulatiemodellen. Zulke modellen zijn bedoeld voor het oplossen van praktische problemen, en die problemen hebben ethische implicaties voor de diverse 'stakeholders', zoals de OR-deskundigen, hun klanten en de samenleving. Verder ook een overzicht van ethische gedragsregels in een aantal wetenschappelijke disciplines verwant aan OR. De auteur benadrukt de validatie van modellen, vooral documentatie van de vooronderstellingen van die modellen. Sommige ethische normen en waarden kunnen worden gekwantificeerd door (multipale) modeluitkomsten. Ethische vragen kunnen vooral urgent zijn in militaire modellen, inclusief 'war games'. Computerspellen en de gerelateerde experimenten in de Experimentele Economie kunnen ook een middel zijn voor de bestudering van ethische vragen.

ETHISCHE VRAGEN IN OPERATIONS RESEARCH

persoonlijke overpeinzingen

JACK P. C. KLEIJNEN

Mijn verhaal is gebaseerd op diverse OR-projecten waaraan ik heb meegewerkt die – achteraf gezien – ethische vragen inhielden. Een voorbeeld is het project Waste Isolation Pilot Plant (WIPP) in Carlsbad in New Mexico. Deze WIPP is een ondergrondse opslagplaats voor nucleair afval dat resteert na bepaalde behandelingen van kankerpatiënten. De opdrachtgever van

dit project was het Environmental Protection Agency van het Department of Energy. Het simulatiemodel voor het schatten van de kans op lekkage van dit afval vanuit de ondergrondse opslagplaats naar de oppervlakte is ontwikkeld door een team deskundigen van Sandia National Laboratories in Albuquerque. Andere belanghebbenden zijn de mensen die door dat afval

besmet kunnen worden in de komende 10.000 jaar: een duidelijk ethisch probleem! Daarnaast geniet de lokale bevolking meteen van economische voordelen zoals werkgelegenheid. Samenvattend, kankerpatiënten hebben kans op genezing; de mensen in Carlsbad en elders hebben kans op huidige en toekomstige gezondheidsproblemen én economische voordelen. Het simulatiemodel kwantificeert alleen de kans op lekkage van het afval naar de oppervlakte. Mede op basis van de uitkomsten van dit model heeft de overheid toestemming gegeven om de WIPP te bouwen. Sandia heeft veel gepubliceerd over projecten voor de opslag van nucleair afval; zie de recente publicatie Helton en Pilch (2011). Dit artikel geeft alleen additionele referenties; veel meer referenties zijn te vinden in Kleijnen (2011).

Een ander 'persoonlijk' voorbeeld is de stormvloedkering in de Westerschelde. Tegenwoordig is een centrale vraag in watermanagement het afwegen van de belangen van *stakeholders* zoals vissers, lokale bewoners, toeristen en de ecologie.

Nog een voorbeeld betreft de mijnenjacht door de Koninklijke Marine (de klant) met behulp van sonar. FEL/TNO bouwde het simulatiemodel. Mijn inbreng betrof de validatie van dat model. Zoals bij elk militair model kunnen ethische vragen gesteld worden; immers een belangrijk doel van militairen is het doden van vijanden.

Een ander voorbeeld betreft het adviseren van het RIVM bij problemen die ontstaan door de opwarming van de aarde. Deze problemen zijn tegenwoordig urgent en wereldwijd, en hebben ethische implicaties voor huidige en toekomstige generaties. De klimaatwetenschappers worstelen met de validatie van de verscheidene klimaatmodellen; daarnaast dienen die wetenschappers de belangrijkste factoren op te sporen, zodat de overheid realistische maatregelen kan bedenken.

Een laatste voorbeeld betreft robots die koeien melken. Een praktisch probleem is te bepalen, hoeveel robots op een gegeven boerderij nodig zijn. Een ethische vraag betreft het dierenwelzijn (dat overigens niet in het model zat).

Ethische gedragscodes in diverse wetenschappelijke disciplines

OR-organisaties (inclusief de VvS+OR) hebben geen eigen gedragsregels. Toch zijn er wel diverse publicaties over OR-gedragsregels (zie Kleijnen, 2011). Verwante organisaties hebben wel gedragsregels: American Statistical Association, Association for Computing Machinery en Society for Computer Simulation. Ethische gedragsregels voor ingenieurs worden uitvoerig besproken in Wikipedia <http://en.wikipedia.org/wiki/Engineering_ethics>; zie ook de gedragscode van KIVI NIRIA (Koninklijk Instituut Van Ingenieurs) <www.kiviniria.net/a/PAG000002106/Gedragscode.html> en bij Kleijnen (2011). Andere disciplines met zulke regels zijn de rechtswetenschap, psychologie, sociologie en de medische wetenschap. De huidige financiële crisis leidt tot debatten over de noodzaak van ethische regels voor financiële analisten; zie de MBA Ethics Oath <<http://mbaoath.org>>.

Modellen: documentatie van vooronderstellingen

Een OR-model betreft een systeem in the reële wereld; het doel van dit model is een probleem in die wereld op te lossen en dat doel heeft vaak ethische implicaties voor de stakeholders. Elk model berust op specifieke vooronderstellingen (bv. lineariteit, normaliteit). Daardoor zullen de uitkomsten geldig zijn als die veronderstellingen opgaan. Dus rijst de cruciale vraag: wat gebeurt er als die veronderstellingen niet opgaan? Die vraag blijft vaak onbeantwoord, bijvoorbeeld doordat de modeluitkomst de klant goed bevalt. Maar *Garbage In, Garbage Out* (GIGO)! Een recent voorbeeld is de discussie betreffende klimaatmodellen: zijn die modellen realistisch? Als lid van diverse panels over validatie concludeer ik dat de belangstelling voor de validatie van modellen groter is in het publieke domein dan in het bedrijfsleven; zie ook de recente publicatie van Simulation Interoperability Standards Organization (2012).

Modellen kunnen worden gebruikt op goede of slechte wijze, door de modelbouwers of hun klanten – en de samenleving kan de dupe worden. Die klanten en de samenleving begrijpen wellicht niet welke causaliteit het model bevat; ze hebben bijvoorbeeld niet de ‘handleiding’ (de bijsluiter) gelezen. De documentatie van het model zou de onderliggende redeneringen moeten uitleggen, inclusief de criteria (uitkomsten) en de veronderstellingen met hun validatie. Ik herinner me nog steeds met plezier, hoe IBM voor voorraadbeheer via het pakket IMPACT een expliciete lijst met veronderstellingen gaf, voor verschillende gebruikersgroepen (het dunste boekje was voor het management en het dikste voor de programmeurs).

Documentatie is ook noodzakelijk om andere onderzoekers in staat te stellen de wetenschappelijke uitkomsten te reproduceren. Zulke reproductie – of de antithese, falsificatie – is een basisprincipe in de wetenschap. De affaire Stapel is een recent wrang voorbeeld.

Het lijkt me een interessante uitdaging om online modeldocumentatie voor de computer te ontwikkelen betreffende criteria, veronderstellingen en validatie. Die documentatie moet dan via de help-functie op te roepen zijn. Veel simulatiemodellen geven nu wel al documentatie via computeranimatie die het gesimuleerde systeem aan de gebruikers uitlegt. Maar die animatie kan ook een misleidende validatietechniek zijn, doordat de animatie een te korte periode toont. Ik voeg toe dat terwijl de klanten het model gebruiken, ‘rode’ waarschuwingslichtjes dienen te gaan branden als de klanten gegevens invoeren buiten het toegestane (geldigheids)gebied. Verder dient een model regelmatig terug te gaan naar de bouwers om bijgewerkt te worden. Deze problemen worden nog belangrijker als de modelbouwers niet weten wie de gebruikers zullen zijn, zoals in het geval van veel belanghebbenden.

Normen en waarden, en model criteria

‘Normen en waarden’ is een gevleugelde uitdrukking in de Nederlandse politiek. In wiskundige model-

len gaat het om de normen en waarden van klanten, modelbouwers, en andere belanghebbenden. Een persoonlijk voorbeeld betreft de simulatie van de financiële gevolgen van veranderingen in bepaalde sociale zekerheidswetten – voor de overheid die geld wil besparen en voor individuele werknemers die niet allen even hard worden getroffen.

In de praktijk geven simulatiemodellen meestal uitkomsten voor meerdere criteria, terwijl andere soorten modellen vaak beperkt blijven tot een enkel criterium. Die simulatie kan dus voor meerdere belanghebbenden waarden kwantificeren. Simulatiemodellen berekenen geen optimale oplossing (maar zijn wel realistischer dan vele andere modellen). Wel berekent een simulatiemodel meerdere criteria voor verschillende scenario’s betreffende verschillende veronderstelde toekomsten en maatregelen. De gebruikers kunnen vervolgens hun favoriete oplossing kiezen, rekening houdend met normen en waarden.

Militaire modellen, computerspelletjes en ‘experimental economics’

Terrorisme is tegenwoordig een probleem dat in veel modellen wordt bestudeerd. Maar wat terrorisme is, hangt af van normen en waarden: is een zelfmoordenaar een martelaar of een misdadiger? Niet alle wetenschappers zijn bereid voor Defensie te werken (pikant is dat de oorsprong van OR de bestrijding is van Duitse onderzeeboten tijdens de Tweede Wereldoorlog). Persoonlijk denk ik dat – in het algemeen gesproken – modellering voor Defensie moreel acceptabel is; maar er zijn uitzonderingen, namelijk het ontwikkelen van bepaalde wapens (bijv. gifgas?). Een hedendaags voorbeeld zijn drones, die het voeren van oorlog doen lijken op een computerspelletje.

Deze drones leiden mij naar *serious gaming*: de spelers nemen beslissingen die ingevoerd worden in een simulatiemodel, waarna de computer uitrekent wat de consequenties van deze beslissingen zijn. Deze gaming is dus een methode voor het bestuderen

van menselijk gedrag inclusief ethische aspecten; bijvoorbeeld gaan de spelers voor hun persoonlijke volle winst of gaan ze samenspannen tegen de consumenten? Naast deze bedrijfsspelen zijn er ook *war games* en andere vormen van gaming. (Momenteel ben ik betrokken bij een project van het Nationaal Lucht- en Ruimtevaartlaboratorium NLR waarin de effecten van verkeersmaatregelen worden bestudeerd).

Verwant aan *serious gaming* is *experimental economics* (EE). EE gebruikt heel eenvoudige economische modellen, maar de spelers ontvangen wel – per ronde – echt geld (zij het kleine bedragen). EE bestudeert bijvoorbeeld altruïstisch tegenover egoïstisch gedrag, rationeel tegenover emotioneel gedrag.

Agent Based Simulation (ABS) gebruikt geen menselijke spelers, maar geautomatiseerde *agents*. Deze *agents* voegen dus nog een abstractieniveau toe aan het model, zodat validatie problematischer wordt – met ethische consequenties. (Ik heb bijgedragen aan de toepassing van ABS bij bedrijfseconomische en militaire problemen.)

Samenvattend is gaming een populaire methode voor de bestudering van menselijk gedrag; gaming is nog niet populair voor de bestudering van ethische vragen in OR.

Klokkenluiders

‘Wiens brood men eet, diens woord men spreekt’, maar klokkenluiders doen dat dus toch niet! In Nederland hebben enkele klokkenluiders hun baan verloren. Een ander voorbeeld betreft een geval in 1999, toen Kamerleden vragen stelden over de toestemming voor de uitbreiding van Schiphol, nadat een RIVM-medewerker beweerde dat deze toestemming gebaseerd was op een verkeerd model en niet op echte metingen van vliegtuiglawaai en luchtvervuiling op Schiphol (ik verwijs ook naar de toestemming voor de WIPP in New Mexico, waar echte metingen niet mogelijk waren).

In het algemeen hebben echte metingen als nadeel dat ze veel tijd en geld kosten; bovendien kunnen dan

maar een paar scenario's gemeten worden en ze kunnen gevaarlijk zijn. Metingen via modellen eisen validatie, gevoeligheids-, onzekerheids- en robuustheidsanalyses. Ik merk nog op dat in geval er meerdere belanghebbenden zijn, de financiële kosten en baten gealloceerd kunnen worden via speltheorie (bijv. Nash evenwicht).

Conclusies

Ethische vragen in OR zijn essentieel voor elke bouwer van een praktisch OR-model, omdat die bouwer als mens morele problemen onder ogen dient te zien! Toch zijn die ethische vragen nog steeds geen deel van de standaard academische OR-curricula en de gedragscode van onze VVS. Nu en dan komen deze vragen aan de orde in de populaire media – bijvoorbeeld door klokkenluiders – maar dan worden die vragen niet wetenschappelijk bestudeerd. Er zijn te weinig specialisten in het interdisciplinaire gebied gevormd door ethiek en OR. In mijn bijdrage heb ik persoonlijke overpeinzingen gepresenteerd, gebaseerd op mijn ervaring met praktische OR-studies. Ik hoop dat mijn bijdrage een stimulans is voor verdere discussie van ethiek en OR.

Deze bijdrage is een Nederlandstalige bewerking van Kleijnen (2011).

LITERATUUR

- Helton, J. C. & Pilch, M. (2011). Guest editorial: Quantification of margins and uncertainty. *Reliability Engineering and System Safety*, 96, 959–964
- Kleijnen, J. P. C. (2011). Ethical issues in engineering models: an operations researcher's reflections. *Science and Engineering Ethics*, 17(3), 539–552
- Simulation Interoperability Standards Organization (2012). *Generic methodology for verification and validation (GM-VV) to support acceptance of models, simulations and data. GM-VV Vol. 1: Introduction and overview*. SISO-GUIDE-001.1-2012. Orlando: SISO

JACK KLEIJNEN is hoogleraar Simulatie en Informatie-systemen aan de Universiteit van Tilburg. Hij is lid van het Departement Informatiekunde en de Operations Research groep van CentER van de Tilburg School of Economics and Management (TiSEM). E-mail: <kleijnen@tilburguniversity.edu>

ETHIEK EN OR

ALEXANDER RINNOOY KAN

Het spijt mij voor alle belanghebbenden, maar de gedachte dat de Operations Research een ethische dimensie zou kennen lijkt mij te berusten op een fundamenteel misverstand. Voor zuivere wiskunde geldt dat net zo. Algebra en ethiek, meetkunde en moraal, het zijn combinaties waar niemand zich iets bij kan voorstellen. Waarom zou dat voor toegepaste wiskunde anders zijn?

Het enig mogelijke antwoord op deze vraag zou moeten voortvloeien uit de toepasbaarheid van die wiskunde. Ja, natuurlijk kan die toepassing een ethisch verwerpelijk doel dienen. Lineaire programmering kan ingezet worden om mensenlevens te redden, maar ook om ze te vernietigen. Wachtrijtheorie kan de toegankelijkheid van de eerste hulp ten goede komen, maar ook bevorderen dat boodschappeninpakkers daar met regelmaat belanden. Maar met de OR zelf heeft dat niets te maken. Het wiskundige model is ethisch neutraal. Als het in slechte handen valt, dan is dat het model niet aan te rekenen. Er zit niets in het wiskundig model dat het vermijdbaar zou kunnen maken. Er zit ook niets in het model dat het onvermijdelijk maakt.

Zijn er dan geen immorele toepassers van de OR? Natuurlijk. Er zijn ook algebraïci die hun kinderen slaan en meetkundigen die er met de kas vandoor gaan. Maar de ethische overtredingen die OR-toepassers plegen bij de uitoefening van hun vak hangen samen met het terrein van de toepassing en niet met de aard van het wiskundig model. Wie geheeltallige optimalisering gebruikt om patiënten zo gauw mogelijk het ziekenhuis uit te duwen, krijgt (misschien) problemen met de Eed van Hippocrates. Wie het gebruikt om

gevangeniszellen vol te proppen met delinquenten, handelt (misschien) in strijd met de rechten van de mens. In beide gevallen kan een passende reflectie op de onderliggende ethische vraagstelling heel verhelderend zijn. Maar de begeleiding daarvan is bij de geheeltallige-optimaliseringsspecialisten niet per se in de beste handen.

De belangstelling voor de combinatie van ethiek en OR weerspiegelt een kennelijke bezorgdheid die interessanter is dan de combinatie zelf. Als er reden is te vermoeden dat OR toepassers met enige regelmaat in lastig ethisch vaarwater belanden, dan kan er een goede reden zijn in kaart te brengen waar en wanneer en waarom dat gebeurt. Dat zou een vervolg kunnen krijgen in het curriculum van de opleiding, in de vorm van een basistraining praktische ethiek dan wel van een gespecialiseerde cursus die aansluit bij de professionele moraal in een of ander toepassingsgebied. Het kan zeker geen kwaad om bij vlagen wat wereldvreemde wiskundigen te wennen aan het idee dat de wereld waarin zij zullen belanden bij de uitoefening van hun vak rijkelijk voorzien is van grijze gebieden waar de keus tussen goed en fout niet zo eenvoudig te maken is. Ethiek voor OR toepassers is zo gezien een relevante uitdaging. Maar de ethiek van Markov ketens en polytopen bestaat niet.

ALEXANDER RINNOOY KAN is universiteitshoogleraar Economie en Bedrijfskunde aan de Universiteit van Amsterdam. Hij is wiskundige en econometrist die na een loopbaan in de wetenschap in 1991 voorzitter werd van het VNO (later VNO/NCW). Werd in 1996 bestuurder van de ING Groep. Was van 2006-2012 SER-voorzitter. E-mail: < a.h.g.rinnooykan@uva.nl >

DE WERELD DER WAARDEN

statistiek, evolutie en ethiek

RICHARD STARMANS

Een kritische reflectie op de verwevenheid van statistiek en ethiek behelst een wijsgerige exercitie die vanuit diverse perspectieven een aanvang kan nemen. We zullen hier niet afdalen naar de spelonken van de moraliteit, maar het thema vooralsnog beschouwen als een verbijzondering van een aloude problematiek uit de wetenschapsfilosofie, de vraag of liever gezegd de queeste naar een normatieve fundering van de kennis. Is wetenschap *sui generis* waarde vrij, universeel en cultuuronafhankelijk of dankt zij haar betekenis aan een welbepaalde plaats in een waardehiërarchie, dan wel een netwerk of 'web van waarden' en is zij derhalve fundamenteel niet waarde vrij, contingent en cultureel bepaald? Op een 'metaniveau' dient zich

dan een analogo dilemma aan. Beschouwen wij de ideeëngeschiedenis internalistisch, als opeenvolging van methoden, technieken, experimenten, theorieën, stellingen of modellen? Of proberen we haar externalistisch te duiden vanuit de gesitueerdheid van een maatschappelijke en dus tevens ethisch, levensbeschouwelijk en politiek geïnspireerde context?

Relatie statistiek en biologie

Van oudsher heeft dit spanningsveld een rijk geschaakt palet aan wijsgerige posities en polemieken opgeleverd. Opvallend is daarbij dat met name analy-

tische filosofen nogal eens de stelling verdedigen dat de onderhavige problematiek vooral de empirische wetenschappen en technische disciplines/ingenieurswetenschappen betreft, maar dat wiskunde en logica als zuiver conceptuele en formele wetenschappen zich er grotendeels aan onttrekken. De vraag dringt zich op of dit ook geldt voor de statistiek als wiskundige c.q. in de wiskunde verankerde discipline, die evenwel steevast in wisselwerking met diverse vakdisciplines wordt beoefend en zelfs op die manier tot ontwikkeling is gekomen. We beperken ons hier tot één saillant aspect van deze problematiek: de historisch cruciale relatie tussen statistiek en biologie. Beide vooronderstellen een conceptie van de werkelijkheid waarin variatie en verandering centraal staan. Deze verwante noties kenden van oudsher een pejoratieve duiding in de Westerse ideeëngeschiedenis, maar doorliepen een emancipatieproces dat in de 19e eeuw een voltooiing bereikte. Toen vond de door Ian Hacking eloquent getypeerde ‘erosie van het determinisme’ plaats, tekende zich een historisering van het wereldbeeld af en beargumenteerde Friedrich Nietzsche dat ook de moraal een genealogie kent.

De opkomst van de evolutietheorie van Wallace en Darwin en later die van de mendeliaanse genetica vormden een hoogtepunt van dit emancipatieproces en tevens een katalysator voor de interactie tussen biologie en statistiek. De kruisbestuiving tussen beide eind 19e eeuw is zonder weerga in de wetenschapsgeschiedenis, maar oversteeg nadrukkelijk de grenzen van de ‘zuivere’ wetenschap. Evolutietheorie en erfelijkheidsleer zetten aan tot vele debatten over de plaats van de mens in de kosmos, eugenetica, sociaaldarwinisme, de grondslagen van de moraal en de zin van een evolutionaire ethiek. Dit alles was allerminst louter voer voor filosofen. Ideologische en ethische denkbeelden en vooronderstellingen inspireerden en beheersten ook het werk van belangrijke pioniers van de statistiek. De biometrici Francis Galton (1822-1911) en Karl Pearson (1857-1936) en later de mendeliaan Ronald Fisher (1890-1962) droegen actief bij aan de ontwikkeling van de biologie als prille wetenschappe-

lijke discipline, maar kregen evenzeer te maken met de voetangels en klemmen van een evolutionaire ethiek.

Ethiek en evolutie

Het verschijnen van Darwins *The Origin of Species* in 1859 leidde al spoedig tot wijsgerig-ethische beschouwingen, ook onder vooraanstaande evolutiedenkers als Wallace, Huxley en Spencer, die overigens onderling diepgaand van mening verschilden over de reikwijdte van de evolutieleer en de implicaties ervan voor de ethiek. Zelf ging Darwin pas in zijn *The Descent of Man* (1871) in op de consequenties van zijn theorie voor de menselijke moraal. Daarbij zocht hij overigens niet zozeer een alternatieve rechtvaardiging van de moraal, maar wilde hij slechts aantonen dat zijn natuurwetenschappelijke theorie een betere verklaring voor uitingen van moreel gedrag bood dan vigerende ethische kaders als de plichtethiek van Immanuel Kant of het utilitarisme van John Stuart Mill.

De Britse ethicus en ingenieur Herbert Spencer (1820-1903) koesterde die eerste ambitie heel nadrukkelijk wel en liet deze gepaard gaan met een sterk moreel optimisme. In het voorwoord van zijn *Principles of Ethics* uit 1879 manifesteert hij zich als exponent van het 19e-eeuwse vooruitgangsgeloof en sciëntisme en betoogt hij zelfs dat de evolutietheorie de bestaande ethische kaders en theorieën onder druk zet en ons noopt tot een nieuwe fundering van de moraal. Spencer beschouwt het concept van evolutie als een universeel, constituerend en welhaast kosmisch principe, dat ertoe zal leiden dat de mensheid een verregaande morele ontwikkeling doorloopt, die uitmondt in een soort eudaimonische voltooiing, waarin altruïsme en egoïsme samenvallen, en waarin de gezochte rechtvaardiging van een evolutionaire ethiek kan worden gevonden. Dit uitgangspunt, gecombineerd met het sciëntistische streven de gehele maatschappij op wetenschappelijke leest te schoeien leidt tot een bijpassende handelingsimperatief, een morele verplichting samenleving en mens-

heid te 'verbeteren'. Het optimisme van Spencer werd door weinigen gedeeld en ook de geschiedenis van de eerste helft van de 20e eeuw toont vooral de keerzijde van de medaille.

Statistiek en evolutie

De triptiek statistiek-biologie-ethiek en de daarmee verbonden grandeur en misère werd dan ook zichtbaar in het werk van Galton, Pearson en Fisher, die allen werkten op het snijvlak van biologie en statistiek. Zij waren ontegenzeggelijk evolutiedenkers en werkten vooral met biologische data. In het *nature versus nurture*-debat namen zij begrijpelijkerwijs stelling ten faveure van *nature*, waarbij het primaat wordt toegekend aan erfelijkheid en evolutie ten koste van de invloed van omgevingsfactoren als opvoeding en milieu. Dit gebeurde deels op ideologische gronden, zonder een helder inzicht in het onderliggende genetische mechanisme. Mendels 'statistische' genetica was slechts in zeer kleine kring bekend en zou pas in het begin van de 20e eeuw worden (her)ontdekt en tot de Nieuwe Synthese van evolutietheorie en genetica leiden. Ook Spencer beriep zich op lamarckiaanse ideeën over overerving van aangeleerde of verworven eigenschappen, hetgeen feitelijk niet met de evolutieleer strookt. Zelfs Darwin greep er af en toe op terug.

De versmelting met ideologie en ethiek kende echter nog verder reikende uitingen. Francis Galton, een volle neef van Darwin, was bovenal een onvermoeibaar pleitbezorger van toepassing van wiskundige methoden in de biologie. De evolutieleer versterkte hem in zijn overtuiging dat variatie en verandering intrinsiek zijn in de natuur, volwaardige concepten vormen en een wiskundige behandeling van de biologie allerminst in de weg staan. Dat bleek bij zijn interpretatie van de normale verdeling. Zijn voorgangers Laplace en Quetelet hadden deze verder ontwikkeld, maar verraadden tevens een pejoratieve duiding van de noties van variatie en verandering. De determinist Laplace gebruikte de verdeling primair in zijn

foutentheorie om 'afwijkingen' te verklaren. Quetelet ontwikkelde zijn concept van de *l'homme moyen*, een metafysische abstractie waarbij de parameter van het gemiddelde een soort essentie vormde en de spreiding als afgeleid of deviant werd opgevat. Galton, die dankzij de analyse van grote aantallen biometrische data het nivellerende principe van *regression to the mean* op het spoor was gekomen, moest niets hebben van Quetelets preoccupatie met het gemiddelde, dat hij veeleer met middelmatigheid dan met essentialisme vereenzelvigde. Hij was juist geïnteresseerd in de 'staarten' van de verdeling. Galton koesterde het afwijkende, het uiterste, het uitzonderlijke en het uitmuntende. En zocht – als uitvinder van de vingerafdrukken voor forensische identificatie – wellicht zelfs het hyperindividuele. Soms schoot hij te ver door; hij meende dat de evolutie grote discontinue variaties nodig had – Galton noemde deze *sports* – juist om de *regression to the mean* te compenseren.

Hoe dan ook, zijn anti-egalitarisme lijkt bescheiden vergeleken met het geniebegrip van Goethe en de *Übermensch* van Nietzsche, maar zijn opvattingen over de fundamentele ongelijkheid van mensen maakten hem een groot voorstander van eugenetica. Zijn zorg over de teloorgang van het Engelse ras en de toekomst van de samenleving bracht hem ertoe het eugenetisch genootschap op te richten, maar ook stelselmatig data te verzamelen om langs biometrische weg de beginselen van de genetica te bestuderen. Kansrekening en statistiek vormden het fundament van de eugenetica. Zijn beschermeling Karl Pearson wilde zelf geen evolutionaire ethiek ontwikkelen, maar was niet minder dan Spencer een sciëntist en evolutiedenker van het eerste uur. In zijn filosofische werk *The Grammar of Science* (1892) stelt hij dat '*The insight which the investigations of Darwin, seconded by the suggestive but far less permanent work of Spencer, have given us into the development of both individual and social life, has compelled us to remodel our historical ideas and is slowly widening and consolidating our moral lards.*' Pearson corrigeerde Galton inzake de *sports* en toonde aan dat klassen van scheve kansverdelingen nodig waren om

recht te doen aan de variabiliteit in de natuur. Zijn stellingname in het *nature nurture*-debat leidde tot de overtuiging dat het weinig zin heeft te investeren in rassen die minder *fit* zijn. Dit alles ondanks het feit dat zijn politieke opvattingen in dat tijdsgewricht in vele opzichten als progressief konden worden aange-merkt. Ook Fisher sloot zich aan bij de eugenetische beweging en bepleitte een landelijk beleid om het geboortecijfer onder 'getalenteerden' te bevorderen en voortplanting van de *poor classes* te ontmoedigen. Zijn bezorgdheid dat *social welfare*-programma's ertoe zouden leiden dat de slechtere genen geselecteerd zouden worden, vormden nadrukkelijk een motivatie voor zijn genetisch/statistische werk.

Het gaat in dit korte essay niet om een stellingname inzake al dan niet vermeende nobele motieven van wetenschappers, die zich eind 19e eeuw onderwierpen aan de spenceriaanse morele verplichting samenleving en mensheid te 'verbeteren'. In dat tijdsgewricht worstelden velen met de ongemakken van een evolutionaire ethiek. Ook Darwin was er niet vrij van. Hij probeerde op evolutionaire gronden de superioriteit van de Engelse moraal te verdedigen ten opzichte van die der tribale samenlevingen die de Britse koloniën bevolkten. Maar sommigen gingen verder en manifesteerden zich als wegbereiders of prominente aanhangers van de eugenetica en het sociaaldarwinisme. Het verloop van de 20e eeuwse (politieke) geschiedenis leert dat naïeve toepassing van de evolutieleer op de moraal kan leiden tot uitwassen, zoals de rassen-theorieën van Ernst Haeckel. Zoals bekend ontwikkelde ook het nazisme een ideologie die teruggreep op argumenten ontleend aan de evolutieleer en het hier geschetste debat.

Epiloog

De dikwijls gepostuleerde scheiding tussen de wereld der waarden en die van de feiten krijgt ook gestalte in de vorm van een methodologisch maxime, of liever gezegd profylactische waarschuwing, die in de

19e eeuw reeds bekend was. Wie tracht normen af te leiden uit feiten begaat de naturalistische drogreden die al bij Hume naar voren komt, maar door G. E. Moore het meest stringent is verwoord. De kloof tussen 'is' en 'ought' is wezenlijk niet te overbruggen, en normstellingen of morele imperatieven zijn geen premissen die uit feiten logisch kunnen worden afgeleid. Met enige goede wil kan men zeggen dat de kwalificatie *fittest* uit *survival of the fittest* bijna onvermijdelijk aanleiding geeft tot deze drogreden. Ofschoon de kwestie uiteraard beduidend subtieler ligt dan hier geschetst, lijkt het erop dat alle besproken protagonisten, inclusief Darwin, Galton, Pearson en Fisher de valkuilen van dit maxime niet geheel hebben weten te vermijden. Dat de drie statistici daarin veel explicieter waren dan Darwin kan deels een verklaring vormen voor het feit dat zij in de populairwetenschappelijke en filosofische literatuur thans niet altijd de status genieten die andere grensverleggende denkers wel ten deel viel.

Hoe dan ook, het ontoereikende van een louter internalistische wetenschapbeschuwing lijkt hiermee wel geïllustreerd. Feit is dat als gevolg van voornoemde aberraties het denken over ethiek en biologie na de Tweede Wereldoorlog lange tijd uit de gratie raakte en het debat werd soms gegijzeld door het verleden. Opmerkelijk en fascinerend is dat veel van de oude stellingnamen en knelpunten heden ten dage opnieuw opduiken en wel in filosofische debatten over zin en onzin van transhumanisme. NBIC-convergentie – de toepassing van Nanotechnologie, Biotechnologie, Informatica en Cognitieve wetenschap – geeft vorm aan het oude ideaal van *human enhancement*. Thema's als klonen, bio-implantaten, *lab-on-a-chip* en *battle against aging* vinden steeds vaker hun weg naar het publieke debat en bieden als een soort eigentijdse eugenetica nieuwe uitdagingen voor de ethiek.

RICHARD STARMANS is verbonden aan de Faculteit Bèta-wetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica.
E-mail: <starmans@cs.uu.nl>

ETHIEK IN DE OR-BUSINESS

GERRIT T. TIMMER

Mijn eerste gedachte bij de relatie tussen Operations Research (OR) en ethiek is dat deze gelijk is aan de relatie tussen een baksteen en ethiek. Gebruik je de baksteen om een prachtig gebouw neer te zetten, dan ben je goed bezig; sla je er iemand de kop mee in, dan is dat duidelijk minder.

Op zich levert de OR een aanpak, methoden en technieken die waarde vrij zijn. Net als bij iedere toolkit kan deze voor goede en slechte doeleinden worden gebruikt. Toch is deze constatering wellicht te gemakkelijk en valt er meer te zeggen over het gebruik van OR in relatie tot ethische vraagstukken.

Bij ORTEC staat OR hoog in het vaandel. De eerste twee letters van het bedrijf zijn niet toevallig zo gekozen. Al meer dan dertig jaar ondersteunt ORTEC haar klanten met behulp van OR. In deze bijdrage ga ik in op mijn ervaringen bij ORTEC met aspecten van de 'OR-business' die een raakvlak hebben met ethiek.

De toepassing en het doel

Een eerste voor de hand liggende observatie is dat er toepassingen zijn die als meer of minder ethisch kunnen worden bestempeld. Ik denk dat dit bij de toepassing van de OR nogal meevalt. Over het algemeen zijn de 'gebruikers' van OR-modellen grote organisaties en zijn de toepassingsgebieden maatschappelijk volledig geaccepteerd.

Maar zelfs gegeven dat het om grote maatschappelijke organisaties gaat, zouden er natuurlijk ethische zorgen kunnen zijn. Zo is discutabel of het berekenen van de optimale opstelling van tanks ethisch verantwoord is.

Aardiger vind ik discussies over toepassingen als revenue management. Al geruime tijd betalen verschillende mensen voor eenzelfde stoel in hetzelfde vliegtuig een sterk verschillende prijs. OR-technieken

hebben dit mogelijk gemaakt. Het was het antwoord van de gevestigde vliegtuigmaatschappijen op de eerste prijsvechters (People Express in de VS begin jaren tachtig vorige eeuw). De gevestigde airlines leken met hun duidelijk hogere kostenniveau kansloos tegen de *low-cost carriers*. Echter, met American Airlines voorop versloegen ze de eerste aanvalsgolf door gebruik te maken van revenue management. Gemiddeld hadden de gevestigde *airlines* een hogere prijs nodig, maar dat verhinderde ze niet om onder bepaalde omstandigheden onder de prijs van de *low-cost carriers* te gaan. Dit vereiste echter wel sterke stafafdelingen met OR- en IT-expertise. En daarin waren de gevestigde maatschappijen beter.

Revenue management is uitgegroeid tot een vakgebied. Het is relevant bij producten die je niet in voorraad kunt houden, zoals capaciteit op een bepaalde vlucht of een vakantiehuisje in een gegeven week. Na de vlucht of na de week is deze capaciteit waardeloos. Met een lage prijs krijg je de gemiddelde kostprijs niet vergoed, maar iedere euro is beter dan leegstand. Het wordt dan aantrekkelijk om met de prijs te gaan 'spelen'. Gevolg is dat wij voor eenzelfde product (twee identieke stoelen naast elkaar) soms heel verschillende prijzen betalen. Is dat eerlijk? Dit is een niet te beantwoorden vraag. Is het 'eerlijk' dat je allemaal dezelfde prijs betaalt, of is het juist 'eerlijk' dat je eenzelfde fractie betaalt van hetgeen je voor het product over hebt? Revenue management komt meer in de buurt van dat laatste.

Naast de toepassing kan het doel dat je met de OR-toepassing nastreeft ethisch discutabel zijn. Dat geldt nagenoeg voor alle doelen. OR-toepassingen beogen vaak winst te maximaliseren of kosten te minimaliseren. Is dat goed? Als gevolg worden mensen mogelijk ontslagen. Aan de andere kant: wat is het nut van een op zich overbodige baan?

Bovenstaande mogelijke ethische zorgen met betrekking tot de toepassingen c.q. het gehanteerde doel zijn overzichtelijk. In de regel is duidelijk wat er aan de hand is. Men kan een verschillende mening hebben over de wenselijkheid van een toepassing of

een beoogd doel. Maar het is wel duidelijk waar men mee bezig is. Discussies over de (ethische) wenselijkheid zijn dan goed mogelijk en moeten zonedig ook vooral worden gevoerd.

Aanpak en modellering

Lastiger ligt het bij een ander mogelijk gebruik van OR-skills. Dat is waar de crux van ons werk zit in de modellering. Ik acht het vermogen van OR-professionals om een complexe werkelijkheid te modelleren veel waardevoller dan het vermogen om de modellen vervolgens doorgerekend c.q. geoptimaliseerd te krijgen. Niet zelden komt kritiek op de OR-beroepsgroep er op neer dat het verkeerde model goed wordt geoptimaliseerd. Per saldo heb je daar niets aan. De uitkomsten hebben geen praktische waarde, omdat de gebruikte modellering onvoldoende aansluit bij de werkelijkheid. Gevaar is dat de uitkomsten toch worden geloofd. Het is tenslotte met de computer berekend! Denk bijvoorbeeld aan het doorrekenen door het Centraal Plan Bureau van verkiezingsprogramma's. Hier wordt relatief klakkeloos aangenomen dat die modellen kloppen. Nu wil ik zeker niet het tegenovergestelde beweren, maar het is natuurlijk niet triviaal dat de gekozen modellering bij extreem afwijkende politieke keuzes altijd correct blijft. Hier moeten de betrokken professionals zich zeer goed van bewust zijn en mogelijke twijfel duidelijk maken. De minder geschoolde toehoorder zal anders ten onrechte teveel waarde hechten aan de uitkomsten.

Ik roer dit onderwerp niet aan omdat ik tegen de modellering van een complexe werkelijkheid ben. Ik vind dat juist het belangrijkste nut van onze professie. Die modellering is echter veelal geen transparant proces en daarbij ontstaan ethische valkuilen.

Laat ik als voorbeeld het werk noemen dat ik voor de NS doe. Sinds de opwindende om het 'rondje om de kerk' begin deze eeuw, speel ik vanuit ORTEC een rol in discussies tussen werknemersvertegenwoordi-

gers onderling en werknemers-vertegenwoordigers en directie. Het gaat hier over een goede toewijzing van het werk aan het rijdende personeel (machinisten/conducteurs). Het verdelen van tienduizenden taken per dag aan duizenden medewerkers is complex. Op iedere trein moet gekwalificeerd personeel worden ingezet (waarbij treintypes en bekendheid van lokale situaties relevant zijn). Dat moet efficiënt, het moet leiden tot diensten per medewerker die uitvoerbaar zijn (de taken op de verschillende treinbewegingen moeten aansluiten) en aan regelgeving voldoen (CAO, maar ook regels over variatie in het werk). Alle partijen zijn het eens dat de oplossing zodanig moet zijn dat de klant goed wordt geholpen, de kosten voor het bedrijf niet te hoog zijn en het personeel voldoende aantrekkelijke roosters moet hebben.

Er zijn natuurlijk verschillen in het belang dat partijen aan verschillende doelen stellen. Maar nog wezenlijker is dat iedereen verschillende ideeën heeft over het effect van mogelijke keuzes op de doelen. Als bijvoorbeeld het aantal malen dat machinisten van trein wisselen op Utrecht lager wordt, neemt dan de variatie in het werk juist toe of af? Dat het effect van keuzes heel lastig is in te schatten en ieder dit op eigen wijze doet, maakt het nog moeilijker voor partijen (zoals Ondernemingsraad en Directie) om het eens te worden over de te maken keuzes. Ten tijde van het 'rondje om de kerk' liep dit hoog op. ORTEC heeft toen een rol gespeeld om partijen bij elkaar te brengen. Onze bijdrage was om de discussies te objectiveren. Door de complexe werkelijkheid te modelleren was het mogelijk diverse scenario's te analyseren en effecten van keuzes te kwantificeren. Daarmee waren er nog altijd meningsverschillen, maar werd het speelveld veel overzichtelijker. Partijen geloofden onze modellen en rekenpartijen. Hiermee werd veel kaf van het koren gescheiden en konden de discussies terug worden gebracht tot de onderwerpen waar partijen echt een verschillend belang hadden. Een oplossing bleek daarmee mogelijk. Dezelfde rol heb ik in daarop volgende jaren opnieuw kunnen spelen.

Mijns inziens maakt dit voorbeeld een heel belang-

rijke rol duidelijk die OR-professionals kunnen spelen. Dat is modellen gebruiken om de werkelijkheid begrijpelijker te maken en daarmee discussies te objectiveren. Hier liggen echter wel ethische kwesties op de loer. De modellering die de werkelijkheid begrijpelijker moet maken is veelal niet transparant en niet te volgen voor minder deskundigen op ons gebied. Daarmee ligt manipulatie op de loer. Dat kan bewust of onbewust. Bewust kan het gebeuren als men voor een opdrachtgever een gewenste keuze met modellen/berekeningen gaat goedpraten. Maar ook onbewust kan er veel misgaan. Als er sympathie is voor een groep leidt dit er gemakkelijk toe dat sommige aspecten in de modellering juist wel of juist niet worden vergeten. In het NS-voorbeeld was het niet aan de orde is, maar juist omdat het onbewust kan gebeuren, is het iets waar we altijd attent op moeten zijn.

Tot slot

Ethiek speelt overal een rol. Zo ook in de OR-business. Dat kan zijn op een meer en op een minder voor de hand liggende wijze. In de eerste categorie kan men denken aan twijfelachtige toepassingsgebieden en twijfelachtige doelen die worden geoptimaliseerd. Dit is relatief overzichtelijk. Vaak is duidelijk wat men doet. Iedereen kan dat goed beoordelen en men moet persoonlijk de keuze maken of iets ethisch wel of niet verantwoord is.

De tweede categorie moeten we echter niet vergeten. Hier gaat het erom dat we niet via een discutabele modellering onze professie (die voor een ander slecht toetsbaar is) moeten misbruiken om 'gewenste' antwoorden te vinden, c.q. recht te praten wat krom is. Gelukkig zie ik vanuit onze beroepsgroep weinig zorgelijke voorbeelden. Alertheid is echter op zijn plaats. Vooroordelen, sympathieën en antipathieën staan objectiviteit gemakkelijk in de weg.

GERRIT T. TIMMER is medeoprichter en CFO ORTEC en hoogleraar Bedrijfseconometrie aan de Vrije Universiteit. E-mail: <Gerrit.Timmer@ortec.com>

ETHIEK MET SAMBAL

FRED STEUTEL

VvS+OR organiseert een symposium over fraude in de wetenschap, STAtOR wijdt een nummer aan ethiek. Het zijn keerzijden van dezelfde medaille, maar ik vind fraude vele malen interessanter dan ethiek. Dat laatste is eigenlijk een deftig woord voor 'fatsoen' – moet je doen, zegt Balkenende. Maar fraude, dat is avontuur, spanning, sensatie!

Mijn eerste ervaring met statistische fraude had ik als student bij het maken van verslagen van practicumproeven: de meetresultaten werden aangepast aan de te verwachten uitkomsten, vaak bekende natuurconstanten.

Een 'aangrijpend' fraudegeval was te lezen in een vers aan de muur van een zaaltje van het Mathematisch Centrum (CWI): een statisticus slaat de hand aan zichzelf om de zelfmoordstatistieken up-to-date te brengen. Ik vertelde er al eens over.

Een verrassend geval van wetenschapsfraude werd jaren geleden gemeld door C. R. Rao op een Statistische Dag. Hij had statistisch aangetoond dat Mendel, van de erfelijkheidswetten, geknoeid had met zijn waarnemingen; dan pasten ze beter bij zijn theorie. Net als Diederick Stapel dus.

Minder sympathie heb ik voor verwoede fraudebestrijders zoals Robbert Dijkgraaf, die met Pim Levelt tegen Stapel ten strijde trok. Hij gaat mij ook te vaak te biecht bij professor Matthijs van Nieuwkerk van *De Wereld Draait Door*. Ik schreef in *de Volkskrant* dat er wel ernstiger misstanden waren aan de universiteiten: rare wetenschappen met rare hoogleraren (politici, koningshuizers en cabaretiers). Daar moesten ze eens wat aan doen. Ik werd bijgevallen door collega en medecolumnist Henk Tijms.

Een van de mooiste verslagen van wiskundig plagiaat komt uit de mond van *singer-songwriter* en statisticus Tom Lehrer, bekend van opwindende nummers als *Poisoning Pigeons in the Park* en de onvergetelijke Irish

Ballad: *'About a maid I'll sing a song. She didn't have her family long. Not only did she do them wrong, she did everyone of them in.'* Zou Robbert het plagiaatlid kennen? Gerard 't Hooft, die ik eens ontmoette in De Witte in Den Haag, kent het vast. Hij kende ook de eerste regels van *The Elements*, waar Lehrer alle 102 toen bekende chemische elementen de revue laat passeren. Hier is de plagiaatmeester zelf.

Let no one else's work evade your eyes
Remember why the good Lord made your eyes
So don't shade your eyes
But plagiarize, plagiarize, plagiarize.

Fraude in de wetenschap is zeldzaam; daarom krijgt het zoveel aandacht. Over grootschalige financiële fraude wordt nauwelijks bericht; geen nieuws.

Andere ethische kwesties in de wetenschap betreffen de onderwerpen waaraan wordt gewerkt. De ontwikkeling van de atoombom was begin jaren veertig zo'n kwestie. Nog onlangs werd bericht over het niet-publiceren van gegevens over akelige ziekten. Steeds blijkt dat wetenschappers zich niet laten tegenhouden door ethische bezwaren; ze hebben nog betere ethische argumenten dan hun tegenstanders.

In de statistiek hebben we weinig ethische problemen; toch was er Lucia de Berk. En, moeten we toepassing van de Student-toets niet afwijzen, omdat de bedenker werkte voor een bierbrouwer die honderden Ierse gezinnen tot de bedelstaf heeft gebracht? Denk aan *Angela's Ashes*.

Tenslotte, de wereld is meer gediend met simpel fatsoen dan met deftige ethische regels.

FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven.

E-mail: <f.w.steutel@tue.nl>

INFORMED CONSENT:

NOODZAKELIJK KWAAD?

Als onderzoeker, statistisch consultant of beoogd studiedeelnemer – allen kunnen we te maken krijgen met een 'informed consent'. In deze bijdrage gaan we kort in op het belang van een informed consent en geven we wat praktische tips.

GERBEN BEKKER & HILDE TOBI

Door middel van een *informed consent* geeft de onderzoeker aan dat zij bewust instemming van de deelnemer aan het onderzoek zoekt en de deelnemer ook afdoende wil informeren over het onderzoek en zijn rechten binnen het onderzoek. Over het algemeen zal de onderzoeker in de informed consent het doel van het onderzoek, de globale gang van zaken, mogelijke risico's en rechten van de deelnemer uitleggen. De (aspirant) deelnemer kan als reactie zijn *consent* (toestemming) geven of weigeren deel te nemen aan het onderzoek. Een eenmaal gegeven consent kan ten alle tijden door de respondent ingetrokken worden. Een informed consent is bijna altijd nodig bij experimenteel medisch onderzoek maar vaak ook bij sociaal-wetenschappelijk experimenteel onderzoek met mensen.

Ethisch handelen

Waarom wordt er eigenlijk van de onderzoeker verwacht dat zij de deelnemer volledig informeert over een experiment? De Declaratie van Helsinki in 1964 heeft in grote mate het denken over ethisch handelen bij medisch onderzoek bepaald. Deze verklaring heeft zijn basis in de afschuw over de medische experimenten van het naziregime zoals die in Neurenberg openbaar werden. De Verklaring van Helsinki wordt gere-

geld herzien en uitgebreid, wat de laatste 10 jaar ook tot kritiek heeft geleid, met name op het terrein van de informed consent. Bij deelname aan een medisch experiment lijkt het logisch dat de deelnemer (de patiënt of zijn directe vertegenwoordiger) geïnformeerde instemming moet worden gevraagd. Maar hoe zit het met sociaalwetenschappelijke experimenten?

In de sociale wetenschappen is het denken hierover sterk gevormd door experimenten zoals die van Milgram (1963). Milgram onderzocht de gehoorzaamheid van gewone mensen tot het uitvoeren van immorele gedragingen middels het volgende experiment. De deelnemer moest samen met een andere deelnemer, die in werkelijkheid een mede-onderzoeker was, een leertaak uitvoeren. De deelnemer werd altijd aangewezen als de leraar en de ander als leerling. De leraar was met de experimentleider in één kamer en kon de leerling niet zien. De leraar stelde de leerling vragen. Wanneer de leerling een fout beging moest de leraar de leerling een elektrische schok toedienen, te beginnen bij 15 volt en toenemend met iedere opvolgende fout. Bij het toedienen van 300 en 315 volt hoorde de leraar de leerling bonzen op de muren. Vanaf 330 volt stopte de leerling met reageren op de vragen. Wanneer de leraar wilde stoppen, spoorde de experimentleider hem aan door te gaan. Er werd verteld dat de schokken pijnlijk zouden zijn, maar geen blijvende weefsel-

schade veroorzaakten. Geen enkele deelnemer stopte voor het bereiken van 300 volt en 65 procent ging door tot de maximale 450 volt. De emotionele stress die het experiment bij de deelnemers veroorzaakte is uitvoerig bekritiseerd. Het experiment heeft duidelijk laten zien hoeveel overwicht een experimentleider kan hebben op deelnemers. Zelfs zoveel dat deelnemers zich bij experimentele excessen niet goed kunnen verweren.

Reikwijdte van de informed consent

Wat hoort er in een informed consent te staan? De American Psychological Association (2010) stelt voor dat minimaal het doel, de duur en de procedures van het onderzoek worden gegeven. Daarbij wordt de deelnemer onder andere op zijn rechten gewezen tijdens het onderzoek, gewezen op mogelijke risico's, wat er tegenover onderzoekdeelname staat, en wat het privacybeleid is.

Tot zover lijkt het geen probleem om aan deze eisen te voldoen. Maar het vertellen van het doel van het onderzoek kan de interne validiteit van het onderzoek ernstig bedreigen. Neem bijvoorbeeld een onderzoek waarin wordt gekeken wat de invloed van de emotionele staat van de deelnemer is op het invullen van extreme waarden op een antwoordschaal. De deelnemers zullen zich door de informed consent beter bewust worden van hun emotionele staat bij aanvang van het experiment en mogelijk anders reageren op de experimentele condities, bijvoorbeeld de video's waarmee de onderzoeker de emotionele staat van de deelnemer probeert te manipuleren. Ook kan een deelnemer vinden dat zijn emotionele toestand geen invloed mag hebben op zijn mening en daarom voor de zekerheid alle vragen neutraal gaan beantwoorden. Een ander voorbeeld is een onderzoek waarbij deelnemers gevraagd wordt om een categoriseringstaak uit te voeren, maar waar in werkelijkheid wordt gemeten of deelnemers meer positieve of meer negatieve associaties met kweekvlees hebben (zie kader 1). Wanneer de respondent zich hiervan bewust is, kan hij proberen de positieve of negatieve associaties

te maskeren. Dit zijn voorbeelden waarbij informeren over het onderzoeksdoel de interne validiteit van een experiment bedreigt, maar dit fenomeen kan zich ook voordoen bij puur observationeel onderzoek.

Is misleiding dan misschien toegestaan onder specifieke voorwaarden? De Verklaring van Helsinki (2002) stelt dat overwegingen betreffende het welzijn van de deelnemer boven de belangen van wetenschap en maatschappij moeten gaan. Deze gedachte is nader uitgewerkt door de American Psychological Association (2010) die schrijft dat misleiding enkel kan wanneer er geen andere mogelijkheid is om een significante wetenschappelijke bijdrage te leveren. Daarbij mag misleiding niet tot fysieke pijn of zware emotionele stress leiden en moet de deelnemer zo snel mogelijk geïnformeerd worden over het werkelijke doel van het onderzoek door middel van een debriefing. De vraag blijft natuurlijk wanneer een onderzoek een significante wetenschappelijke bijdrage levert en wat de rol van replicatie is. Verder is de scheidslijn tussen emotionele en zware emotionele stress moeilijk te definiëren, en daarna, te voorspellen. Daarom wordt in de praktijk bij twijfel vaak een vorm van nazorg in de debriefing genoemd, zoals het telefoonnummer of webadres van een hulporganisatie of van een onafhankelijke deskundige. In het eerder genoemde onderzoek naar associaties met kweekvlees konden deelnemers na afloop van het experiment aangeven of zij meer over het onderzoek wilden weten.

Knelpunten

Een mogelijk praktisch knelpunt van informed consent is de combinatie van beloofde anonimiteit en studieopzet, vooral bij longitudinaal onderzoek. Dat kan vrij eenvoudig worden opgelost door versleuteling van de data door een externe databeheerder. De statisticus kan dan in de analyses bijvoorbeeld een onderscheid maken tussen de deelnemers die zich actief terugtrekken uit het onderzoek, de *no shows* en zij die uitgevalen zijn door overlijden.

Een lastiger methodologisch knelpunt van de infor-

med consent is het volledig geïnformeerd zijn van de deelnemer wanneer de verwachtingen van de deelnemer de uitkomsten van het onderzoek kunnen beïnvloeden, zoals hierboven beschreven. Volledige transparantie over de doeleinden van het onderzoek brengt nog een ander lastig methodologisch probleem met zich mee. De deelnemer kan na geïnformeerd te zijn zelf inschatten of het wenselijk is om deel te nemen aan het onderzoek. Het probleem hiervan is dat een bepaalde groep deelnemers misschien eerder geneigd is om niet aan het onderzoek deel te nemen na informed consent. Dit resulteert daarmee in een selectie bias. Dit probleem kan ook spelen bij observationeel onderzoek. Bijvoorbeeld in opinieonderzoek naar de Hedwigepolder zullen zelfbewuste mensen met een uitgesproken mening veel eerder geneigd zijn mee te doen aan het onderzoek, dan mensen die vinden dat ze er te weinig vanaf weten of mensen die geen uitgesproken mening hebben.

Volledig geïnformeerd zijn heeft dus invloed op zowel de interne als de externe validiteit.

Het mag duidelijk zijn: de informed consent is er niet voor niets, maar heeft wel een prijs. De afweging tussen wel, geen of een informed consent met deceptie, is lastig te maken voor individuele onderzoekers. Gelukkig hoeven ze deze afweging ook niet alleen te maken. In toenemende mate zijn er ethische commissies die er op helpen toezien dat deze afwegingen zorgvuldig worden gemaakt.

LITERATUUR

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Retrieved from <http://www.apa.org/ethics/code/index.aspx>.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67(4), 371–378.

GERBEN BEKKER is promovendus bij MCB/RME van Wageningen Universiteit.

E-mail: <gerben.bekker@wur.nl>

HILDE TOBI is universitair hoofddocent Research Methodologie aan Wageningen Universiteit en redacteur van *STAtOR*.

E-mail: <hilde.tobi@wur.nl>

CONSENT FORM

Welkom bij dit onderzoek over de snelheid en accuraatheid waarmee woorden gecategoriseerd kunnen worden. Dit onderzoek wordt uitgevoerd door de promovendus Gerben Bekker van Wageningen University.

Procedure:

Deelname aan het onderzoek duurt gemiddeld 25 minuten. Hiervoor ontvang je een vergoeding van €3,00. Het onderzoek bestaat uit drie onderdelen. Je begint met een computertaak waarin je verschillende woorden zo snel mogelijk en precies mogelijk moet categoriseren. Daarna krijg je vragen over jouw mening ten opzichte van verschillende onderwerpen. Tot slot worden enkele algemene vragen gesteld. In dit onderzoek zijn we geïnteresseerd in jouw mening, er zijn hierbij geen goede of foute antwoorden. De deelname aan het onderzoek is volledig vrijwillig en gedurende het experiment ben je vrij om te stoppen zonder dat je hiervoor een verklaring hoeft te geven.

Data management en vertrouwelijkheid:

Je krijgt een willekeurig proefpersoon nummer toegewezen en jouw data zal enkel met dit nummer bewaard worden. De data zal op de server van de universiteit opgeslagen worden. Jouw antwoorden zullen niet voor andere doeleinden dan wetenschappelijk onderzoek gebruikt worden. In geen geval zullen antwoorden van individuele respondenten kenbaar worden gemaakt. Wageningen University garandeert dat jouw data anoniem verwerkt zal worden en alleen gebruikt wordt voor onderzoeksdoeleinden.

Betrokkenheid van derden:

Bij dit onderzoek zijn geen derde partijen betrokken. Dit onderzoek is enkel voor wetenschappelijke doeleinden.

Contact:

Wanneer je vragen of opmerkingen over het onderzoek hebt kun je contact opnemen met:

Gerben Bekker, promovendus MCB / RME, gerben.bekker@wur.nl;

Dr. Hilde Tobi, universitair hoofddocent RME, hilde.tobi@wur.nl;

Dr. Arnout Fischer, universitair docent MCB, arnout.fischer@wur.nl;

Door te tekenen geef je aan dat je de hierboven genoemde verklaringen begrijpt en uit vrije wil instemt deel te nemen aan dit onderzoek.

Datum: _____

Handtekening participant: _____

Kader 1. Voorbeeld van een informed consent

RANDOMISED RESPONSE: over het hoe en waarom van privacybescherming

Randomised response is een methode voor het meten van sensitieve attitudes of gedrag, zoals wetsovertredingen, drugs- en alcoholgebruik en seksualiteit. In deze bijdrage geven we voorbeelden van randomised response en staan we stil bij de ethiek van de bescherming van privacy. We zullen laten zien dat het vanuit ethisch perspectief niet alleen belangrijk is dat randomised response de privacy beschermt, maar dat het evenzo essentieel is dat de respondenten de gehanteerde methode vertrouwen.

ARDO VAN DEN HOUT & MARIJE ALTORF

Randomised response is een statistische techniek om gegevens te verzamelen in een situatie waar respondenten terughoudend zijn in het geven van informatie (Warner 1965). De terughoudendheid kan verschillende redenen hebben. In sommige gevallen gaat het om overtredingen van de wet, bijvoorbeeld uitkeringsfraude of corruptie. Bij andere gaat het om privé-zaken zoals seksuele oriëntatie, abortus, gokverslaving.

Het basisidee van randomised response is dat er bescherming wordt geboden op individueel niveau, maar dat er desondanks statistische conclusies kunnen worden getrokken op het niveau van de groep die wordt onderzocht (de populatie). De veronderstelling is dat door de bescherming mensen eerder en beter zullen meewerken aan het onderzoek.

Randomised response met dobbelstenen

Er zijn verschillende technieken voor randomised response. We bespreken er een die illustratief is. Stel we willen wetsovertreding met betrekking tot een uitkering onderzoeken met de vraag 'Heeft u wel eens gefraudeerd ten aanzien van uw uitkering?'. Het is duidelijk dat dit een moeilijke vraag is om gegevens mee te verzamelen. Mensen die frauderen geven in het algemeen niet graag informatie over het niet volgen van de regels. Stel dat de vraag wordt gesteld aan een vrouwelijke respondent. Voordat ze ja of nee antwoordt, gooit ze met twee dobbelstenen. De uitkomst van de worp houdt ze voor zich. Als de uitkomst 2, 3 of 4 is, antwoordt ze met ja, ongeacht of ze wel of niet

heeft gefraudeerd. Als de uitkomst 5, 6, 7, 8, 9, of 10 is, antwoordt ze ja of nee naar waarheid. Als de uitkomst 11 of 12 is, antwoordt ze met nee, ongeacht of ze wel of niet heeft gefraudeerd.

Omdat de ondervrager de uitkomst van de worp niet kent, is het niet te achterhalen of een ja-antwoord daadwerkelijk correspondeert met fraude. Dit is de privacy-bescherming op individueel niveau. Echter, omdat we de kansverdeling van de uitkomst van de worp met de dobbelstenen kennen, kunnen we de kans berekenen dat fraudegedrag van een respondent leidt tot een ja-antwoord - mits de respondent de randomised response-instructies volgt. De volgende conditionele kansen kunnen worden afgeleid:

$$P(\text{ja}|\text{fraude}) = 33/36 \quad \text{en} \quad P(\text{ja}|\text{geen fraude}) = 6/36.$$

Vervolgens geldt:

$$\begin{aligned} P(\text{ja}) &= P(\text{ja}|\text{fraude}) P(\text{fraude}) + P(\text{ja}|\text{geen fraude}) P(\text{geen fraude}) \\ &= P(\text{ja}|\text{fraude}) P(\text{fraude}) + P(\text{ja}|\text{geen fraude}) (1 - P(\text{fraude})) \end{aligned}$$

Waaruit volgt dat:

$$P(\text{fraude}) = \frac{P(\text{ja}) - P(\text{ja}|\text{geen fraude})}{P(\text{ja}|\text{fraude}) - P(\text{ja}|\text{geen fraude})}$$

De conditionele kansen in deze vergelijking zijn gegeven met kansverdeling van de uitkomst van de worp, de kans $P(\text{ja})$ kan worden geschat door de proportie ja-antwoorden in de steekproef. Aldus hebben we een schatting van $P(\text{fraude})$ en een statistische conclusie over het fraudegedrag in de populatie. Er is een formule voor de variantie van de schatter. De methode kan ook worden toegepast als er meer dan twee antwoordcategorieën zijn. Ook zijn er statistische modellen ontwikkeld waarmee het gedrag dat wordt onderzocht met randomised response kan worden verklaard aan de hand van variabelen zoals geslacht, leeftijd, of sociaal-economische status.

Statistisch gezien is het idee van randomised response dat data wordt verzameld met een techniek die misclassificatie toelaat. In het voorbeeld is $P(\text{ja}|\text{geen fraude})$ een misclassificatiekans, namelijk de kans dat een latent nee (geen fraude) wordt geobser-

veerd als een ja. Omdat het stochastisch gedrag van de misclassificatie bekend is, kan de statistische analyse hiervoor worden gecorrigeerd.

Als respondenten niet de randomised response-instructies volgen, heeft dit natuurlijk een ernstig effect op de data analyse. Wanneer het om een beperkte groep gaat die niet meewerkt door altijd nee te beantwoorden, dan is er statistisch nog wel iets te corrigeren (Böckenholt & Van der Heijden, 2007). In het algemeen echter ondergraaft het niet volgen van de instructies de data-analyse.

Post-randomisatie

Randomised response kan ook worden gebruikt om de privacy van respondenten te beschermen in een bestaand databestand. Dit heet post-randomisatie: misclassificatie wordt door middel van randomised response uitgevoerd nadat de gegevens zijn verzameld (Gouweleeuw et al., 1998). Post-randomisatie kan worden toegepast als de dataverzamelaar gegevens aan een derde partij wil doorgeven, dat wil zeggen aan onderzoekers die buiten de vertrouwensrelatie staan tussen de vragensteller en de respondent.

Stel dat het gaat om een bestand met gegevens over individuele spaartegoeden en dat naast het spaartegoed, ook een aantal persoonlijke gegevens wordt verzameld zoals leeftijdsgroep, geslacht, woonplaats en geboorteland. Als gegevens worden doorgegeven, dan worden directe indicatoren zoals naam en huisadres sowieso weggelaten, maar dat beschermt niet altijd afdoende. Als het gaat om een respondent in de leeftijdsgroep 70-80, die woont in Broek op Langendijk en geboren is in Peru, dan is het heel goed mogelijk dat deze combinatie van kenmerken uniek is in de steekproef én in de populatie. De identiteit van deze respondent is niet beschermd zonder extra maatregelen.

Het toepassen van post-randomisatie bestaat eruit dat voor bepaalde variabelen in het bestand geobserveerde waarden worden misgeclassificeerd en dat deze misclassificatie wordt uitgevoerd met condi-

onele kansen die bekend zijn. Vervolgens wordt het bestand met de misgeclassificeerde gegevens vrijgegeven voor een derde partij tezamen met informatie over de conditionele kansen.

Verschillen tussen randomised response en post-randomisatie

Het grote verschil tussen randomised response en post-randomisatie is natuurlijk dat bij de eerste misclassificatie wordt uitgevoerd door de respondent zelf en bij de tweede de misclassificatie wordt uitgevoerd door een computer. Een ander verschil is dat randomised response typisch wordt toegepast op variabelen met latente waarden (bijvoorbeeld fraudegedrag) en post-randomisatie op variabelen met manifeste waarden (bijvoorbeeld geslacht, leeftijdsgroep). Een interessant verschil vanuit een statistisch oogpunt is dat bij post-randomisatie de misclassificatieparameters kunnen worden bepaald aan de hand van de (al verzamelde) gegevens. Als er een maat is voor de bescherming, dan kan de misclassificatie daarop worden afgesteld (Van den Hout & Elamir, 2006). Bij randomised response is dit niet mogelijk omdat de misclassificatieparameters moeten worden vastgesteld voordat de gegevens worden verzameld.

Het belang van privacy-bescherming: de scheiding tussen privé en publiek

Hoewel randomised response is ontworpen om in bepaalde situaties statistische resultaten te verbeteren, dringen zich ook ethische vragen op. We zullen deze bespreken aan de hand van het begrip privacy.

Als recht speelt privacy een belangrijke rol in de relatie tussen burgers en overheid, tussen burgers en bedrijven, alsmede tussen burgers onderling. In dit verband wordt privacy begrepen als controle over informatie over jezelf. Het lijkt niet meer dan vanzelfsprekend dat gegevens over godsdienst, politieke

gezindheid of seksuele voorkeur niet zomaar worden doorgegeven aan een derde partij (zie ook de Wet Bescherming Persoonsgegevens <www.rijksoverheid.nl/onderwerpen/persoonsgegevens>).

Deze vanzelfsprekendheid suggereert een ethisch voorschrift: privacy moet worden beschermd. De redenen hiervoor zijn sterk verbonden aan een besef van wat het is om mens te zijn en deel te nemen aan de maatschappij. Privacy wordt fundamenteel geacht voor een open en vrije samenleving. Voor de Duits-Amerikaanse filosoof Hannah Arendt is deze eis zo wezenlijk dat ze geen enkele rol ziet voor de overheid in het privé-leven. Zij stelt hierbij de open samenleving recht tegenover de totalitaire samenleving, waar de overheid de huiskamer binnendringt door bijvoorbeeld kinderen tot klikken over hun ouders aan te zetten.

De positie van Arendt is sterk bekritiseerd vanuit verschillende perspectieven. Bowring (2011) geeft hiervan een goed overzicht. Arendts strenge tegenstelling tussen privé en publiek beschrijft ook niet de werkelijkheid van de Nederlandse samenleving. De overheid heeft toegang tot ons privé-leven, bijvoorbeeld in de vorm van regelgeving voor een paspoort, of voorwaarden voor een uitkering. Maar de tegenstelling helpt bij het nadenken over privacy en kan dienen als een waarschuwing voor wat mis kan gaan.

Privacy wordt niet alleen gezien als van belang voor de open samenleving, maar ook voor de ontwikkeling van het individu. Relaties tussen individuen kunnen zich alleen ontwikkelen wanneer de privé-sfeer niet wordt binnengedrongen of bekeken. Evenzo zijn spontaniteit, autonomie, creativiteit en persoonlijke verantwoordelijkheid gebaat bij de bescherming van privacy en bij het besef van deze bescherming. Dit laatste aspect doet een extra beroep op de onderzoeker. Bij statistisch onderzoek naar gevoelige zaken gaat het er niet alleen om te voorkomen dat informatie wordt verspreid. De onderzoeker moet ook voorkomen dat de deelnemer dit zal vrezen. Privacy is in deze context sterk verbonden met gemoedsrust (*peace of mind*).

De bescherming die randomised response en post-randomisatie geven betreft beide aspecten. Allereerst

wordt geprobeerd te voorkomen dat individuele informatie wordt verspreid. Die wordt onbekend gehouden (randomised response) of verdoezeld (post-randomisatie). Daarnaast moet de respondent beseffen dat privacy is gewaarborgd. Bij randomised response kan dit worden bewerkstelligd door actieve meewerking van de respondent.

Privacy lijkt zo beschermd bij het verzamelen van de gegevens. Maar dat is niet het hele verhaal. Gegevens van een randomised response-onderzoek kunnen leiden tot maatregelen die toch in het privé-leven van de respondent ingrijpen. Dit is in het bijzonder het geval bij wetsovertredingen, waar inzichten verkregen met randomised response kunnen worden gebruikt om de controle op regelnaleving te veranderen. In dit geval zijn wel de individuele gegevens van de respondenten beschermd, maar medewerking aan het onderzoek kan nadelig uitwerken voor de groep en dus uiteindelijk ook voor de respondent.

Bij privé-zaken kan het evenzo zijn dat inzichten verkregen met randomised response worden gebruikt om veranderingen door te voeren. Maar deze veranderingen kunnen voordelig zijn voor de groep. Informatie over latente gokverslaving bijvoorbeeld, kan leiden tot een uitbreiding van zorgverlening. Dit is een optimale situatie voor het toepassen van randomised response. Dit voorbeeld maakt ook duidelijk dat er goede redenen zijn om niet aan Arendts strenge onderscheid tussen publiek en privé vast te houden. In deze laatste situatie is een beroep op de respondenten om mee te werken soms mogelijk. Als het onderzoek zaken betreft, waar privé-leven publiek kan worden bij ernstige gevallen (bijvoorbeeld faillissement bij gokverslaving), dan kunnen resultaten van het onderzoek leiden tot maatregelen die het aantal ernstige gevallen beperken en daarmee een betere scheiding bewerkstelligen tussen privé en publiek.

Conclusie

Randomised response biedt bescherming op het niveau van de respondent, maar niet op het niveau

van de groep (de populatie). Voor de statistische analyse is het van uiterst belang dat de respondenten het randomised response-design volgen. Vanuit ethisch perspectief is het minstens even belangrijk dat onderzoek de privacy waarborgt. Informatie mag niet verder worden doorgegeven en respondenten moeten de methode hierin kunnen vertrouwen.

Deze ethische overwegingen geven tot slot inzicht in het functioneren van randomised response en post-randomisatie. Randomised response zal meer moeilijkheden opleveren wanneer medewerking van respondenten kan leiden tot maatregelen die op groepsniveau nadelig zijn. De optimale situatie voor randomised response is die waar respondenten beseffen dat hun privé-leven is beschermd en bovendien inzien dat het onderzoek kan leiden tot maatregelen die voordelig zijn voor hun groep en dus uiteindelijk misschien ook voor hen.

LITERATUUR

- Arendt, H. (1998). *The Human Condition*. Chicago: University of Chicago Press.
- Bowring, F. (2011). *Hannah Arendt: A Critical Introduction*. Londen: Pluto Press.
- Böckenholt, U. & Van der Heijden, P. G. M. (2007). Item randomized-response models for measuring non-compliance: risk-return perceptions, social influences and self-protective responses. *Psychometrika*, 72, 245–262.
- Gouweleeuw, J. M., Kooiman, P., Willenborg, L. C. R. J. & De Wolf, P.-P. (1998). Post randomisation for statistical disclosure control: theory and implementation. *Journal of Official Statistics*, 14, 463–478.
- Van den Hout, A. & Elamir, E. A. H. (2006). Statistical disclosure control using post randomisation: Variants and Measures for Disclosure Risk. *Journal of Official Statistics*, 20, 711–731.
- Warner, S. L. (1965). Randomized response: a survey technique for eliminating answer bias. *Journal of the American Statistical Association*, 60, 63–69.

ARDO VAN DEN HOUT is wiskundige en werkt als docent-onderzoeker in het Department of Statistical Science, University College London.
E-mail: <ardo.vandenhout@ucl.ac.uk>

MARIJE ALTORF is filosoof en werkt als docent-onderzoeker in de School of Theology, Philosophy, and History, St Mary's University College, London.
E-mail: <marije.altorf@smuc.ac.uk>

ONAFHANKELIJKE STATISTIEK

KEES ZEELBERG

Eind januari was er in het 8 uurjournaal een reportage over Elstat, het Griekse statistiekbureau. Griekse onderzoeksrechters hebben het parlement en het hoogerechtshof aanbevolen om te onderzoeken of het statistiekbureau de cijfers over het begrotingstekort gemanipuleerd heeft. Om in aanmerking te komen voor Europese en IMF-steun, zou eind 2010 het begrotingstekort kunstmatig verhoogd zijn. Daar doorheen speelt dat de zaak aangebracht is door voormalige leden van de raad van bestuur van Elstat, die nu juist ontslagen zijn omdat ze eerder de cijfers gemanipuleerd zouden hebben. De geloofwaardigheid van de huidige Griekse directeur-generaal van Elstat, Andreas Georgiou, is in elk geval veel hoger dan die van zijn voorgangers. Eurostat, het Europese statistiekbureau, heeft duidelijk geconstateerd dat onder zijn leiding de kwaliteit van de Griekse statistieken aanzienlijk is verhoogd. Maar het verhaal staat in een veel langere rij van verhalen uit de afgelopen 10 jaar over politieke bemoeienis met de Griekse statistiek. Keer op keer moest Eurostat vaststellen dat er maat-

regelen nodig waren om het cijfer van het Griekse begrotingstekort te verbeteren.

Ook in Argentinië is er politieke bemoeienis met de statistiek. De regering heeft de berekeningswijze van de inflatie (de consumentenprijsindex) opgelegd. Dat is op zich al een enorme ingreep in de onafhankelijkheid van de statistiek, omdat de onafhankelijkheid van een statistisch bureau nu juist tot uitdrukking komt in zelfstandigheid wat betreft methoden en publicaties. In Argentinië gaat het echter nog verder: iedereen die zich hierin niet wil schikken wordt bedreigd met een boete tot 125.000 US dollar of zelfs een gevangenisstraf. Enkele voormalige medewerkers van het Argentijnse bureau hebben dat ervaren: zij hebben openlijke kritiek geuit en komen nu voor de rechter. Het Internationale Monetaire Fonds (IMF) heeft – hoogst uitzonderlijk – openlijk een dwingend methodologisch advies aan Argentinië gegeven voor de berekeningswijze van de consumentenprijsindex. Dit advies is niet opgevolgd en daarop is het IMF-kantoor in Argentinië gesloten.

CBS-wet

Hoe is in Nederland de onafhankelijke statistiek geregeld? Het Centraal Bureau voor de Statistiek (CBS) is in 1899 opgericht om te voorzien in statistieken voor 'praktijk, beleid en wetenschap'. In de loop van de twintigste eeuw is de onafhankelijkheid van het CBS steeds duidelijker geformuleerd. Vooral in de jaren 90 van de vorige eeuw en eerste jaren van deze eeuw is ook in het parlement daarover gediscussieerd bij debatten over wetsvoorstellen. Dit heeft in 2003 geleid tot de nieuwe CBS-wet, waarbij het CBS een zelfstandig bestuursorgaan werd. In die wet is ook de onafhankelijkheid geformuleerd: artikel 8 zegt dat de directeur-generaal van het CBS niet ondergeschikt aan de minister van Economische Zaken, Landbouw en Innovatie kan zijn en artikel 18 dat 'De directeur-generaal bepaalt de methoden waarmee de in de werk- en meerjarenprogramma's opgenomen onderzoeken worden uitgevoerd en de wijze waarop de resultaten van die onderzoeken worden openbaar gemaakt'. Het CBS is dus onafhankelijk in de wijze waarop de statistieken worden samengesteld en in de wijze van publiceren ervan.

Maar waarom is er eigenlijk onafhankelijke statistiek? Statistische bureaus zijn immers opgericht om in het bijzonder de overheid te voorzien van beleidsrelevante informatie; waarom mag die overheid zich dan niet bemoeien met de manier waarop die statistieken gemaakt worden? De toelichting op de CBS-wet geeft als belangrijkste motivatie voor onafhankelijkheid dat voorkomen moet worden dat bepaalde delen van de samenleving, i.c. groepen die niet tot de overheid behoren, een informatieachterstand hebben bij onderhandelings- en besluitvormingsprocessen. Als elke deelnemer aan die processen dezelfde informatie heeft en over dezelfde gegevens beschikt, dan kan de discussie gaan over de inhoud van het beleid, en niet over de gegevens die ten grondslag liggen aan de keuze voor een bepaald beleid. Dit past in de traditie van *evidence-based policy making* die in Nederland zo

sterk is; denk bijvoorbeeld aan de planbureaus waar politieke partijen hun verkiezingsprogramma's laten doorrekenen, iets wat in de meeste andere landen ondenkbaar zou zijn.

Praktijkcode

Ook in veel andere landen bestaan vergelijkbare statistiekwetten waarin de onafhankelijkheid van het statistiekbureau vastgelegd is. In de Europese Unie (EU) bestaat sinds 2005 de Praktijkcode voor Europese Statistieken (Europese Unie, 2005), waarin bepaald wordt dat de Europese statistiekbureaus onafhankelijk moeten zijn. Deze Praktijkcode is tot stand gekomen naar aanleiding van eerdere problemen met de Griekse begrotingsstatistieken. De achtergrond van de Europese aandacht voor statistieken is het toegenomen administratieve gebruik ervan voor Europese doeleinden. Zo wordt bijvoorbeeld de bijdrage van de EU-landen aan de EU-begroting gebaseerd op het bruto-binnenlands product van de landen, dat berekend wordt door de nationale statistische bureaus. Ook in het economische en financiële beleid spelen statistische gegevens een belangrijke rol, zoals in de normen voor het begrotingstekort en de inflatie. Daarom noemt de Statistiekverordening van de Europese Unie (Europese Unie, 2009) als statistische beginselen: professionele onafhankelijkheid, onpartijdigheid, objectiviteit, betrouwbaarheid, geheimhouding van individuele gegevens en kosteneffectiviteit. Meer specifiek schrijft ze onder andere voor dat 'beroepsregels en ethische normen moeten worden nageleefd' en dat 'bij de keuze van bronnen, methoden en procedures, wetenschappelijke criteria moeten worden toegepast'.

Naleving van de Praktijkcode wordt serieus genomen. Zo wordt ongeveer eens per vijf jaar bij elk statistisch bureau in de EU in een *peer review* onderzocht welke verbeteringen er nodig zijn om aan de Praktijkcode te voldoen. Verder zijn er regelmatig Europese missies, van het Europees Statistisch Stelsel,

dat bestaat uit Eurostat en de nationale statistische bureaus in de EU, en van de Europese Rekenkamer. Hierin wordt bij een statistisch bureau de kwaliteit van specifieke thema's zoals de macro-economische of inflatiestatistieken bekeken. Eurostat heeft ook in het openbaar stelling genomen in de 'Griekse kwestie' (Eurostat, 2011). In feite zijn de toetsen door Eurostat de voornaamste kwaliteitsonderzoeken voor de statistische bureau's in de EU. Samen met de eigen professionaliteit van de statistici en van de gebruikers waarborgen ze de kwaliteit van de publieke statistiek.

Onafhankelijkheid

In Nederland hebben situaties zoals in Griekenland en Argentinië zich nooit voorgedaan. De overheid heeft de onafhankelijkheid van de statistiek altijd uitgedragen. Maar daarnaast is het belangrijk dat statistici begrijpen hoe ze in bepaalde situaties moeten handelen. Daarom heeft het International Statistical Institute – de internationale vereniging van statistici – de *Declaration on Professional Ethics* (International Statistical Institute, 1985) opgesteld, waarin professionele waarden en ethische beginselen voor statistici worden geformuleerd; het CBS kent een *Gedragcode* (Centraal Bureau voor de Statistiek, 2003) die daarop is gebaseerd. Op basis van deze statistische principes heeft het ISI herhaaldelijk in het openbaar stelling genomen in de Griekse en Argentijnse kwesties (International Statistical Institute, 2011; 2012).

Het is ook belangrijk om te beseffen dat onafhankelijkheid niet zomaar vanzelf komt en blijft. Politiek en wetten weerspiegelen nu eenmaal meningen in de samenleving, en die kunnen veranderen. De *Fundamental Principles of Official Statistics* (Statistical Commission of the United Nations, 1994), opgesteld door de Statistische Commissie van de Verenigde Naties, zeggen: *Official statistics provide an indispensable element in the information system of a democratic society*. Het bestaansrecht van statistiek is niet alleen

maar technocratisch, statistiek is niet alleen maar van belang om beleid te funderen, maar de onafhankelijkheid van de statistiek is juist nodig voor het publieke debat in een democratische samenleving. We kunnen dat samenvatten als 'Statistiek is een democratisch recht'.*

* Zo geformuleerd door mijn CBS-collega, Symon Algera.

LITERATUUR

- Centraal Bureau voor de Statistiek (2003). *Gedragcode*. <<http://www.cbs.nl/NR/rdonlyres/7D28C075-E6F7-4DDB-BE56-7BEC81B3BFBB/o/gedragcode.pdf>>.
- Eurostat (2011). *Communication from Eurostat to the National Statistical Institutes of the EU regarding the matter of the 2009 fiscal deficit of Greece*. <http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/partners/european_union/el/tab_news>.
- Europese Unie (2005). *Praktijkcode voor Europese Statistieken (Code of Practice for European Statistics)*. Herzien in 2011 <http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/code_of_practice>.
- Europese Unie (2009). *Verordening betreffende de Europese statistiek (Regulation on European statistics)*. Publicatieblad van de Europese Unie L 87, 31 maart 2009, pp. 164–173.
- International Statistical Institute (1985). *Declaration on Professional Ethics*. Herzien in 2010. <<http://www.isi-web.org/about-isi/professional-ethics>>.
- International Statistical Institute (2011). *ISI's Position on Methodological Concerns about the Argentine CPI & the Position of Statisticians in Argentina*. <<http://www.isi-web.org/about-isi/professional-ethics>>.
- International Statistical Institute (2012). *Letter by the ISI President to the President of the Hellenic Statistical Authority*. <<http://www.isi-web.org/about-isi/professional-ethics>>.
- Statistical Commission of the United Nations (1994). *Fundamental Principles of Official Statistics*. <<http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>>.
- Wet op het Centraal Bureau voor de Statistiek*. Staatsblad 2003, nr. 516. <<http://wetten.overheid.nl/BWBR0015926>>.

KEES ZEELBERG is directeur Methoden & Statistisch Beleid bij het Centraal Bureau voor de Statistiek.
E-mail: <k.zeelenberg@cbs.nl>.

DE MENSELIJKE KANT VAN STATISTIEK

JELTE M. WICHERTS

Naïeve lesboeken

Veel lesboeken in de statistiek lijken te zijn geschreven met een aan naïviteit grenzend beeld van hoe gedragswetenschappers statistiek gebruiken. Zo omschrijven Agresti en Franklin (2007) het gebruik van een eenzijdige toets nadat men de resultaten al heeft gezien als 'cheating [and] unethical behavior'. Wat zouden Agresti en Franklin eigenlijk vinden van een onderzoeker die besluit om iets meer gegevens te verzamelen nadat in een eerdere toets de nulhypothese (bij $\alpha=0,05$) niet verworpen kon worden? Ik weet niet of Agresti en Franklin dit gedrag ook als vals spelen zouden omschrijven want ze bespreken het niet. In de wandelgangen van menig psychologieafdeling heet dit gedrag gewoon 'even bijdraaien'. Andere trucs om de p-waarde onder de heilige grens van 0,05 te krijgen (zie bijv. Simmons, Nelson, & Simonsohn, 2011) zijn onder meer het selectief gebruik van uitkomstmaten, spelen met covariaten, en tamelijk subjectieve bepaling van welke proefpersonen niet in de analyse worden meegenomen omdat ze uitbijters zijn of omdat ze zich niet aan de experimentele instructie (lees: onderzoekshypothese) hebben gehouden. Recent groot-

schalig vragenlijstonderzoek laat zien dat dergelijke praktijken onder onderzoekers eerder regel dan uitzondering zijn (John, Loewenstein, & Prelec, 2012).

Wetenschap is mensenwerk en onderzoekers die deze trucs gebruiken zijn geen fraudeurs. Onderzoekers hebben verwachtingen en zullen dus oprecht denken dat er iets mis is met hun analyseplan als dit niet de verwachte resultaten oplevert. Daarnaast worden onderzoekers geconfronteerd met bikkelharde *peer reviewers* die hun artikel voor publicatie in een vaktijdschrift tegenhouden als er niet ten minste een aantal p-waarden van onder de .05 in staan (Mahoney, 1977). Het is de hoogste tijd dat we in het onderwijs en het onderzoek wat meer aandacht besteden aan de menselijke kant van de statistiek. Veel te lang is namelijk gedaan alsof onderzoekers zich houden aan de normatieve opzet van statistische toetsen terwijl ze in werkelijkheid veel zaken onjuist interpreteren (Hoekstra, Finch, Kiers, & Johnson, 2006), vrij opportunistisch zijn (Hoekstra, Kiers, & Johnson, 2012; John et al., 2012), en vaak fouten maken (Bakker & Wicherts, 2011).

Oeps!

Stelt u zich een onderzoeker voor die verwacht dat stress en motivatie de cognitieve prestaties beïnvloeden. Hij heeft een experiment uitgevoerd waarbij proefpersonen een cognitieve test maakten onder condities die verschilden in stress en motivatie (zeg, door middel van het opvoeren van tijdsdruk en het beloven van een financiële beloning voor goede prestaties). De gegevens zijn verzameld en de onderzoeker voert een variantieanalyse uit met de factoren sekse, motivatie, stress en met leeftijd als covariaat. Zijn SPSS-output staat in figuur 1.

Onze onderzoeker is laaiend enthousiast over deze fantastische resultaten (let wel: ik heb deze resultaten gefabriceerd) en begint direct met het schrijven van een onderzoeksartikel waarin hij het volgende rapporteert: 'As expected, the main effect of stress was significant: $F(1,96) = 3,59, p < 0,05$ '. Ik weet niet of u de fout al heeft ontdekt, maar een blik op figuur 1 laat duidelijk zien hoe foutengevoelig SPSS eigenlijk is. In de output staan allerlei redundante gegevens en de kolommen

en rijen worden eenvoudig verward. Vergeet niet dat onze onderzoeker een sterke verwachting had over de onderzoeksuitkomsten en dat dit de kans op selectieve waarnemingsfouten vergroot.

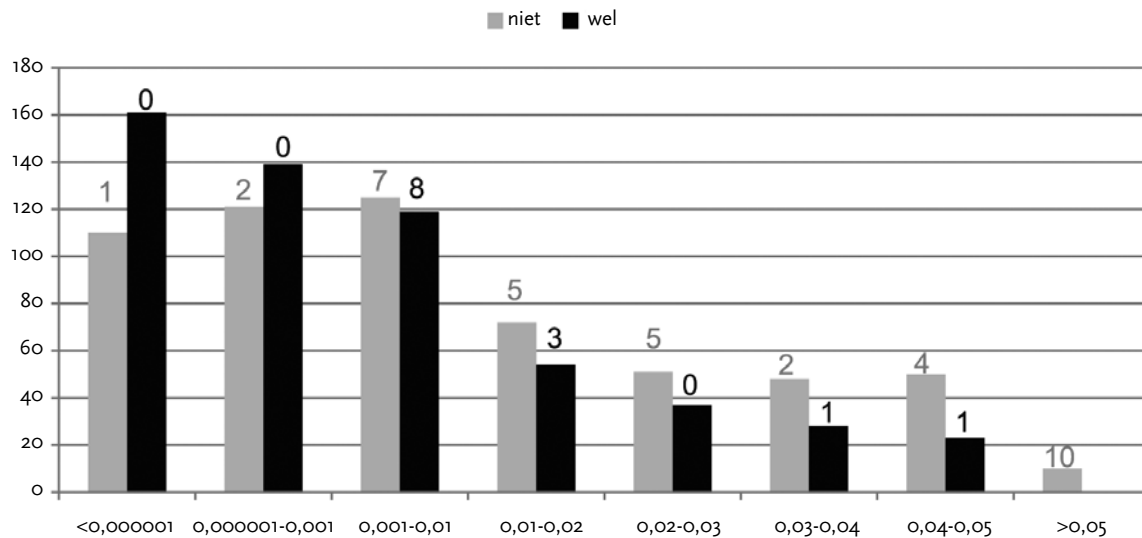
Nu kun je eenvoudig nagaan dat het gerapporteerde resultaat ' $F(1,96) = 3,59, p < 0,05$ ' niet deugt door de p-waarde te berekenen die hoort bij de F verdeling met genoemde vrijheidsgraden. Marjan Bakker en ik hebben deze berekening uitgevoerd bij 4720 toetsresultaten uit 257 psychologische onderzoeksartikelen en vonden in 128 van die artikelen (49,8%) ten minste een resultaat waarbij de gerapporteerde p-waarde niet gerepliceerd kon worden (Bakker & Wicherts, 2011). Nu waren veel van die gevonden discrepanties niet erg groot (bijv. een $p=0,0012$ gerapporteerd als $p<0,001$) maar een niet onaanzienlijk deel van de discrepanties was wel degelijk relevant (althans binnen de regels van nulhypothese significantietoetsing). In 39 artikelen vonden wij ten minste een resultaat dat rond de 0,05 een andere uitkomst opleverde dan wij naberekend hadden. In bijna een op zeven artikelen in de door ons bekeken psychologische literatuur zat (niettegen-

Dependent Variable:score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	388.662 ^a	8	48.583	5.285	.000	.306
Intercept	88.794	1	88.794	9.660	.002	.091
age	101.520	1	101.520	11.044	.001	.103
mot	76.726	1	76.726	8.347	.005	.080
sex	3.642	1	3.642	.396	.531	.004
stress	33.011	1	33.011	3.591	.061	.036
mot * sex	43.203	1	43.203	4.700	.033	.047
mot * stress	41.832	1	41.832	4.551	.035	.045
sex * stress	.426	1	.426	.046	.830	.000
mot * sex * stress	8.137	1	8.137	.885	.349	.009
Error	882.423	96	9.192			
Total	11002.910	105				
Corrected Total	1271.086	104				

a. R Squared = .306 (Adjusted R Squared = .248)

Figuur 1. Een output uit SPSS van een ANCOVA op grond van gefabriceerde gegevens



Figuur 2. Verdeling van de p-waardes die als significant werden gepresenteerd in artikelen waarvan wel (in zwart) of niet (in grijs) de gegevens werden gedeeld. Boven elke staaf staat het aantal gevonden inconsistenties in de rapportage van toetsresultaten (bron: doi:10.1371/journal.pone.0026828.g002)

staande die kritische reviewers) wel een toetsresultaat dat als significant werd gepresenteerd maar dat niet leek te zijn (ten minste als we de gerapporteerde toetsingsgrootheid en de vrijheidsgraden mogen geloven).

Geheimzinnigheid?

Nu schrijft de onderzoeksethiek voor dat onderzoekers na publicatie van een onderzoeksartikel hun onderliggende gegevens moeten vrijgeven voor heranalyse door vakgenoten. Als je een artikel weet te publiceren in een vaktijdschrift van de *American Psychological Association* (APA) moet je hier zelfs een formulier voor ondertekenen. Mijn collega's en ik vroegen in 2005 met een vriendelijke e-mail gegevens op uit 141 artikelen die anderhalf jaar eerder waren gepubliceerd in een van vier toptijdschriften van de APA. Hoewel weinig onderzoekers het delen van de gegevens echt weigerden (al is het bizarre antwoord '*I am afraid your request is not possible*' mij altijd bijgebleven), gaven veel onderzoekers aan dat ze de gegevens uit het oog waren verloren of dat ze gewoonweg de tijd niet had-

den om de gegevens bij elkaar te zoeken en goed te documenteren. Een onderzoeker beloofde de data binnen enkele dagen te sturen maar laat ons inmiddels al zo'n 2595 dagen wachten. Nu waren er ook onderzoekers die de gegevens uitstekend gedocumenteerd hadden en deze binnen korte tijd met ons deelden, maar zulke gewetensvolle onderzoekers waren in de minderheid. Na verschillende reminders kwam 27% van de onderzoekers uiteindelijk met gegevens over de brug (Wicherts, Borsboom, Kats, & Molenaar, 2006).

Mogelijke verklaringen voor deze magere vangst zijn dat veel onderzoekspsychologen slordig met hun gegevens omspringen, of dat ze bang zijn dat analysefouten worden ontdekt of dat een kleine aanpassing van de analyse een andere kijk op de resultaten oplevert. We besloten daarom eens terug te gaan naar (een deel van) de artikelen waarvan we ooit gegevens hadden opgevraagd. We herberekenden alle 1148 p-waardes van t, F, of chi-kwadraat toetsen die in die artikelen waren gerapporteerd als zijnde $p < 0,05$. Onze resultaten (Wicherts, Bakker, & Molenaar, 2011) staan weergegeven in figuur 2. We vonden aanzienlijk meer fouten in de rapportage van de toetsresultaten in de

28 artikelen (met 587 toetsresultaten) waarvan geen gegevens werden gedeeld dan in de 21 artikelen (met 561 toetsresultaten) waarvan we wel de gegevens hadden. Vooral interessant waren de 10 toetsresultaten uit zeven artikelen die als $p < 0,05$ waren gerapporteerd maar dat niet leken te zijn. Spijtig genoeg waren gegevens van die artikel niet (meer) beschikbaar. Daar kwam nog eens bij dat de verdeling van de p-waardes tussen beide type artikelen verschilde; de p-waardes van delende onderzoekers waren gemiddeld lager dan de p-waardes van niet-delende onderzoekers. Vooral interessant voor heranalyse zijn de resultaten die *net* onder de grens van 0,05 vallen. Maar uiterekend die resultaten werden niet gedeeld. Kortom: naarmate een onafhankelijke heranalyse relevanter wordt, is de kans dat gegevens worden gedeeld voor zo'n heranalyse kleiner.

Hoe het beter kan

Deze resultaten stemmen weliswaar niet vrolijk maar willen ook weer niet zeggen dat een groot deel van de psychologische onderzoekers de kluit bedondert. Het gaat eigenlijk al fout bij het statistiekonderwijs. Zo schenken Agresti en Franklin en veel andere statistiekboekenschrijvers nauwelijks aandacht aan hoe gegevensbestanden netjes moeten worden gedocumenteerd. De oplossing is wat mij betreft simpel. Het moet standaard worden dat onderzoekers in een onderzoeksgroep (of collega's) elkaars analyses repliceren en dus controleren. In het zogenaamde co-pilootmodel van co-auteurschap heeft elke coauteur het gegevensbestand en draaien ten minste twee onderzoekers de analyses daarvan. Dit voorkomt niet alleen veel fouten, maar maakt de kans op frauderen of verfraaien van de resultaten aanzienlijk kleiner. Daar komt bij dat de onderzoeker de gegevens dan al klaar heeft gemaakt (middels documentatie etc.) om te worden gedeeld met vakgenoten *na* publicatie. Uiteindelijk horen die gegevens zelf (daar waar het kan met oog op privacy en eigendomsrechten) gepubli-

ceerd te worden met het artikel of bijvoorbeeld in een online *data repository*. Onderzoeksgegevens zijn een stuk interessanter dan een enkele p-waarde en moeten uiteindelijk deel gaan uitmaken van het publiceren van onderzoeksresultaten.

LITERATUUR

- Agresti, A., & Franklin, C. (2007). *Statistics. The art and science of learning from data*. Upper Saddle River, NJ, US: Pearson Education, Inc.
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43, 666–678. doi: 10.3758/s13428-011-0089-5
- Hoekstra, R., Finch, S., Kiers, H. A. L., & Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of p values. *Psychonomic Bulletin & Review*, 13, 1033–1037.
- Hoekstra, R., Kiers, H., & Johnson, A. (2012). Are assumptions of well-known statistical techniques checked, and why (not)? *Frontiers in Quantitative Psychology and Measurement*, 3, 137. doi: 10.3389/fpsyg.2012.00137
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524–532. doi: 10.1177/0956797611430953
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1, 161–175.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366. doi: 10.1177/0956797611417632
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6, e26828. doi: 10.1371/journal.pone.0026828
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726–728. doi: 10.1037/0003-066X.61.7.726

JELTE M. WICHERTS is als universitair hoofddocent verbonden aan het departement Methoden en Technieken van de Tilburg School of Social and Behavioral Sciences van Tilburg University.
E-mail: <j.m.wicherts@uvt.nl>

STATISTIEK, DATA MONITORING EN ETHIEK

KIT C. B. ROES

In de zomer van 2010 lanceerde de Nederlandse Federatie van Universitair Medisch Centra (NFU) de beleidsnota *Kwaliteit van mensgebonden onderzoek*. Dit beleid beoogt de kwaliteit van onderzoekergeïnitieerd klinisch onderzoek naar een hoger plan te brengen. Het raakvlak tussen statistiek en ethiek wordt belangrijker, maar ook complexer. In toenemende mate wordt gebruik gemaakt van *Data Monitoring Committees*. Dit geeft uitdagingen voor de statistiek en statistici, die om ontwikkeling van en discussie binnen ons vak vragen.

Schaalgrootte van klinisch onderzoek

Medisch wetenschappelijk onderzoek met mensen is essentieel om onze kennis te vergroten en de grenzen van nu beschikbare therapieën te verleggen. De hoeveelheid klinische studies is onvoorstelbaar groot. Op de website <www.clinicaltrials.gov> stonden op 17 oktober 2012 134.268 studies met locaties in 180 landen geregistreerd (rond 40% industrie gesponsord en 60% niet-industrie gesponsord). Het is een veilige schatting dat miljoenen mensen wereldwijd participe-

ren in klinische studies. Deze schaal, het ingrijpende karakter voor patiënten en het feit dat niet-industrie gesponsorde studies duidelijk de overhand hebben leggen een grote verantwoordelijkheid bij de verantwoordelijke academische centra.

Kwaliteit van mensgebonden onderzoek

Op basis van het Propatria onderzoek hebben de Inspectie voor de Gezondheidszorg (IGZ) en de Centrale Commissie Mensgebonden Onderzoek (CCMO) geconstateerd dat voor het academisch mensgebonden onderzoek de kwaliteitsborging naar een hoger plan moet. Onlangs gaf de NFU daartoe het boekje *Kwaliteitsborging van mensgebonden onderzoek* uit. Belangrijke kernpunten uit het beleid zijn opleiding van alle klinisch onderzoekers, een risico-

gebaseerd onafhankelijk toezicht op de uitvoering van het onderzoek en een beter gebruik van onafhankelijke *Data Monitoring Committees*. Het opleiden van onderzoekers is landelijk in volle gang: er zijn al meer dan 4500 klinisch onderzoekers van de 8 deelnemende UMC's gecertificeerd.

Risico-gebaseerd kwaliteitstoezicht

Het voorgestelde toezicht op de kwaliteit van de uitvoering is *risico-gebaseerd*. Er zijn criteria aangegeven waarmee onderzoekers kunnen bepalen of hun onderzoek in de categorie verwaarloosbaar, minimale overschrijding van verwaarloosbaar, matig of hoog risico valt, op basis van de ingeschatte kans op schade en de ernst van de schade (tabel 1). Dit heeft betrekking op het risico dat door het onderzoek wordt toegevoegd

GROOTTE VAN KANS	MATE VAN SCHADE	LICHTE SCHADE	MATIGE SCHADE	ERNSTIGE SCHADE
		KLEINE KANS	Verwaarloosbaar risico	Minimale overschrijding van verwaarloosbaar risico
MATIGE KANS	Minimale overschrijding van verwaarloosbaar risico	Matig risico	Hoog risico	
GROTE KANS	Matig risico	Hoog risico	Hoog risico	
KWETSBARE PATIËTENGROEP		Opwaarderen naar hoger risico	Opwaarderen naar hoger risico	

Tabel 1. Risicoclassificatie in relatie tot de kans op schade, de ernst van die schade en de kwetsbaarheid. Uit: *Kwaliteitsborging mensgebonden onderzoek*, NFU, Juni 2010.

aan de eventuele belasting en gevaren die al kleven aan de reguliere behandeling. De lokale of centrale medisch-ethische toetsingscommissie toetst en stelt de risicoclassificatie vast. De risicoclassificatie bepaalt de intensiteit van het kwaliteitstoezicht gedurende de uitvoering van het onderzoek.

Al het klinische onderzoek zal in principe enige vorm van *monitoring* moeten hebben, al kan dat minimaal zijn. Het gebeurt onafhankelijk van de betrokken onderzoekers. Bij een onderzoek van de risicoclassificatie 'matig' of 'hoog' is er intensiever toezicht op het totale proces, bij voorkeur onafhankelijk van de instelling waar het onderzoek plaatsvindt. Dit monitoren is een arbeidsintensief proces, met nauwkeurige controle van brondocumenten, studie van documentatie en gevolgde procedures en accurate rapportage van (ernstige) voorvallen. De monitoring maakt deel uit van een kwaliteitssysteem: instellingen dienen (conform Good Clinical Practice) te beschikken over Standard Operating Procedures (SOPs) die het proces van onderzoek doen voorschrijven. Voor geneesmiddelenonderzoek is dat al sinds begin jaren negentig van de vorige eeuw het geval. Voor onderzoekergeïnitieerd onderzoek is het betrekkelijk nieuw. Door de grote arbeidsintensiteit zijn de kosten relatief hoog en veel financierende instanties (NWO/ZonMW) voorzien daar niet in. Deze lacune vraagt serieuze aandacht om innovatief onderzoek dat niet door de farmaceutische industrie wordt gesponsord te kunnen laten plaatsvinden. Het is noodzakelijk voor het stimuleren van een sterke, onafhankelijke positie van de academische onderzoekscentra ten opzichte van het onderzoek dat wel door die industrie wordt gesponsord.

Data Monitoring Committees

Volgens het NFU-beleid moet voor studies die geclassificeerd zijn als 'hoog risico' gebruik worden gemaakt van onafhankelijke *Data Monitoring Committees* (DMC), ook wel *Data and Safety Monitoring Committees* genoemd. Ook hier komt de term 'monitoring' in voor,

maar dit is wezenlijk anders dan de kwaliteitsgerichte monitoring die hiervoor beschreven staat. Klinische studies hebben vaak een looptijd van jaren, en in die tijd zijn er al veel gegevens beschikbaar. Het kan van groot belang zijn dat tussentijds al resultaten met betrekking tot effectiviteit en veiligheid worden geëvalueerd. Als er al eerder overtuigend bewijs is dat de behandeling effectief is, is het verantwoord te stoppen en de behandeling ter beschikking te stellen. Of er kunnen zich veiligheidsproblemen aftekenen, die doorgaan niet ethisch maken. Het is van belang dat een dergelijke tussentijdse analyse het verloop van de studie niet beïnvloedt (als de studie gewoon door moet gaan). Daarom worden dergelijke tussentijdse analyses uitgevoerd door een DMC (Ellenberg et al., 2002; Sydes et al., 2004), bestaande uit wetenschappers die onafhankelijk zijn, i.e. geen belangenverstrengeling hebben ten aanzien van het betreffende onderzoek. De DMC brengt advies uit aan de studieleiding (de 'verrichter' volgens de Wet medisch-wetenschappelijk onderzoek met mensen, en 'sponsor' volgens *good clinical practice*), echter zonder de tussentijdse resultaten prijs te geven. Dit advies betreft de veiligheid van de deelnemers, van nog te rekruteren deelnemers alsook de wetenschappelijke waarde van het voortzetten van de studie. De DMC bestaat uit klinische wetenschappers en een statisticus, die gezamenlijk op basis van een goede wetenschappelijke beoordeling tot een advies aan de studieleiding moeten komen.

De DMC zal zich doorgaans (mede) laten leiden door statistische criteria. Die zijn nodig, omdat de herhaalde analyses zoals bekend leiden tot inflatie van de type I fout (significantieniveau). De te hanteren (groep) sequentiële criteria die de type I fout op het gewenste niveau houden worden in het studieprotocol beschreven, en de statisticus voert de analyses op de geplande tijd uit. Enige achtergrond van de onderliggende statistische beginselen is recent in *STAtOR* gegeven door van der Tweel (2010). De statistische criteria zijn vastgelegd in het studieprotocol en kunnen het advies bepalen: doorgaan of stoppen. Een computer zou dan 'de was kunnen doen'. In werkelijk-

Meeting	Subject-jaren	PROGRESSIE NAAR ARC/ AIDS/STERFTE		STERFTE		P-WAARDEN	
		Onmiddellijk	Uitgesteld	Onmiddellijk	Uitgesteld	Progressie	Sterfte
1	-	-	-	-	-	-	-
2	?	1	4	0	0	0,220	-
3	443	7	13	0	0	0,190	-
4	-	-	-	-	-	-	-
5	920	24	34	0	0	0,190	-
6	1711	56	74	14	7	0,120	0,130
7	2181	76	104	25	14	0,037	0,081
8	3042	114	148	29	24	0,036	0,500
9	-	-	-	-	-	-	-
10	4258	202	227	59	45	0,230	0,110
11	5328	263	284	95	76	0,370	0,150

Tabel 2. Data beschikbaar voor de Concorde DMC (Armitage, 1999)

heid is het niet zo simpel. Dat heeft te maken met de totaliteit van data en informatie die een DMC in haar overwegingen mee moet nemen, zoals het volgende voorbeeld laat zien. De Concorde-studie (Armitage, 1999) betrof een gerandomiseerde klinische studie onder mensen die HIV besmet waren, maar nog symptoomvrij. In deze studie werd voor die groep onmiddellijke behandeling met zidovudine (AZT) vergeleken met uitgestelde behandeling. Uitgesteld hield in dat deze groep eerst placebo ontving, tot het moment van optreden van symptomen: AIDS, AID-gerelateerd complex (ARC), of voortdurende lage CD4 tellingen. De belangrijkste interim resultaten gedurende de studie zagen er als volgt uit (tabel 2). Uit de tabel lezen we tussentijds trends af, maar in tegengestelde richting: onmiddellijke behandeling van symptoomvrije HIV-patiënten lijkt een voordeel op te leveren wat betreft uitstellen van progressie (primaire eindpunt),

maar er is een suggestie van nadeel met betrekking tot sterfte (rode data).

De DMC heeft in dit geval de moed gehad door te gaan, waardoor we de betekenis van variatie, *nature's guessing game* kunnen zien: een trend die zich lijkt af te tekenen zet zich niet vanzelfsprekend door. De DMC heeft het besluit tot doorgaan niet licht genomen. In aanvulling op de vooraf gedefinieerde criteria zijn diverse additionele analyses uitgevoerd om na te gaan hoe robuust de tussentijdse resultaten zijn en in hoeverre andere factoren, zoals nog ontbrekende gegevens en patiëntkarakteristieken, het resultaat kunnen beïnvloeden.

Statistiek en ethische aspecten

Statistici hebben in onderzoek een vakinhoudelijke

grote verantwoordelijkheid en refereren regelmatig naar de ethische aspecten van hun bijdrage in het onderzoeksontwerp. Door de betrokken statisticus wordt zelden precies duidelijk gemaakt vanuit welke ethische principes hij handelt. De twee belangrijkste dilemma's zijn:

1. de rechtvaardiging van het randomiseren naar verschillende behandelingen (inclusief placebo);
2. welke mate van bewijs voldoende is om een behandeling de voorkeur te geven (waarna randomiseren naar de inferieure behandeling ethisch niet meer acceptabel zou zijn).

Het (impliciet) onder statistici meest gangbare ethische perspectief lijkt dat van afwegen van individuele versus collectieve ethiek: wat is het beste voor de patiënten in de studie versus wat is het beste voor toekomstige patiënten – die van de resultaten zullen kunnen profiteren (Palmer en Rosenberger, 1999). Daarbij lijkt het mogen randomiseren tussen twee (actieve) behandelingen consistent met het meer gewicht toekennen aan collectieve ethiek (Pullman en Wang, 2001), dan wel met de unieke toestand van *equipoise*:

een staat van onverschilligheid of onenigheid in de medische professionele gemeenschap over de medische procedure voor de te onderzoeken aandoening die alles overwegende zowel de voorkeur verdient, als valt binnen de medisch professionele standaard (verbeterde definitie, van der Graaf, 2010).

Het is goed te bedenken dat deze afwegingen niet altijd geheel bij de onderzoeker liggen. Het kan een inrichtingsvraagstuk zijn (Senn, 2002): is onze maatschappij beter af als wij ons voor de voortgang van de medische kennis in belangrijke mate baseren op *randomized controlled trial's* (RCTs) – en dus randomiseren moeten toestaan? Dit is voor geneesmiddelenonderzoek bij wet zo geregeld. Een nieuw geneesmiddel komt in de regel pas beschikbaar voor patiënten buiten het kader van onderzoek als in meerdere RCTs werkzaamheid en veiligheid is bewezen. Zo gesteld,

slaat voor geneesmiddelenonderzoek de individuele versus collectieve ethiek door naar de laatste, waarbij de individuele klinisch onderzoeker slechts kan beslissen of hij in het onderzoek wil participeren.

DMC, statistiek, ethiek en advies

Een DMC ontkomt er niet aan bij tussentijdse evaluatie en advisering ethische perspectieven mee te wegen. Die perspectieven zijn, ondanks de vele beschikbare technieken, (nog) niet volledig te vangen in van tevoren vastgelegde statistische criteria. Een blik op voorbeelden die recent aandacht hebben getrokken geeft inzicht waar uitdagingen kunnen liggen. Het Propatria onderzoek (Besselink et al., 2008) is uitgebreid beschreven, ook door statistici (Gill, 2009); dit zal ik hier niet herhalen. De betreffende studie is afgerond, waarbij bij uiteindelijk 24 patiënten van de 152 behandeld met probiotica zijn overleden, tegen 9 van de 144 behandeld met placebo. Op grond van het nader verrichte onderzoek (IGZ, CCMO en VWA, 2009) is vastgesteld dat er geen oorzakelijk verband gelegd kan worden tussen de geconstateerde tekortkomingen van de studie en de sterfgevallen in de studie. Ook is er geen oorzakelijk verband aangetoond met de behandeling. Een belangrijke vraag die gesteld kan worden is: had het onderzoek al eerder stopgezet moeten worden, op advies van de DMC? In lijn met Gill (2009), zou vanuit statistisch perspectief gesteld kunnen worden dat de interim uitkomst (na evaluatie van 184 van de geplande 300 patiënten) consistent was met een advies de studie te stoppen vanwege 'futiliteit' op het primaire eindpunt: *infectueuze complicaties*. Het verschil in percentage infectueuze complicaties tussen probiotica en placebo was 5% ten nadele van probiotica op dat punt in de studie. De DMC had het dus mogelijk eenvoudig: op basis van de vooraf geformuleerde criteria – mits juist en gedeblindeerd toegepast – volgt een advies aan de studieleiding om te stoppen. De tussentijds zichtbare, zij het zwakke, trend van hogere mortaliteit in de probiotica groep

hoefde daarin niet eens een doorslaggevende rol te spelen. Is daar de kous mee af? Dat is zeer de vraag.

Probiotica voor deze groep ernstige patiënten was experimenteel en dus ‘niet beschikbaar’ als behandeling voor patiënten buiten dit onderzoek, ook niet onmiddellijk na een positief resultaat. Bij eventueel tussentijds stoppen vanwege futiliteit waren mogelijke consequenties in sterfte op zijn best zwak zichtbaar. Tussentijds stoppen had randomisatie en behandeling van naar schatting ongeveer 70 patiënten ‘voorkomen’ (rekrutering was gespreid over 3 jaar, eindpunt na follow-up van 3 maanden, gegevens beschikbaar en geanalyseerd door DMC kost ook enige tijd). De eerste ethische vraag ligt binnen de studie: zouden deze 70 patiënten beter af geweest zijn, als de studie gestopt was en zij buiten het kader van onderzoek behandeld waren? Over de twee groepen gezamenlijk was de mortaliteit lager dan verwacht, een effect dat vaker in klinische studies gezien wordt. Dit effect wordt wel toegeschreven aan selectiebias, maar ook aan de intensievere begeleiding die het gevolg is van het uitvoeren van de studie. Kortom, het is niet vast te stellen of dit het geval is.

Een tweede ethische punt ligt buiten de studie. Voor de medische gemeenschap resteert bij tussentijds stoppen een grotere residuele onzekerheid over potentiële effecten van probiotica. Op basis van de interim-resultaten is natuurlijk niet aangetoond dat probiotica niet werkt, en er is ook niet aangetoond dat het onveilig is (mortaliteit). In dergelijke gevallen dient er rekening mee gehouden te worden dat nieuwe klinische studies kunnen volgen. Voor veel effectieve en waardevolle behandelingen zijn immers meerdere studies gedaan, waarvan een aantal geen effect hebben laten zien. We ontwerpen studies zo dat dit voor de prima werkzame behandelingen naar verwachting in tenminste 20% van de studies het geval is. De DMC moet zich voor haar verantwoordelijkheid naar toekomstige studiedeelnemers ervan vergewissen dat het tussentijdse bewijs afdoende is om de onderzoekende gemeenschap te weerhouden een volgende studie te starten. Gebeurt dit wel, dan kan het netto effect van

eerder stoppen zijn dat juist meer mensen worden blootgesteld aan een experimentele behandeling die niet werkt, en mogelijk niet veilig is. Het is immers denkbaar dat juist een grotere studie wordt opgezet, omdat men bescheidener verwachtingen heeft over het positieve effect.

Het tweede voorbeeld is de studie van rivastigmine versus placebo voor de behandeling van post-operatief delirium bij intensive-care-patiënten (Van Eijk et al., 2010). Delirium is ernstige verwardheid, die het algemene herstel kan bemoeilijken. In de dubbelblinde, gerandomiseerde studie werd rivastigmine vergeleken met placebo – beide gegeven aanvullend op bestaande standaardzorg voor het delirium. De primaire uitkomst in deze studie was de duur van het delirium gedurende opname in de IC. De geplande omvang was 440 patiënten. De ingestelde DMC evalueerde veiligheid elke drie maanden. Vroegtijdig stoppen vanwege overtuigende effectiviteit werd niet voorzien.

Na inclusie van 104 patiënten (54 rivastigmine, 50 placebo) adviseerde de DMC de studie te stoppen omdat de sterfte in de rivastigmine-groep (12 patiënten, 22%) hoger lag dan in de placebo-groep (4 patiënten, 8%; sequentiële analyse: $p=0,07$). De duur van delirium was langer in de rivastigmine-groep (mediaan 5,0 dagen) dan in de placebo-groep (mediaan 3,0 dagen, $p=0,06$ – tweezijdig, ongecorrigeerd). De situatie is voor de DMC nu wezenlijk anders. Het protocol bevat een richtlijn waarop de DMC voortijdig stoppen vanwege veiligheid moet overwegen; die was inderdaad overschreden in dit geval. Rivastigmine is een beschikbare, geregistreerde behandeling voor symptomatische behandeling van Alzheimer en andere vormen van dementie. Het wordt – op basis van beschreven case series en kleine studies – door artsen voorgeschreven voor precies de patiëntengroep en aandoening in het onderzoek. In haar besluit kan de DMC daarom niet alleen naar de richtlijn voor veiligheid kijken. Conclusies hebben in dit geval onmiddellijk effect voor de huidige medische praktijk en er is, bij eenzijdige evaluatie, kans dat effectieve behandeling

voor deze ernstige patiënten wordt gestopt. In haar overwegingen moesten dus de tussentijdse resultaten op effectiviteit nadrukkelijk worden meegewogen. De (negatieve) resultaten zijn gepubliceerd, maar ook via andere kanalen naar de beroepsgemeenschap gecommuniceerd, om zeker te stellen dat ernaar gehandeld zou worden. Ook hier is er geen oorzakelijk verband aangetoond tussen de sterfgevallen en de behandeling. Het advies is in dit geval gebaseerd op de totale afweging van het vermoedelijk ontbreken van voordelen (op de duur van het delirium) en potentieel risico in de studie.

Discussie

Statistici zijn veel en nauw betrokken bij het opzetten van klinische studies, inclusief het vaststellen van statistische criteria voor interim-analyses voor effectiviteit en veiligheid. De positieve impact is al groot en kan verder groeien. *Data Monitoring Committees* vervullen een wetenschappelijke en ethische rol in klinische studies. Een statistische gereedschapskist is beschikbaar en met de ontwikkeling van adaptieve ontwerpen groeit de inhoud van die kist exponentieel. Een DMC kan niet zonder 'echte' statisticus en hetzelfde geldt ook voor medisch ethische toetsingscommissies, inclusief de CCMO. Zowel voor de opzet als voor de tussentijdse beoordeling die een DMC doet, is gedegen statistische input onmisbaar voor de wetenschappelijke beoordeling van de data en het daaruit voortvloeiende advies. Maar ook voor ons vakgebied zullen we verder moeten ontwikkelen. Ik durf te stellen dat we de kunst van het ontwerpen van de studies en criteria, op een dusdanige wijze dat die adequaat rekening houden met de gevolgen buiten de studie, nog niet onder de knie hebben. Dit blijkt ook uit recente ethische overwegingen voor *adaptive clinical trials* (Van der Graaf et al., 2012). Een onderzoeksuitdaging en opleidingsuitdaging voor toekomstige en huidige klinische (DMC-)statistici: gegeven de ingezette ontwik-

kelingen zullen we er veel nodig hebben. Ik zou meer discussie binnen ons vakgebied toejuichen en kijk dan ook uit naar reacties.

LITERATUUR

- Armitage, P. on behalf of the Concorde and Alpha Data and Safety Monitoring Committee (1999). Data and Safety Monitoring in the Concorde Alpha Trials. *Controlled Clinical Trials*, 20, 207–228.
- Besselink, M. G. H. et al. (2008). Probiotic prophylaxis in predicted severe acute pancreatitis: a randomised, double-blind, placebo-controlled trial. *Lancet*, 371, 651–659.
- Eijk, M. J. van, Roes, K. C. B., et al. (2010). Effect of rivastigmine as an adjunct to usual care with haloperidol on duration of delirium and mortality in critically ill patients: a multicentre, double-blind, placebo-controlled randomised trial. *Lancet*, 373, Published online November 5, 2010, doi: 10.1016/S0140-6736(10)61855.
- Ellenberg, S. S., Fleming, T. R. and DeMets, D. L. (2002). *Data Monitoring Committees in Clinical Trials: A Practical Perspective*. Chichester: John Wiley & Sons.
- Gill, R. D. (2009). Statistics, ethics and probiotica. *Statistics in medicine*, 63, 1-12.
- Graaf, R. van der. (2010). *On equipoise and human dignity*. PhD Thesis, University of Utrecht.
- IGZ, CCMO en VWA (2009). Onderzoek naar de PRO-PATRIA studie. (www.igz.nl).
- Senn, S. (2002). Ethical considerations concerning treatment allocation in drug development trials. *Statistical Methods in Medical Research*, 11, 403–411.
- Sydes, M. R., Spiegelhalter, D. J., Altman, D. G., Babiker, A. B., Parmar, M. K. B. & DAMOCLES Group (2004). Systematic qualitative review of the literature on data monitoring committees for randomized controlled trials. *Clinical Trials*: 1, 60-79.
- Tweel, van der, I. (2010). Wanneer is er genoeg 'bewijs'? *STATOR*, 11(1), 17–21.
- Van der Graaf, R., Roes, K. C. B., Van Delden, J. M. (2012). Adaptive Trials in Clinical Research. Scientific and Ethical Issues to Consider. *Journal of the American Medical Association*, 307(22), 2379–2380.

KIT C.B. ROES is hoogleraar Biostatistiek bij het Julius Centrum voor Gezondheidswetenschappen en Huisartsgeneeskunde van het UMC Utrecht. Als hoofd van het Bureau Kwaliteitsborging Onderzoek is hij verantwoordelijk voor de implementatie van het NFU-beleid Kwaliteit van mensgebonden onderzoek in het UMC Utrecht.
E-mail: <k.c.b.roes@umcutrecht.nl>

OOK IK HEB GEZONDIGD...

GERRIT STEMERDINK

Dit nummer gaat over Ethiek. Een belangrijk onderwerp waar we allemaal mee te maken hebben in de dagelijkse praktijk van ons werk. Er zijn vast wel mensen die nooit onethisch hebben gehandeld, maar ik betwijfel of dat er veel zijn. Als je de regels helemaal strikt handhaaft is de scheidslijn soms maar dun. Eigenlijk zou iedere onderzoeker een soort ethische waakhond op de achtergrond beschikbaar moeten hebben om hem/haar bij te staan bij een beslissing over grensgevallen.

Is het echt zó moeilijk? Ja, want soms is het lastig je te realiseren dat je strikt genomen in de fout gaat. Wie heeft er nooit een cijfer van een student een klein beetje in positieve zin aangepast? De reden daarvoor kan velerlei zijn: je kunt de indruk hebben dat de student door een toevalsfluctuatie beneden het gebruikelijke niveau scoorde, of je wist van een persoonlijk probleem dat net speelde, maar je hebt al met al toch maar een uitkomst 'vervalst'!

Soms speelt ook mee dat de meetinstrumenten lastig af te lezen zijn, of dat nu vragenlijsten zijn of meer fysieke metingen. Dan kun je ongemerkt en geheel onopzettelijk een beetje scheef kijken. Dat is een gevaar waar bijvoorbeeld fysici uitgebreid voor worden gewaarschuwd tijdens practica.

In de Ethische Richtlijnen van het International Statistical Institute (zie <www.isi-web.org>) wordt aanbevolen dat onderzoekers een volledige openbaarheid van hun gegevens én hun methoden, inclusief computerprogramma's, dienen na te streven. Recente ontsparingen bij Nederlandse universiteiten hebben aangetoond hoe nuttig zo'n openbaarheid zou kunnen zijn. Helaas zitten we maar al te vaak met tegenstrijdige belangen. Je zult bijvoorbeeld maar werken bij een farmaceutisch bedrijf. De baas ziet je al aankomen als je zegt dat alle gegevens openbaar moeten worden.

Een ethisch ongeluk zit in een klein hoekje. Zelf

kan ik me herinneren minstens eenmaal in de fout te zijn gegaan. Weliswaar een klein beetje, maar toch: fout is fout.

Het speelde rond 1965 toen ik bij een marktonderzoeksbureau werkte. Het bedrijf had het financieel moeilijk en de directeur had een tamelijk kostbaar onderzoek binnengehaald waarbij een aantal kleine, middelgrote en grotere drukkerijen moest worden ondervraagd. Hij was met de opdrachtgever een ingewikkelde prijsberekening overeen gekomen, waarbij het financieel nogal wat uitmaakte of het ons lukte 20 middelgrote bedrijven te strikken voor het onderzoek. Helaas kwamen we maar tot 19. Maar we hadden er bij de kleinere bedrijven eentje die net twee personeelsleden onder de grens van klein/middelgroot zat. In opdracht van de directie is toen de ingevulde vragenlijst vervalst, waardoor dit bedrijf als middelgroot werd gerekend en we meer broodnodige guldens binnenhaalden. Volgens alle regels is dat natuurlijk keiharde fraude, maar wat doe je in zo'n geval als net binnengekomen jongste medewerker? Ik heb het maar gedaan, ook al omdat ik in een voorlopige analyse had gezien dat de totale uitkomst nauwelijks beïnvloed werd door deze ingreep. Als ik toen had kunnen overleggen met een meer ervaren collega zou ik me vast beter hebben gevoeld.

Wel heb ik zo snel mogelijk daarna de in mijn ogen ruige wereld van het marktonderzoek vaarwel gezegd. Later las ik bij toeval (ja, toeval bestaat!) enkele boeken van de Nederlandse schrijver Hans Vervoort, gebaseerd op zijn ervaringen in de marktonderzoekwereld. Ik vond het heel herkenbaar.

Dus, als u mij weer eens ziet en denkt dat ik een keurige oudere heer ben: besef dat schijn bedriegt en dat ook ik heb gezondigd!

GERRIT STEMERDINK is eindredacteur van *STATOR*
E-mail: <gjstemerding@hotmail.com>

ETHISCH OPTIMALISEREN VAN HET RUILEN VAN NIEREN

KRISTIAAN GLORIE, JORIS VAN DE KLUNDERT & ALBERT WAGELMANS

Het tekort aan donoren voor nieren en andere organen vormt al jaren een probleem, zowel binnen als buiten Nederland. Sinds 2004 wint een mogelijkheid voor het verlichten van dit tekort aan populariteit, het zogeheten cross-overprogramma waarin nierpatiënten met een incompatibele levende donor kunnen ruilen van nier(donor). Dankzij dit programma heeft ruim 40% van de patiënten met een levende donor, die anders niet door zouden kunnen gaan met transplantatie, weer perspectief. Het op deze manier ruilen van nieren brengt echter ook een aantal belangrijke ethische en optimalisatievraagstukken met zich mee. Wie mag er bijvoorbeeld meedoen in een ruil en wie niet? In dit artikel laten we zien hoe operations-researchtechnieken hierbij uitkomst bieden.

Verlichten van het orgaantekort

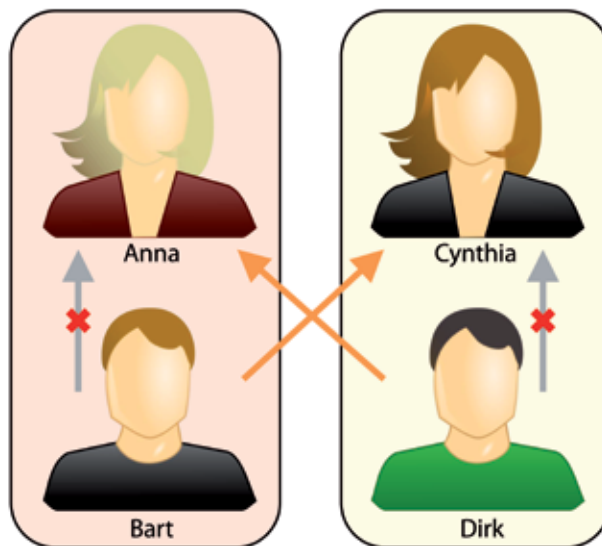
Er zijn in Nederland circa 60.000 patiënten met ernstige nierproblemen (Nierstichting, 2011). 6400 van hen zijn aan het dialyseren, wat betekent dat ze drie tot vijf keer per week naar een dialysecentrum moeten om hun bloed gedurende vier uur te laten zuiveren en onderworpen zijn aan een streng dieet. De kwaliteit van leven met dialyse is erg laag en het jaarlijkse overlijdenspercentage is 20%. Het geprefereerde alternatief, transplantatie, is helaas niet voor iedereen beschikbaar. Per jaar kunnen slechts 860 patiënten op deze manier geholpen worden (Nederlandse Transplantatie Stichting, 2012)

De helft van het aantal transplantaties is postmor-

taal. Dat wil zeggen dat de transplantaties plaatsvinden met een orgaan van een overleden donor, waarvoor patiënten gemiddeld 4 jaar op de wachtlijst staan. De andere helft vindt plaats met een levende donor, zoals een broer of zus van de patiënt. Ruim 40 % van de levende donoren is echter incompatibel met de patiënt waaraan ze in eerste instantie zouden willen doneren. Om transplantatie voor deze groep patiënten toch mogelijk te maken is in 2004 het zogeheten cross-overprogramma opgericht. Dit programma stelt incompatibele paren in staat te ruilen van donor om zodoende op een indirecte manier toch door te gaan met transplantatie.

Het cross-overprogramma werkt als volgt. Stel, patiënt Anna heeft een nier nodig. Haar broer Bart wil haar graag helpen en een van zijn twee gezonde nieren doneren. Na enkele tests in het ziekenhuis blijkt echter dat de bloedgroepen van Bart en Anna niet compatibel zijn. Een transplantatie zou vrijwel onmiddellijk tot afstoting leiden. Elders in het land zitten Cynthia en Dirk met een soortgelijk probleem. Cynthia heeft na haar zwangerschap antistoffen aangemaakt tegen de cellen van Dirk. Hierdoor kunnen ook zij niet doorgaan met transplantatie. In een nationale database constateren transplantatiedeskundigen echter dat Cynthia en Bart wél compatibel zouden zijn, evenals Anna en Dirk. Zij stellen daarom voor dat beide paren, weliswaar anoniem, ruilen van donor via het cross-overprogramma (zie figuur 1). Op deze manier zouden zowel Anna als Cynthia toch een transplantatie kunnen krijgen.

Tien jaar geleden hadden Anna en Cynthia niet geholpen kunnen worden. Gelukkig voor hen en vele andere patiënten was Nederland in 2004 het eerste land ter wereld met een nationaal cross-overprogramma. Maar net als bij de toewijzing van postmortale organen rijst de vraag: wie krijgt een transplantatie en wie niet? In ons voorbeeld ruilden Anna en Bart bijvoorbeeld met Cynthia en Dirk. Maar wellicht hadden



Figuur 1. cross-overtransplantatie

zij ook kunnen ruilen met Eduardo en Floor. Wie gaat er dan voor, Cynthia of Eduardo?

Van ruilen komt huilen

Het ruilen binnen het cross-overprogramma hoeft zich niet te beperken tot twee patiënt-donor paren. Er kunnen ook grotere ruilcycli gevormd worden. Een risico is echter dat een donor zich terugtrekt nadat zijn patiënt een nier heeft ontvangen. Dit zou de andere patiënten in de cyclus die nog niet getransplanteerd zijn, maar waarvan de donor reeds heeft gedoneerd, ernstig benadelen. Dit terugtrekken hoeft niet opzettelijk te gebeuren en kan bijvoorbeeld veroorzaakt worden door ziekte of zwangerschap van de donor.

Om benadeling van patiënten op deze manier te voorkomen zijn er een paar mogelijkheden:

1. alle transplantaties in een cyclus gelijktijdig uitvoeren;
2. in plaats van een ruilcyclus een ruilketen vormen die start met een Samaritaanse donor (een levende donor zonder specifieke patiënt) en eindigt met een donatie aan een patiënt op de postmortale wachtlijst; en
3. een patiënt in plaats van een directe transplantatie voorrang geven op de wachtlijst in ruil voor donatie door zijn levende donor.

Opties 2 en 3 hebben als voordeel dat zij de ruilmogelijkheden vergroten, dit in tegenstelling tot optie 1. Bij optie 1 is er namelijk een beperking op hoeveel paren kunnen deelnemen in een cyclus, ingegeven door het aantal operatiekamers en operatieteams dat simultaan beschikbaar is.

Hoe wordt hier in Nederland mee omgegaan? Wel, simultaneiteit is een standaardvereiste voor ruilcycli (die we om logistieke redenen uitvoeren met maximaal vier paren) en Samaritaanse donorketens worden waar mogelijk regionaal – in tegenstelling tot nationaal - per transplantatiecentrum gecoördineerd. De derde optie, wachtlijstprioriteit, is echter bij wet verboden omdat dit patiënten die reeds op de wachtlijst staan zou kunnen benadelen. Met deze maatregelen wordt gepoogd te zorgen dat er binnen een van de grootste cross-over programma's ter wereld 'van ruilen geen huilen komt'.

Geoptimaliseerd ruilen

Een belangrijk criterium bij het bepalen wie met wie ruilt is het helpen van het maximale aantal patiënten. Dit kan worden bereikt door het ruilvraagstuk te formuleren als een optimaliseringsprobleem. Laten we alle patiënt-donor paren en Samaritaanse donoren in het programma aanduiden met N en de mogelijke ruilcycli en -ketens van lengte maximaal k , respectievelijk l , met $C(k,l)$. Dan kunnen we met een binaire variabele X_c voor elke cyclus of keten $c \in C(k,l)$ aangeven of deze uitgekozen wordt. We lossen daartoe het volgende binair programmeringsprobleem op:

$$\max \sum_{c \in C(k,l)} |c| \cdot X_c \quad (a)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{c \in C(k,l): \epsilon c} X_c \leq 1 \quad \forall n \in N \quad (b) \\ & X_c \in \{0,1\} \quad \forall c \in C(k,l) \end{aligned}$$

De doelstelling (a) is hier het maximaliseren van het aantal transplantaties onder de restrictie (b) dat elk patiënt-donor paar en elke Samaritaanse donor in hoogstens één cyclus of keten deelneemt.

Complexiteit

Een bottleneck bij het oplossen van dit probleem is dat het aantal variabelen zeer groot kan zijn. Het aantal cycli en ketens in de formulering hierboven is namelijk exponentieel. Voor $k=4$ en $l=6$ zijn er in een pool met slechts 80 paren en 20 Samaritaanse donoren al meer dan 300 miljoen verschillende cycli en ketens mogelijk. Interessant is dat de complexiteit van het probleem vooral ingegeven is door de beperking op het aantal paren in een cyclus of keten. Voor $k, l \in [3, \infty]$ is het probleem NP-Compleet. Als $k=l=2$ of $k=l=\infty$ dan is het probleem, zij het in verschillend gedefiniëerde grafen, in polynomiale tijd op te lossen als maximum-matchingprobleem.

Voor de waarden van k en l waar we in de praktijk mee te maken hebben is het door gebruik te maken van een zogeheten *branch-and-price* algoritme gelukkig ook mogelijk snel een optimale oplossing te vinden (Abraham et al., 2007). Het idee achter dit algoritme is dat niet alle mogelijke cycli en ketens van tevoren gegenereerd worden, maar alleen aan het probleem worden toegevoegd wanneer dit op basis van hun *price* berekend met de duale waarden in de LP relaxatie gerechtvaardigd is. Door dit herhaaldelijk te doen in elke node van een *branch-and-bound*-boom kan een optimale geheeltallige oplossing worden verkregen.

Ethisch ruilen

Maximaliseren van het aantal transplantaties is niet het enige criterium. Want is een maximale oplossing ethisch gezien wel juist? Op Europees niveau is afgesproken dat er sprake moet zijn van zowel een optimale als een *rechtvaardige* verdeling (Council of Europe, 2002). Dit houdt in dat factoren als de kans op transplantatie en de wachttijd ook meegenomen moeten worden. Om op ons voorbeeld terug te komen: stel dat Eduardo anti-stoffen heeft tegen een zeer hoog percentage van alle donoren waardoor de kans op het vinden van een geschikte match zeer klein is, dan is het wellicht eerlijker om Bart aan Eduardo te laten doneren dan aan Cynthia.

De Nederlandse transplantatiestichting heeft de volgende zes hiërarchische beslissingsregels opgesteld waaraan een cross-over ruil moet voldoen:

1. het aantal transplantaties is maximaal;
2. het aantal bloedtype identieke transplantaties is maximaal;
3. de patiënt met de laagste matchkans wordt gematcht (iteratief voor elke transplantatie);
4. het aantal paren in de langste ruilcyclus is zo klein mogelijk;
5. de spreiding over transplantatiecentra in de minst gespreide ruilcyclus is zo groot mogelijk;
6. de patiënt met de langste dialysetijd wordt gematcht.

Het doel van criterium 2 is om patiënten met een lastig te matchen bloedtype te beschermen. Dit wordt vervolgens verfijnd door criterium 3 dat rekening houdt met een meer specifiek gedefiniëerde matchkans. Criteria 4 en 5 zijn logistiek van aard en criterium 6 spreekt voor zich. De hiërarchische opzet van de criteria sluit aan bij de traditionele prioriteitsmechanismen die worden gebruikt bij de toewijzing van postmortale organen aan patiënten op de wachtlijst.

Hoewel het toevoegen van deze criteria het optimalisatieprobleem computationeel lastiger maakt, blijft

het gelukkig door het op een slimme manier uitvoeren van *branch-and-price* iteraties, waarbij de doelfunctiewaarden gepropageerd worden door het toevoegen van restricties, ook mogelijk om dit probleem snel tot optimaliteit op te lossen (Glorie et al., 2012b).

Alternatieven

De hierboven genoemde criteria vormen niet de enige mogelijkheid om tot een rechtvaardige ruil te komen. Als alternatief wordt ook wel een lotingsprocedure genoemd. Hoewel een dergelijk stochastisch mechanisme mooie theoretische eigenschappen heeft, zoals het bieden van zoveel als mogelijk gelijke kansen voor patiënten, wordt zij in de praktijk echter (nog) nergens omarmd.

Daarnaast is het bijvoorbeeld mogelijk te matchen op basis van gezondheidswinst. Naast het feit dat dit de voordelen van transplantatie maximaliseert, schept het ook een mogelijkheid voor de inclusie van compatibele paren in het cross-overprogramma. Stel dat een compatibel echtpaar van wat oudere leeftijd, Greet en Henk, de mogelijkheid geboden wordt om te ruilen met de jongere Cynthia en Dirk. Dit zou er toe kunnen leiden dat Greet een betere kwaliteit orgaan krijgt en dat Cynthia niet ongematcht hoeft achter te blijven. Door compatibele paren een verbetering in levensverwachting te garanderen zouden de kansen voor alle patiënt-donor paren in het programma zo verbeterd kunnen worden.

Uitdagingen

Bij het bespreken van degenen die door het ruilen benadeeld kunnen worden hebben we ons direct gericht op de patiënt-donor paren. Maar zij zijn niet de enigen die er op achteruit kunnen gaan. De transplantatiecentra zelf dienen ook in acht genomen te worden, met name wanneer het gaat om het ruilen met Samaritaanse donoren. Een centrum steekt namelijk

veel tijd en geld in de voorbereidende onderzoeken en opwerking van donoren. Met een Samaritaanse donor zou het centrum dan ook het liefst zoveel mogelijk van de eigen patiënten helpen. Dit vormt in veel landen een groot praktisch probleem bij het opzetten van een nationaal programma en is mogelijk een van de redenen waarom ruilketens in Nederland tot op heden lokaal worden gecoördineerd.

In een recente studie is gekeken naar de toevoeging van participatierestricties (garanties dat een transplantatiecentrum in een nationaal mechanisme minstens zoveel transplantaties kan verrichten als in een lokaal mechanisme) voor verschillende configuraties van het cross-overprogramma (Glorie et al., 2012a). Hieruit blijkt dat dergelijke restricties geen negatieve consequenties hebben op de lange termijn. Bovendien wordt aangetoond dat onder landelijke coördinatie een toename in het aantal transplantaties mogelijk is van bijna 60% voor de meest lastig te matchen patiënten.

Andere uitbreidingen

Hoewel we het cross-overprogramma tot dusver in een statische toestand hebben beschouwd, is het dat in werkelijkheid niet. Maandelijks melden zich nieuwe patiënten en donoren aan en verdwijnen reeds aangemelde paren (bijvoorbeeld door verergering van de ziekte, transplantatie buiten het programma om etc.). Deze dynamische context heeft implicaties voor de optimaliteits- en rechtvaardigheidscondities van optimale oplossingen. In toekomstig onderzoek zullen we ons daarom ook richten op het oplossen van dit dynamische probleem.

Tot slot

Het aloude gezegde 'waar twee ruilen, moet er een huilen' hoeft niet op te gaan voor het ruilen van nieren binnen het cross-overprogramma. Integendeel, het programma biedt deelnemers de kans om te overleven

in betere gezondheid. Dat brengt belangrijke optimaliseringsvraagstukken met zich mee. Ten eerste om te zorgen dat het maximaal aantal patiënten geholpen wordt, ten tweede om ervoor te zorgen dat dit rechtvaardig, of zelfs zo rechtvaardig mogelijk gebeurt. Dankzij recente ontwikkelingen op het gebied van operations research blijven deze optimaliseringsvraagstukken praktisch oplosbaar, nu en in de toekomst als het cross-overprogramma nog verder groeit. Zo draagt operations research bij aan de kwaliteit en het behoud van leven van een groeiende populatie van patiënten.

LITERATUUR

- Abraham, D., Blum, A. & Sandholm, T. (2007). Clearing algorithms for barter exchange markets: enabling nationwide kidney exchanges. In *Proceedings of the ACM Conference on Electronic Commerce (EC)*.
- Glorie, K. M., De Klerk, M. and Wagelmans, A. P. M., Van de Klundert, J. J., Zuidema, W. C., Claas, F. H. J. & Weimar, W. (2012a). *Unspecified donation in kidney exchange: when to end the chain?* (Econometric Institute report, 2012-19). Rotterdam: Erasmus Universiteit Rotterdam.
- Glorie, K., Wagelmans, A. P. M. & Van de Klundert, J. J. (2012b). *Iterative branch-and-price for large multi-criteria kidney exchange* (Econometric Institute report, 2012-11). Rotterdam: Erasmus Universiteit Rotterdam.
- Nierstichting Nederland (2011). *Feiten en cijfers*. Bussum: Auteur.
- Nederlandse Transplantatie Stichting (2012). *Annual report 2011*. Leiden: Auteur.
- Council of Europe (2002). Additional protocol to the convention on human rights and biomedicine concerning transplantation of organs and tissues of human origin (*European Treaty Series*, 186). Straatsburg: Auteur.

KRISTIAAN GLORIE is promovendus aan het Econometrisch Instituut van de Erasmus Universiteit Rotterdam, op een project over optimale orgaanallocatie.
E-mail: <glorie@ese.eur.nl>

JORIS VAN DE KLUNDERT is hoogleraar Bedrijfsvoering van Zorgorganisaties, bij het instituut Beleid en Management Gezondheidszorg van de Erasmus Universiteit Rotterdam. Hij is sectieleider van de sectie Health Service.
E-mail: vandeklundert@bmg.eur.nl

ALBERT WAGELMANS is hoogleraar Management Science en directeur van het Econometrisch Instituut van de Erasmus Universiteit Rotterdam.
E-mail: <wagelmans@ese.eur.nl>

WAT IS KREDIET IN HET CASINO WAARD?

EEN ETHISCHE VRAAG

HENK TIJMS

In de kansrekening wordt een prominente rol gespeeld door de zogenoemde *gambler's ruin*-formule, een formule die al teruggaat tot Christiaan Huygens. Deze formule kent vele toepassingen en een verrassende toepassing heeft betrekking op de zaak 'Zarin'. De verslaafde gokker David Zarin kreeg in 1980 een vrijwel onbeperkt krediet in een casino in Atlantic City om te gokken. Het casino zette het krediet pas stop toen de gokschuld van Zarin was opgelopen tot 3 miljoen dollar. Mede vanwege de wetgeving in New Jersey om verslaafde gokkers te beschermen, kon het casino niet via de rechter de schuld verhalen bij Zarin en was het casino gedwongen het overgrote deel van de schuld kwijt te schelden. Echter hiermee was de kous niet af voor Zarin. Kort daarna viel op zijn deurmat een aanmaning van de Amerikaanse belastingdienst om inkomstenbelasting te betalen over ongeveer 3 miljoen dollar. Hierop stapte Zarin naar de rechter. Met het argument dat Zarin nimmer cash geld ontvangen had

maar alleen chips om te spelen, wist de advocaat van Zarin te bereiken dat de rechtbank als oordeel uitsprak dat Zarin geen inkomstenbelasting verschuldigd was.

De rechtbank stelde echter niet de simpele vraag: 'wat is in geld uitgedrukt de waarde van een krediet van 3 miljoen dollar aan chips om te mogen spelen in het casino?' Hoewel de kans niet groot is, de gokker had ook winst kunnen maken en dus, na terugbetaling van het voorschot van 3 miljoen dollar, het casino kunnen verlaten met door een lening verdiend geldsbedrag. De *gambler's ruin*-formule stelt ons in staat om dit geldsbedrag nader te kwantificeren. Zarin had in het casino het in Amerika uiterst populaire gokspel *craps* gespeeld. Craps is een dobbelspel dat gespeeld wordt met twee dobbelstenen. In craps zijn verschillende weddenschappen mogelijk, maar verreweg de meest gespeelde weddenschap is de pass-line weddenschap. Op details van craps en de pass-line weddenschap gaan we niet in, maar het volstaat op te merken dat bij

de pass-line weddenschap de winkans van de speler gelijk is aan $244/495 \approx 0,493$ en de winkans van het casino gelijk is aan $251/495 \approx 0,507$. Bij winst van de speler krijgt deze twee maal de inzet terug en anders verliest de speler de inzet. Dit is precies de situatie van het *gambler's ruin*-probleem. In dit probleem start de gokker met a geldseenheden, zet elke keer één geldseenheid in en ziet dan met kans p zijn kapitaal met één geldseenheid toenemen en met kans $q = 1 - p$ zijn kapitaal met één geldseenheid afnemen. De gokker stopt als zijn kapitaal een van te voren vastgelegd bedrag van $a + b$ geldseenheden heeft bereikt of als de gokker blut is. Geven we met $P(a,b)$ de kans aan dat de gokker zijn doel van $a + b$ geldseenheden bereikt voordat de gokker blut is, dan wordt de beroemde *gambler's ruin*- formule gegeven door

$$P(a,b) = \frac{1 - (q/p)^a}{1 - (q/p)^{a+b}}$$

waarbij $P(a,b) = a/(a+b)$ als $q = p$. Deze formule stelt ons in staat te berekenen dat de 'waarde' van een krediet van 3 miljoen dollar ongeveer gelijk is aan 195 duizend dollar in het geval van David Zarin. Daarvoor moet nog wel het volgende opgemerkt worden. Als het doel van de speler is om een vooraf gesteld doelbedrag met maximale kans te bereiken en het spel is ongunstig voor de speler (een positief huisvoordeel voor het casino), dan is het voor de speler optimaal om zo brutaal mogelijk te spelen, dat wil zeggen elke keer de inzet gelijk te kiezen aan het maximaal inzetbare bedrag. Intuïtief is dit duidelijk: bij brutaal spel stelt de speler zijn kapitaal zo kort mogelijk bloot aan het huisvoordeel van het casino. In het geval van Zarin had het casino een huislimiet van 15 duizend dollar voor de pass-line weddenschap bij het spel craps. Het is redelijk te veronderstellen dat Zarin elke keer 15 duizend dollar inzette. In de *gambler's ruin*-formule nemen we derhalve 15 duizend dollar als geldseenheid. Laten we verder veronderstellen dat Zarins doel was b additionele geldseenheden te winnen bovenop

het startkapitaal van $3.000.000/15.000 = 200$ geldseenheden, waarbij Zarin de waarde van b van te voren gekozen had.

Wat is een redelijke keus voor b voor de situatie dat gokker het casino niets verschuldigd is als hij blut raakt, terwijl de gokker het casino met b geldseenheden in zijn zak verlaat als zijn kapitaal het streefbedrag van $a + b$ geldseenheden bereikt bij een startkapitaal van a geldseenheden? De waarde van het krediet van a geldseenheden kan gedefinieerd worden als de verwachtingswaarde

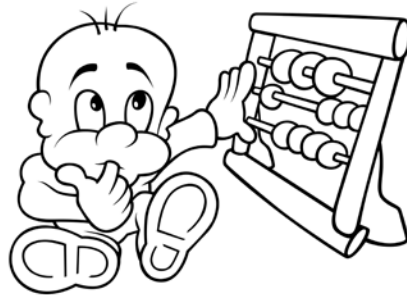
$$u(a, b) = 0 \times [1 - P(a, b)] + b \times P(a, b).$$

Het is redelijk het maximum van $u(a,b)$ als functie van b te beschouwen als de waarde van het krediet van a geldseenheden. Het is een kwestie van simpele algebra om aan te tonen dat voor voldoende grote a de waarde van b waarvoor de functie $u(a,b)$ maximaal is bij goede benadering gegeven wordt door $1/\ln(q/p)$ onafhankelijk van de waarde van a . Verder is voor voldoende grote a de maximale waarde van $u(a,b)$ bij goede benadering gelijk aan $e^{-1/\ln(q/p)}$ waarbij $e=2,71828\dots$ de basis is van de natuurlijke logaritmes.

Passen we bovenstaande resultaten toe op het geval van Zarin met $a=200$, $p = 244/495$ en $q = 251/495$, dan vinden we $b = 35,355$ geldseenheden en een waarde van ongeveer 13 geldseenheden voor het krediet van het casino. Elke geldseenheid vertegenwoordigt 15 duizend dollar. Dus we kunnen concluderen dat het krediet van 3 miljoen dollar dat het casino aan David Zarin gaf ongeveer 195 duizend dollar waard is. Het was ethisch gerechtvaardigd geweest als de Amerikaanse belastingdienst onze vriend voor dit bedrag van 195 duizend dollar had aangeslagen.

HENK TIJMS is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening.
E-mail: <tijms@quicknet.nl>

Young Statisticians



Hi STATOR readers!

The second year of the Young Statisticians has gone by so fast! Very soon we will need to evaluate our old strategy and form a new one for 2013. For that we still have to look for a nice environment where our creativity is set free: perhaps a nice pub? But first, let Maarten van Smeden and Karin Visser update us on the two upcoming activities. It is very likely that we will combine these activities. But, at the moment of writing this small column, a good location still is being searched for! We'll keep you posted.

Hope to see you at our activities!

Brunilda Balliu

Maarten van Smeden on the Statistical Science Café:

'In the middle of December, the Statistical Science Café will be held on "Statistics and Research Ethics". This event is based on the recent fraud-affairs in science. Do we need strict controls by the statistics police, or, are the recent fraud-affairs just incidental and being hyped in media? On both the Stapel-affair and Smeesters-affair, dr. Marcel van Assen and prof. dr. Richard Gill give their opinion. Prof. dr. Peter Grünwald will lead the discussion. Come along! Keep a close eye on the VvS+OR website for more information or send an e-mail at <sciencecafe@vvs-or.nl>.'

Karin Visser on the Book-market:

'Find the statistics books you have always been looking for... for free! Visit the book market organized by the Young Statisticians of the VvS+OR. Our call for books that have been standing on the shelf collecting dust and needed a new home has been a success. Due to a number of generous offers Young Statisticians managed to collect a few hundred books that we will proudly present in November. Many thanks to dr. A.H. de Jonge, C. Kool, G. Stemerink, prof. dr. F. Steutel, Ron Wolterbeek and dr. Hae-Won Uh! Further details can be found on our website!'



Katerina Papadimitropoulou en Karin Visser bij een auto volgeladen met boeken. Foto: Gerrit Stemerink

IN MEMORIAM

Rien van der Leeden (1956-2012)

Op 24 juli 2012 is op 55-jarige leeftijd overleden Rien van der Leeden, sinds 1982 verbonden aan het Instituut Psychologie van de Universiteit Leiden. Hij was reeds geruime tijd ernstig ziek. Van der Leeden kwam in 1975 in Leiden aan om wiskunde te studeren, maar in het tweede jaar schakelde hij over naar psychologie, welke studie hij in 1986 afsloot, met als specialisatie Methoden en Technieken (M&T) en als bijvak Sociaal-Wetenschappelijke Informatica. Hierop volgde een promotieonderzoek, dat in 1990 succesvol werd afgesloten met het proefschrift getiteld *Reduced Rank Regression with Structured Residuals*. Vanaf 1990 was hij aangesteld als wetenschappelijk staf lid bij de sectie M&T.

Van der Leeden was een begenadigd en veelzijdig docent, die bij studenten zeer geliefd was. Een belangrijke vrucht van zijn onderwijsactiviteiten kwam in 1995 met de verschijning van het boek *Toegepaste data analyse; technieken voor niet-experimenteel onderzoek in de sociale wetenschappen*, geschreven samen met Peter de Heus en Bert Gazendam, en gebaseerd op de cursus 'Survey Onderzoek' die sinds 1986 gegeven werd aan derdejaars psychologiestudenten. Door de glasheldere uitleg en vele praktische tips werd dit boek een bestseller, die herdrukken beleefde in 1999 en 2001. In 2000 werd Van der Leeden senior staf lid van de interuniversitaire onderzoekschool IOPS, waarin promovendi op het gebied van de psychometrie en de sociometrie verenigd zijn. De laatste jaren ontwikkelde hij zich tot de belangrijkste M&T docent in het eerste jaar van het bachelorprogramma, waarvoor hij grootschalige en druk bezochte hoorcolleges gaf. Zijn didactische kwaliteiten werden bekroond toen hij in 2007 de Onderwijsprijs voor Beste Docent van het Instituut Psychologie in ontvangst mocht nemen.

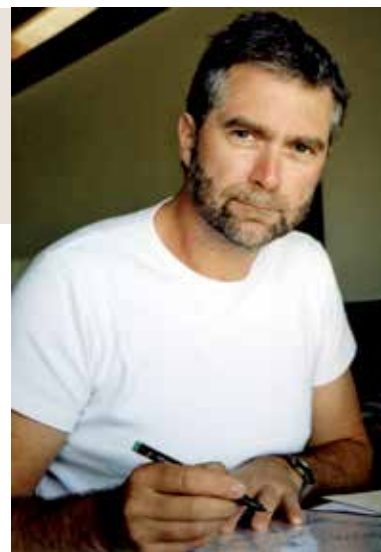
Na het proefschrift richtte Van der Leeden zijn onderzoek zich op multilevel modellen voor analyse van

gegevens die op meerdere aggregatieniveaus verzameld zijn. Hij was betrokken bij de ontwikkeling van multilevel

software, behoorde tot de pioniers die multilevel analyse in Nederland introduceerden, en publiceerde internationaal over diverse technische aspecten. Vanaf 2000 begon hij zich meer te richten op toepassingen van zijn technische kennis. Deze publicaties betroffen een breed spectrum van onderwerpen in de criminologie, de arbeid- en organisatiepsychologie, de onderwijskunde en de ontwikkelingspsychologie, maar hij was vooral een geliefd coauteur bij studies in de gezondheidspsychologie en de klinische psychologie. Zo leverde hij een belangrijke bijdrage aan de acceptatie van multivariate analyse en multilevel modellen binnen de gedrags- en maatschappijwetenschappen.

Naast docent en onderzoeker was Rien van der Leeden ook begaafd fotograaf, en zijn foto's hebben veel uitgaven van de faculteit en het universitaire weekblad *Mare* opgeluisterd. Hij was iemand die in alles ambachtelijke kwaliteit nastreefde en die zijn hang naar het optimale niet gebruikte om anderen te overtroeven, maar juist om de weg er naartoe met anderen te delen. Hij kon terugkijken op een waardevol leven, en heeft in rust afscheid kunnen nemen van zijn dierbaren. Wij wensen zijn vrouw Marike Polak, zijn kinderen Michiel en Ries, zijn ouders en verdere naasten veel sterkte met dit grote verlies.

WILLEM HEISER



Wijbrandt van Schuur (1946-2012)

Een hartaanval tijdens een autorit in Berkeley op 25 juli jongleden heeft een eind gemaakt aan het leven van Wijbrandt van Schuur, een jaar na zijn pensionering. Wijbrandt was daar om de as van zijn Amerikaanse vrouw Melissa, die nog geen jaar eerder overleden was, te verstrooien. Dit was een paar dagen eerder gebeurd.

Wijbrandt was een zeer gedreven persoon met veel gevoel voor rechtvaardigheid. Hij begon als laborant, studeerde politicologie in Amsterdam en werd na zijn afstuderen medewerker aan de Rijksuniversiteit Groningen, vanaf 1989 was hij hoofddocent bij de afdeling sociologie. Zijn onderzoek draaide om twee belangrijke thema's: de methodologie en de politiek.

Bij het eerste ging het vooral om schaalanalyse. Voor het elkaar kunnen begrijpen is meten heel belangrijk, daarbij gaat het vooral om classificaties en ordeningen. Deze twee moeten aan een aantal eisen voldoen, die in meetmodellen omschreven zijn.

Wijbrandts laatste grote product in deze was zijn boek *Ordinal item response theory: Mokken scale analysis* uit 2011. Zijn proefschrift ging over stochastische ontvouwing en zijn grootste puzzel de laatste jaren was het circumplex model. In een artikel met Sipke Huismans uit 2010 wordt beschreven hoe belangrijk dit model is. Tijdens het laatste gesprek dat ik met Wijbrandt heb gehad, vlak voordat hij naar Amerika vertrok, vertelde hij trots de oplossing te hebben voor de mathematische problemen die hij had met dit model. Na terugkomst zou hij het opschrijven. Dat heeft niet meer kunnen gebeuren, heel jammer voor de wetenschap.

Het tweede thema was de politiek. Wijbrandt was geïnteresseerd in hetgeen we weten van de personen en organisaties die beslissingen nemen over de toe-

komst. Dit spitste zich toe op vragen als: wat weten de burgers over belangrijke politieke gebeurtenissen en figuren. Maar ook op de vraag: wie staan waarvoor, dit vooral gemeten aan de hand van de vraag naar wie de leden zijn van bepaalde politieke partijen. Ook wilde hij graag de programma's van partijen met name op lokaal niveau onderzoeken, data heeft hij volop verzameld, maar verder is het niet gekomen.

Soms was het lastig om met Wijbrandt samen te werken. Je kreeg ongevraagd commentaar op teksten. Dat ging voor een deel over de tekst, maar voor een groter deel over allerlei zaken die gerelateerd waren aan de onderzoeksvraag in die tekst. Vaak was het stof voor een aantal artikelen. Dit duidt op zijn enorme betrokkenheid, maar ook op een tomeloze energie.

Naast zijn werk waren er twee andere zaken erg belangrijk. De eerste lag wel dicht bij het werk: de politiek. Wijbrandt heeft ontzettend veel werk verzet voor de partij D66. In Haren heeft hij de afdeling daar, die bijna ingeslapen was, weer overeind getrokken. In Nijmegen was hij actief in allerlei groepen binnen de partij en Wijbrandt kennende was dat echt actief.

Het tweede naast het werk was het gezin. Hij heeft zware tijden meegemaakt die zeker hun sporen hebben nagelaten. Zijn eerste vrouw, Marian, is overleden na een lang ziekbed, zijn tweede vrouw na een heel kort ziekbed. Maar er waren ook de goede momenten: Hij was ontzettend trots op de kinderen en kleinkinderen. Ook over het verblijf in het 'boshuis' in Berkeley kon hij heel lyrisch zijn. Wij verliezen met Wijbrandt een fantastische collega.

ROEL POPPING

AGENDA

12-14 november 2012

De **41e Stochastics Meeting** zal ook dit jaar in De Wereld in Lunteren worden gehouden. Voor informatie en aanmelding <www.cwi.nl/node/2627>.

18-19 december 2012

De **IOPS Winter Conference** wordt georganiseerd door Interuniversity Graduate School of Psychometrics and Sociometrics en vindt plaats in Enschede. Zie: <www.iops.nl/category/conferences>.

5-7 maart 2013

NTTS (New Techniques and Technologies for Statistics) is een tweejaarlijkse conferentie over de invloed van nieuwe technologieën op statistische methoden met als doel het verbeteren van de kwaliteit van officiële statistieken van het European Framework Programme for Research and Development. **NTTS 2013** vindt plaats in Brussel. Zie <www.ntts2013.eu>.

25 maart 2013

De **5th Mplus user's meeting** zal in Utrecht plaatsvinden. Zie voor informatie <www.fss.uu.nl/mplus/>.

27-28 maart 2013

De **9th Multivel Conference** vindt plaats in Utrecht. Daags daarvoor (26 maart) geeft prof. dr. Stef van Buuren een eendaagse cursus 'Multiple Imputation of Multilevel missing data in MICE'. Voor informatie: <<http://multilevel.fss.uu.nl/>>.

10-14 juni 2013

De **9th Conference on Bayesian Nonparametrics (BNP)** vindt plaats in Amsterdam. BNP, een tweejaarlijkse internationale bijeenkomst, wordt georganiseerd door de International Society for Bayesian Analysis (ISBA) en Mathematical Statistics (IMS). Papers zijn welkom tot 15 oktober 2012. Zie <www.bnpg.win.tue.nl/>.

STATOR

Jaargang 13, nummer 3-4, november 2012

STATOR is een uitgave van de Vereniging voor Statistiek en Operationele Research (VvS+OR). 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 Leeuwen, Mirjam Moerbeek, Gerrit Stermerdink (eindredacteur), Hilde Tobi. Vaste medewerkers: Fred Steutel, Henk Tijms

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+OR

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+OR

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+OR-website

www.vvs-or.nl

Sociale media

Wilt u uw vakgenoten ontmoeten en wilt u discussiëren over actuele thema's, volg dan de VvS+OR en de Young Statisticians via LinkedIn, Facebook, Twitter en Flickr. Sluit je aan bij de Linkedingroep van VvS+OR of Young Statisticians; bekijk foto's op <www.flickr.com/photos/vvs-or/sets>; Like onze Facebook-pagina; volg de President van VvS+OR op <<https://twitter.com/#!/dutchstat>>.

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, Nijmegen

Druk

Drukkerij Zoetewij, Yerseke

Uitgever

© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

OPROEP

OM KANDIDATEN TE NOMINEREN VOOR DE VvS+OR THESIS AWARD 2012

Ter bekroning van een uitzonderlijke afstudeerprestatie aan een Nederlandse instelling voor wetenschappelijk onderwijs of 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 duizend (1000) euro.

Genomineerd kunnen worden studenten die tussen september 2010 en september 2012 zijn afgestudeerd en die nog niet eerder zijn genomineerd. Er 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 2012.

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 scriptie 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.

Reglementen en het nominatieformulier zijn te downloaden op de website van de VvS+OR (<www.vvs-or.nl>).

Het nominatieformulier en de aanbevelingsbrief dienen tezamen met de afstudeerscriptie (in pdf-formaat) te worden opgestuurd naar de VVS+OR op <thesisaward@vvs-or.nl>. Het is ook mogelijk de scriptie per post in plaats van e-mail te versturen; neem in dat geval contact op met Mark van der Loo (via thesisaward@vvs-or.nl) voor informatie over de te volgen procedure.

De nominatie dient binnen te zijn vóór 1 december 2012.

Namens de VvS+OR,

Prof. dr. Peter Grünwald, voorzitter Jury Thesis Award 2012

Dr. Mark van der Loo, secretaris Jury Thesis Award 2012

