

STAtOR

periodiek van de VVSOR jaargang 23, nummer 2, juni 2022

100.000 Extra hulpverleners; Het Woeste Doel van Brandweer Amsterdam-Amstelland

Nieuwe vrienden

Vijfenveertig jaar statistiek

Beter inzicht in energie-impact

Stiglers wet, het Mattheüseffect of selectiebias?

Rommel niet met de assen van een grafiek

Lof der luiheid

Social media, a broken legal system, and Micky Mouse statistics; An Italian CSI drama



STATOR

Jaargang 23, nummer 2, juni 2022

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

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Joep Burger, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stemerink (eindredacteur), Vanessa Torres van Grinsven en Sanne Willems. Vaste medewerkers: Jelke Bethlehem, John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Universiteit van Amsterdam Faculteit Economie en Bedrijfskunde, Sectie Operations Management | Amsterdam Business School, Plantage Muidergracht 12, 1018 TV Amsterdam, j.a.s.gromicho@uva.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: Pieter Jongsma MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Algemeen bestuurslid: Thomas Wise MSc, db@vvsor.nl; Webmaster: Eugenio Traini MSc: webmaster@vvsor.nl.

Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Daniel Oberski (Section Data Science); Marije Sluiskes MSc (Young Statisticians) dr. Sanne Willems (Section Statistics Communication).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Maarsbergseweg 20, 3956 KW Leersum, admin@vvsor.nl. Raadpleeg onze website www.vvsor.nl over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STATOR.

Voor advertenties

M. van Hootegem, hootegem@xs4all.nl
STATOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

Verstrooiend en leerzaam

Terugblikkend op ruim 22 jaar STATOR kunnen we constateren dat vrijwel alle nummers zowel verstrooiende als leerzame artikelen bevatten. Vaak zijn beide aspecten gecombineerd in een en hetzelfde artikel.

Ook dit nummer lijkt te passen in die traditie. We openen met een verhaal van Guido Legemaate et al. over de inzet van grote aantallen extra hulpverleners bij de brandweer en de positieve impact die dat heeft op de aanrijtijden. Zeer verstrooiend, maar ook leerzaam.

De recente toekenning van een belangrijke Award aan Judea Pearl was voor Richard Starmans aanleiding een verstrooiend artikel te schrijven over wetenschappelijke prijzen.

Leerzaam is het artikel van Richard Gill waarin hij vertelt hoe met behulp van statistiek een Italiaanse pendant van Lucia de Berks foutieve veroordeling werd voorkomen.

Ronald Does blikst terug op 45 jaar in de statistiek, zijn loopbaan omvat zowel mathematische statistiek als medische statistiek en toepassingen van statistiek in de industrie. In dit nummer vertelt hij over de jaren bij het CWI en de Universiteit Maastricht. In het volgende nummer zal hij zijn werk in de industriële statistiek belichten. Behoorlijk verstrooiend om te lezen.

Drie van de vier columns in dit nummer hebben een hoge mate van leerzaamheid: Jelke Bethlehem legt uit wat er vaak fout is aan de verticale as van grafieken, John Poppelaars vertelt hoe wiskunde kan bijdragen aan een beter inzicht in de energie-impact en Henk Tijms gebruikt een denkbeeldig treffen tussen twee tegenstanders voor een gedegen lesje kansrekening. Als verstrooiend tegenwicht stelt Gerrit Stemerink in de vierde column dat een wiskundige manier van denken een zekere mate van luiheid impliceert.

Aanhakend bij zijn leerzame column over foute grafieken volgt een verstrooiend gesprek van Jelke Bethlehem met Gerrit Stemerink over Jelke's recente *Grafiekenboek*. Hierin worden systematisch alle vormen van grafieken behandeld, met steeds een voorbeeld van een typerende fout.

Verder nieuws over de Young Statisticians, het erelidmaatschap van de VVSOR voor Jacqueline Meulman en de Willem R. van Zwet Prijs voor Rianne de Heide. En helaas ook droevig nieuws over het veel te vroege overlijden van Gerhard Woeginger.

U ziet het: ook dit nummer is verstrooiend én leerzaam. Wij wensen u veel leesplezier!



INHOUD

- 2 Verstrooiend en leerzaam
- 4 100.000 Extra hulpverleners; Het Woeste Doel van Brandweer Amsterdam-Amstelland | GUIDO LEGEMAATE, DAAN SCHÖNBERGER, CHRISTIAAN SMIT & ROB VAN DER MEI
- 9 Nieuwe vrienden – column | HENK TIJMS
- 11 Vijfenvestig jaar statistiek; Deel 1: periode 1976–1989 | RONALD DOES
- 14 Beter inzicht in energie-impact – column | JOHN POPPELAARS
- 17 Stiglers wet, het Mattheüseffect of selectiebias? | RICHARD STARMANS
- 20 Young Statisticians
- 21 Peilingpraktijken. Rommel niet met de assen van een grafiek – column | JELKE BETHLEHEM
- 24 Grafieken kunnen een zegen zijn, maar ook een ramp; Interview met Jelke Bethlehem | GERRIT STEMERINK
- 26 Lof der luiheid – column | GERRIT STEMERINK
- 27 Social media, a broken legal system, and Micky Mouse statistics; An Italian CSI drama | RICHARD GILL
- 35 In Memoriam Gerhard Woeginger, 1964 – 2022 | FRITS SPIEKSA
- 36 Erelidmaatschap voor prof. dr. Jacqueline Meulman
- 36 VVSOR Willem R. van Zwet Prijs voor Rianne de Heide



100.000 extra hulpverleners

HET WOESTE DOEL VAN BRANDWEER AMSTERDAM-AMSTELLAND

‘Iedereen kent de brandweer. Hoogste tijd dat de brandweer ook iedereen leert kennen!’ Okay, dat is misschien wat overdreven, maar wellicht dat een gedach- tenspinsel als dit heeft bijgedragen aan het bedenken van het volgende Woeste Doel: de hoofdstedelijke brandweer hoopt in 2030 zo’n 100.000 extra hulpver- leners te hebben om in te zetten gedurende de eerste cruciale minuten van een incident. Terwijl de brandweer zich nog beraadt over wat ze dit nieuw type hulp- verleners laat uitvoeren, rekenen wij alvast aan vragen als ‘heb je er eigenlijk wel echt 100.000 nodig?’, ‘wat levert het op in termen van tijdswinst voordat hulp ter plaatse is?’ en ‘waar moet je die dan positioneren/werven?’

GUIDO LEGEMAATE, DAAN SCHÖNBERGER, CHRISTIAAN SMIT & ROB VAN DER MEI

Brandweer Amsterdam-Amstelland, opgericht in 1874, is het oudste beroepsbrandweerkorps van Nederland. In de begintijd opereerde het vanuit 7 posten. Tegenwoordig verzorgt het, als onderdeel van veiligheidsregio Amster- dam-Amstelland, de brandweerzorg en crisisbestrijding voor de 7 gemeenten in deze regio. Vanaf momenteel 19 posten rukken zij, met een verscheidenheid aan personeel en materieel, uit naar een breed scala aan type incidenten. Het daadwerkelijk blussen van branden is er slechts één van, en niet eens het meest frequent voorkomende. In het noordelijk deel van de regio, voornamelijk stedelijk gebied, gebeurt dit met een flink aantal beroepsposten. Typerend voor deze posten is dat zij over het algemeen 24 uur per

dag bemenst zijn. Personeel werkt, oefent, studeert, slaapt en rukt uit vanaf deze post. In het zuidelijk deel van de regio zijn voornamelijk vrijwilligersposten te vinden. Deze parttime brandweermensen bewegen zich vrij door het verzorgingsgebied en reageren op alarmeringen vanaf de locatie waar zij zich op dat moment bevinden. Zij gaan niet direct naar het incident, maar maken een tussenstop bij hun post. Daar kleden ze zich om en, wanneer voldoende personeel is gearriveerd, rukken zij uit. Deze *pre-trip delay* bedraagt in de regel ongeveer 3 minuten. Voor beide orga- nisatievormen geldt dat zij zich aan dezelfde opleidingssei- sen en wettelijke normen dienen te houden. En ja, deze brandweervrijwilligers krijgen dus ook gewoon betaald.

Normen voor opkomsttijd

Tot op heden één van de belangrijkste normen waar de brandweer zich aan dient te houden is de opkomsttijd bij het incidenttype brand. Dit is de tijd tussen de brandmelding en het moment dat de brandweer ter plaatse is. Het behelst daarmee de tijd die het kost voor het aannemen van de melding op de meldkamer, de tijd die het kost voor de brandweer om uit te rukken (dus in het geval van vrijwilligers óók de rijtijd naar de post toe) én de rijtijd naar het incident. De wettelijke opkomsttijd bedraagt afhankelijk van bouw en gebruiksdoel van het object 5, 6, 8 of 10 minuten. Gemiddeld genomen is de opkomsttijd van de brandweer in Nederland zo'n 8 minuten (waarbij het noemenswaardig is dat posten met vrijwilligers zijn oververtegenwoordigd). Voor andere incidenttypen waar de brandweer met spoed op reageert geldt dat de opkomsttijden in verreweg de meeste gevallen vergelijkbaar zijn.

Woeste Doelen

De Amsterdamse brandweerorganisatie heeft in het beleidsplan 2021–2024 een aantal woeste doelen opgenomen. 'Woeste doelen zijn ons vergezicht. Ze inspireren en enthousiasmeren de ontwikkeling van onze organisatie.' Eén van die woeste doelen is het uiterlijk 2030 realiseren van een netwerk van 100.000 burgerhulpverleners. Het beleidsplan zegt hierover: '100.000 betrokken inwoners die zich inzetten voor veiligheid in onze regio. Niet omdat het hun werk is, maar omdat ze er willen zijn voor anderen. Bij kleine of grote incidenten en ter voorkoming van incidenten. Als elke seconde telt of met meer rust. Altijd samenwerkend met ons, de veiligheidsprofessionals.'

Burgerhulpverleners

De inzet van burgerhulpverleners is niet nieuw binnen de hulpverleningswereld. Met name binnen de medische sector zijn tal van voorbeelden te noemen van projecten die de verdeling, plaatsing en inzet van automatische externe defibrillatoren (AED) willen verbeteren, al dan niet aan de hand van optimalisatiemethoden (Chan et al., 2016). In Nederland zijn geen voorbeelden bekend van de inzet van burgerhulpverleners binnen de context van brandweezorg waarbij op grote schaal een beroep wordt gedaan op burgerhulpverleners. Vanuit het buitenland is dit soort voorbeelden er wel. Zo had *myResponder* – een *smart city initiative* in Singapore – al succes geboekt met het door middel van een app actief betrekken van getrainde burgers met AED bij reanimaties. Onlangs is dit uitgebreid met het ook betrekken van deze burgers bij branden en ongevallen om in de eerste cruciale minuten levens te kunnen redden (Hasija et al., 2020). Laat het ook voor de Nederlandse context duidelijk zijn: de brandweer kan uiteraard niet zijn wettelijke opkomsttijd verbeteren door de inzet van burgerhulpverleners, maar vanuit het perspectief van de hulpbehoevende burger kan de tijdige komst van welke hulp dan ook wel hét verschil maken.

Verdeling van burgerhulpverleners

Voor een waarheidsgetrouw model moet gekeken worden naar de verdeling van het aantal burgerhulpverleners over de veiligheidsregio. Uiteraard is de positie van een enkele hulpverlener lastig te voorspellen, maar de kans dat een persoon zich in regio bevindt kan wel worden geschat. Een voor de hand liggende manier om dit te doen is door dit gelijk te stellen aan de bevolkingsdichtheid van de regio. Met de aanname dat elke persoon een even grote kans heeft om burgerhulpverlener te zijn, kan geconcludeerd worden dat in elke regio de fractie van hulpverleners op de populatie gemiddeld hetzelfde is. Vervolgens kan de verdeling van de hulpverleners worden gemodelleerd als een *spatial Poisson point process*, waar de intensiteit in proportie is met de populatie van regio.

Een andere manier om hiernaar te kijken is door te optimaliseren over de verdeling. Dit betekent dat de verdeling zo gevormd wordt dat de kans dat een hulpverlener later aanwezig is dan de richttijd van de brandweer wordt geminimaliseerd. De brandweer heeft niet direct invloed op deze verdeling, maar het geeft wel een idee welke gebieden de meeste focus moeten krijgen bij het rekruteren van burgerhulpverleners. Het *greedy* algorit-

me dat hiervoor is gebruikt staat uitgebreid beschreven in het artikel door Van den Berg et al. (2021). *Greedy* leidt in dit geval tot een optimale oplossing, omdat de score-functie een som van convexe formules is. Dit algoritme voegt telkens een hulpverlener toe op de plek waar dit het meeste voordeel zou hebben in het verkleinen van de kans dat men later dan de richttijd bij een incident aankomt. Zowel de proportionele verdeling als de geoptimaliseerde verdeling worden gebruikt bij het berekenen van de resultaten. Op deze manier kan een reële schatting en een bovengrens worden berekend en kan ook de waarde van slim rekruteren worden gekwantificeerd.

Scenario's

In totaal zijn 8 verschillende scenario's opgesteld aan de hand waarvan een conclusie kan worden getrokken over het aantal hulpverleners dat daadwerkelijk nodig is voor een adequate dekkinggraad. De volgende parameters zijn hier van belang:

- n : het totaal aantal burgerhulpverleners in de veiligheidsregio.
- α : de fractie van personen dat op een willekeurig moment beschikbaar is en dus kan reageren op een incident. In de literatuur is dit meestal 0,05 of 0,1.
- v_i : de dichtheid van hulpverleners op locatie. Volgens de proportionele, of de optimale verdeling.

- w : de snelheid van een burgerhulpverlener in km/min.
- τ_i : de verwachte tijd tussen het ontvangen van een melding en het vertrek van een burgerhulpverlener in gebied.
- λ_i : de intensiteit van incidenten op locatie.

De in tabel 1 genoemde scenario's met bijbehorende parameters worden vervolgens als basis gebruikt voor de berekeningen. Belangrijk detail hierbij is dat de winst in opkomsttijd vergeleken met de brandweer wordt gerapporteerd. Als de burgerhulpverlener dus later arriveert dan de brandweer, dan zal dit verschil oedragen. De berekeningen worden gedaan gedaan met behulp van een, door een student in een eerder project uitgebreide en op maat gemaakte simulator én een eigen toepassing van de *Open Source Routing Machine* routeplanner voor het berekenen van de opkomsttijden (Luxen & Vetter, 2011).

Resultaten

Door de genoemde scenario's uit tabel 1 te simuleren krijgt elk scenario 2 resultaten. Eén voor de proportionele verdeling en een voor de optimale verdeling. Elk scenario levert een verbetering ten opzichte van de gemiddelde responstijd van de brandweer. Dat betekent zelfs dat wanneer er 10.000 burgerhulpverleners zouden deelnemen, de responstijd wordt verbeterd. Dit

SCENARIO / PARAMETERS	n	α	w	τ_i
1	100.000	0,1	0,1	3
2	50.000	0,1	0,1	3
3	10.000	0,1	0,1	3
4	100.000	0,05	0,1	3
5	100.000	0,01	0,1	3
6	100.000	0,1	0,2	3
7	100.000	0,1	0,1	4
8	100.000	0,1	0,1	2

Tabel 1. De 8 scenario's met bijbehorende parameters

met 1/10 van het originele Woeste Doel.

De resultaten van scenario 1 laten zien dat er gemiddeld 4 minuten gewonnen kunnen worden ten opzichte van de brandweer. Dit betekent dus een halvering van de responstijd. Figuur 1 laat zien hoe die winst in responstijd is opgebouwd voor de veiligheidsregio Amsterdam-Amstelland. Dit is gedaan met een optimale verdeling van de burgerhulpverleners.

Ook blijkt dat met 'slechts' 50.000 burgerhulpverleners uit scenario 2 er nog steeds een aanzienlijke tijdswinst wordt geboekt. De vertraging ten opzichte van het hiervoor genoemde scenario met 100.000 deelnemers betreft dan 34 seconden.

Een ander belangrijk resultaat was dat wanneer burgerhulpverleners op de fiets/scooter komen uit scenario 6. Dat zorgt ervoor dat variabele groter wordt. Dit geeft bij elk scenario nog eens een winst van ongeveer 50 seconden. Dit is aanzienlijk en aangezien het in Nederland best aannemelijk is dat een burgerhulpverlener met de fiets komt, mag dit resultaat niet verwaarloosd worden.

Ten slotte zorgt het optimalisatie algoritme voor een winst van ongeveer 30 seconden ten opzichte van een proportionele verdeling van de burgerhulpverleners. Dus op de juiste locatie rekruteren zal zeker in het voordeel werken voor de brandweer, of beter gezegd: de hulpbehoevende burger zelf.



Figuur 1. Winst in responstijd bij een optimale verdeling van burgerhulpverleners

LITERATUUR

- Chan, T. C. Y., Demirtas, D., & Kwon, R. H. (2016). Optimizing the Deployment of Public Access Defibrillators. *Management science, INFORMS*, 12, 3617–3635.
- Hasija, S., Shen, Z.-J. M., & Teo, C.-P. (2020). Smart city operations: Modeling challenges and opportunities. *Manufacturing & Service Operations Management, INFORMS*, 22, 203–213
- Luxen, D., & Vetter, C. (2011). Real-time routing with OpenStreetMap data. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM*, 513–516
- van den Berg, P. L., Henderson, S. G., Jagtenberg, C. J., & Li, H. (2021). *Modeling Emergency Medical Service Volunteer Response*. Geraadpleegd op <<https://ssrn.com/abstract=3825060>>.

GUIDO LEGEMAATE is onderzoeker en coördinator van de afdeling informatiemanagement bij brandweer Amsterdam-Amstelland. Hij werkt inmiddels geruime tijd samen met Rob van der Mei op vraagstukken waarbij operationele brandweerprocessen worden verbeterd door het toepassen van inzichten vanuit de wetenschap. Samen met studenten en promovendi zijn de afgelopen jaren verschillende optimalisaties en algoritmen ontwikkeld die in de dagelijkse praktijk worden gebruikt. E-mail: g.legemaate@brandweeraa.nl

DAAN SCHÖNBERGER deed in 2021 zijn afstudeerstage vanuit de masteropleiding Business Analytics (VU) bij brandweer Amsterdam-Amstelland en voerde het in dit artikel beschreven onderzoek uit. Daarnaast heeft hij veel werk verzet in het verbeteren van de gebruikte simulator. Daan heeft zijn studie succesvol afgerond en werkt nu bij een grote bank. E-mail: daanschonberger@hotmail.com

CHRISTIAAN SMIT werkt op dit moment als deel van een afstudeerstage samen met brandweer Amsterdam-Amstelland aan een vervolgstudie van het hier beschreven onderwerp van burgerhulpverlening. Dit vervolgonderzoek zal zich vooral richten op het optimaal verdelen van hulpmiddelen voor burgerhulpverlening over de stad. Hij zit in het laatste jaar van de mastersopleiding Business Analytics (VU) en is vooral geïnteresseerd in optimalisatie-vraagstukken zoals vehicle routing. E-mail: cj.smit@hotmail.nl

ROB VAN DER MEI is als hoogleraar Toegepaste Wiskunde en Manager Research and Strategy verbonden aan het Centrum Wiskunde & Informatica en de Vrije Universiteit Amsterdam. Hij richt zich met name op het modelleren en optimaliseren van logistieke processen in de praktijk, en slaat daarmee een belangrijke brug tussen theorie en toepassing. In 2021 ontving hij voor zijn werk samen met collega Sandjai Bhulai de prestigieuze Huibregtsenprijs. De laatste jaren is hij vooral werkzaam op het gebied van ambulance- en brandweerplanning, het reduceren van wachtlijsten in de ouderenzorg, en het optimaliseren van personenmobiliteit en vrachtlogistiek. E-mail: mei@cw.nl



Nieuwe vrienden

Rein Nobel, mijn gewaardeerde en naaste collega op het gebied van kansrekening en operations research in de tijd dat ik nog aan de universiteit werkzaam was, stuurde mij onlangs een nogal absurdistische opgave over kansrekening. Een opgave die volledig past in het kader van de tentamenopgaven waarmee hij befaamd en berucht was onder de studenten: creatieve opgaven - soms humoristisch maar bij nadere lezing meestal met een serieuze ondertoon - die vaak direct aansloten bij de dagelijkse realiteit en waarin ook nog wel eens instituties en personen op de hak genomen werden. De opgave die hij mij onlangs stuurde luidt als volgt:

De eerste fusiebesprekingen tussen de Noord-Koreaanse en Zuid-Koreaanse voetbalbonden vinden plaats in Jeruzalem. De Israëliëse premier Naftali Bennett, die als groot liefhebber van het Koreaanse voetbal de gastheer is van deze mega-gebeurtenis, besluit de 25 deelnemers van de beide belligerenten voor de lunch plaats te laten nemen aan een grote ronde tafel met 25 genummerde zitplaatsen. In een poging het ijs tussen de onderhandelaars te breken laat Bennett voorafgaande aan de lunch een schaal rondgaan met daarop 25 lootjes en hij verzoekt al zijn gasten een lootje te trekken. Iedere onderhandelaar neemt vervolgens plaats

op de stoel met zijn/haar getrokken nummer. Na de lunch worden de stroef verlopende onderhandelingen in groepjes voortgezet, maar voor het diner worden opnieuw alle onderhandelaars aan de grote ronde tafel genoodd waar dit keer premier Bennett zijn echtgenote Gilat, niet aanwezig bij de lunch en daardoor niet op de hoogte van de eerdere tafelschikking, inschakelt bij de verdeling van de zitplaatsen, in de hoop dat zij met haar grote mensenkennis en vrouwelijke intuïtie een meer de vrede bevorderende tafelschikking uit de hoge hoed zal weten te toveren dan de eerdere door het blinde lot gerealiseerde tafelschikking van de lunch. Tijdens het diner waar de Teperberg en Recanati wijnen rijkelijk geschonken en gedronken worden en de gasten ter afsluiting van het samenzijn kunnen genieten van Gilat's vermaarde desserts (zoals bekend is zij een 'professionele' pastrychef), merkt Naftali Bennett in zijn slottoespraak op dat tijdens het diner iedereen andere gasten naast zich had zitten dan tijdens de lunch. Hij vreest dat daardoor eventuele gedurende de lunch ontstane prille toenaderingen tussen de gasten geen vervolg hebben kunnen krijgen tijdens het diner, maar hij spreekt ook de hoop uit dat de desserts van zijn echtgenote, aangevuld met enkele glazen

Mount Hermon Moscato, de beroemde Israëli-sche dessertwijn van de Golanhoogten, alsnog een pacificerende rol hebben kunnen spelen. Als later op de persconferentie premier Bennett verslag doet van de moeizame onderhandelingen en terloops ook vertelt hoe de twee tafelschikkingen van de lunch en het diner tot stand gekomen zijn vraagt onze sterverslaggever van de Ecotribune - het blad voor econometristudenten aan de Vrije Universiteit - zich af wat de kans is dat tijdens het diner inderdaad niemand naast een lunch-disgenoot gezeten heeft.

Achter deze absurdistische kansopgave gaat een interessant en uitdagend kansprobleem schuil dat in één enkele zin geformuleerd kan worden:

'Stel dat n personen volgens een of ander protocol aan een ronde tafel geplaatst worden en dat vervolgens de tafelschikking veranderd wordt door de personen random een zitplaats toe te wijzen. Wat is de kans dat in de nieuwe tafelschikking geen twee personen naast elkaar zitten die in de oorspronkelijke tafelschikking wel naast elkaar zaten?'

Een combinatorisch kans probleem waarvoor het bijzonder lastig lijkt om tot een analytische oplossing te komen. In een dergelijke situatie brengt Monte Carlo simulatie uitkomst om snel een antwoord te vinden: de gezochte kans is ongeveer 11,4% voor $n = 25$. Voor n voldoende groot kan een simpele benaderingsoplossing gevonden worden met de Poisson heuristiek. Dit gaat als volgt. Denk in dat een Bernoulli-achtig deelexperiment wordt uitgevoerd voor elk van de n personen. Voor een gegeven persoon wordt dit experiment succesvol genoemd als deze persoon in de nieuwe random tafelschikking iemand rechts van hem/haar treft die in de oorspronkelijke tafelschikking rechts of links van hem/haar zat. De gezochte kans is gelijk aan de kans dat geen enkel

deelexperiment succesvol is. De kans dat een gegeven deelexperiment succesvol is wordt gegeven door $2/(n-1)$. De verwachtingswaarde van het aantal succesvolle deelexperimenten is dus $2n/(n-1)$. De deelexperimenten zijn niet onafhankelijk van elkaar, maar voor n voldoende groot zijn de deelexperimenten als 'bijna-onafhankelijk' te beschouwen. Je kunt dan de Poisson heuristiek toepassen. Deze heuristiek geeft dat de kans op geen enkel succesvol deelexperiment bij goede benadering gelijk is aan $e^{-2n/(n-1)}$. Voor $n = 25$ vind je de benaderingswaarde 0,1245. De benaderingswaarde van 12,5% ligt opmerkelijk dicht bij de simulatiewaarde van 11,4%. Voor n voldoende groot geeft de heuristiek een uitstekende benadering (voor $n = 50$ is de benaderingswaarde 13,0% waar simulatie de waarde 12,5% geeft). De benaderingsformule is ook bruikbaar voor de situatie van een lange rechte tafel waar de n personen aan één kant van de tafel geplaatst worden. Voor zowel dit model als het model met een ronde tafel kan exact worden aangetoond dat voor n naar oneindig de formule $2^k e^{-2}/k!$ voor $k = 0, 1, \dots$ geldt voor de limietkans dat in de nieuwe random tafelschikking er precies k personen zijn met rechts van zich iemand die ook directe buur was in de oorspronkelijke tafelschikking. Een andere variant van het tafelschikkingsprobleem is een lange rechte tafel waaraan een even aantal personen aan beide kanten van de tafel geplaatst worden en de kans wordt gezocht dat in de nieuwe tafelschikking geen twee personen naast elkaar of direct tegenover elkaar zitten waar dat in de oorspronkelijke tafelschikking wel het geval was. Werk aan de winkel!

HENK TIJMS is emeritus-hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Zijn meest recente boeken zijn *Basic probability; What every math student should know* (World Scientific Press, 2021, 2e druk) en *Operations Research; An introduction to models and methods* met de co-auteurs R. Boucherie en A. Braaksma, (World Scientific Press, 2021). Homepage: <https://personal.vu.nl/h.c.tijms/> E-mail: h.c.tijms@xs4all.nl

Deel 1: periode 1976–1989

Oosterscheldekering. Foto: Eddy Westveer | www.beeldbank.zeeland.nl



Vijfenveertig jaar statistiek

Als je met emeritaat gaat dan ligt er meer tijd achter je dan voor je. De afgelopen 45 jaar heb ik als statisticus achtereenvolgens gewerkt bij het CWI, de Universiteit van Maastricht, het CQM (Philips) en de Universiteit van Amsterdam. In twee delen zal ik die 45 jaar beschrijven. Het eerste deel zal besteed worden aan mijn carrière als mathematisch en medisch statisticus.

RONALD J. M. M. DOES

Het was een voorrecht om in de vorige eeuw na de lagere school naar de Hogere BurgerSchool (HBS) te gaan. Een brede vijfjarige vervolgopleiding met vijftien eindexamenvakken, die toegang gaf tot een universitaire opleiding. Ook de school van de oud-premiers Drees, Zijlstra, De Jong, Den Uyl en Kok; de school van de ondernemers Fokker, Philips, Heineken en Heijn; de school van dertien

van onze eenentwintig Nobelprijswinnaars; en de school van kunstenaars en schrijvers zoals Vincent van Gogh, Jules Deelder, J. Bernlef, Joris Ivens, Annie M.G. Schmidt, Wim de Bie, Freek de Jonge, Herman Koch en Adriaan van Dis.

Na mijn HBS-b-opleiding begon ik in 1972 aan de studie wiskunde met natuur- en sterrenkunde aan de Uni-

versiteit Leiden. Na twee jaar slaagde ik voor mijn kandidaatsexamen en vervolgde ik met een doctoraalstudie wiskunde met bijvak mathematische statistiek. Het bijvak was een van mijn laatste tentamens en dat ging zo goed dat professor Van Zwet na afloop van het tentamen vroeg of ik interesse had om wetenschappelijk medewerker te worden op het Mathematisch Centrum (nu CWI geheten) in Amsterdam.

Kansmodel voor golf- en vervalcrachten

In september 1976 begon ik mijn carrière als 21-jarige op de afdeling Mathematische Statistiek, die onder leiding stond van professor Hemelrijk. Mijn leven als statisticus was begonnen en mijn eerste taak was om samen met Jelke Bethlehem en Richard Gill onder leiding van professor Oosterhoff een werkweek voor te bereiden met als thema stochastische censurering. Een onderwerp waarin de levensduurverdeling van een object centraal staat. De corresponderende overlevingsfunctie wordt gedefinieerd als $S(x) = P(X > x) = 1 - F(x)$, waarin F de verdelingsfunctie is van de stochastische variabele X . In woorden: $S(x)$ is de kans dat een object een levensduur groter dan x heeft. Een veelvoorkomend probleem is dat men niet van ieder object uit een aselechte steekproef op tijdstip x de levensduur kan meten omdat het object een langere levensduur heeft of bijvoorbeeld uitgevallen is door een andere oorzaak dan die men aan het onderzoeken is. In dergelijke gevallen spreekt men van censurering. In augustus 1977 werd de werkweek gegeven. Centraal stonden de verdelingsvrije methoden voor het twee-steekproevenprobleem met gecensureerde waarnemingen. Tevens werd tijdens de werkweek een overzicht gegeven van de martingaaltheorie en theorie van de stochastische integralen (zie Bethlehem e.a., 1977). Twee jaar later promoveerde Richard Gill op dit onderwerp met een proefschrift getiteld 'Censoring and Stochastic Integrals'.

De afdeling Mathematische Statistiek deed in die tijd ook veel consultatie. Met name Rijkswaterstaat, onderzoeksinstituten en accountancykantoren waren grote opdrachtgevers. De aard van de consultaties was uitdrukkelijk mathematisch statistisch. In mijn tweede jaar kreeg ik samen met Jelke Bethlehem en Roelof Helmers de opdracht om de nota getiteld *Beschrijving van de probabilistische methode voor de bepaling van de golf- en vervalcrachten* van de Deltadienst op zijn wiskundig-statistische merites te voorzien. In Bethlehem e.a. (1978) wordt een kansmodel ontwikkeld voor de golf- en vervalcrachten op de stormvloedkering in de Oosterschelde. Omdat de

golf- en vervalcrachten kunnen worden samengesteld tot een totaalbelasting is een uitdrukking gevonden voor de kans dat de totaalbelasting een gegeven belastingwaarde tenminste eenmaal overschrijdt. Voor het ontwikkelde kansmodel hadden we over de jaren 1949–1977 de beschikking over hoog- en laagwaterstanden te Vlissingen en Burghsluis; significante golfhoogten, tijdens stormen met het oog geschat op het lichtschip of lichteiland Goeree en gemeten bij een meetpaal met een amplitude-schrijver; en windrichtingen gemeten in Hoek van Holland. De bouw van de stormvloedkering in de Oosterschelde begon in 1979 en werd in 1986 voltooid. Sinds de ingebruikname worden de schuiven ongeveer eenmaal per jaar gesloten vanwege hoog water. Het is goed om te constateren dat de stormvloedkering de totaalbelasting iedere keer goed heeft doorstaan.

Naast het uitvoeren van wat kleinere consultatie opdrachten in het tweede jaar kreeg ik ook voldoende ruimte om bij te scholen. Zo werd in het academisch jaar 1978–1979 een werkgroep gestart, bestaande uit statistici van het Mathematisch Centrum en een aantal promovendi in de statistiek uit Nederland en België, om alle opgaven van het boek *Testing Statistical Hypotheses* van E. L. Lehmann op te lossen. Toen we klaar waren met het oplossen van de 192 opgaven, kregen we bericht dat er een tweede editie met extra opgaven zou verschijnen. Die verscheen echter pas in 1986 en toen was de werkgroep al opgeheven.

Hogere orde asymptotische ontwikkelingen voor een klasse rangtoetsen

Een aanstelling bij het Mathematisch Centrum was in principe voor vier jaar met een mogelijkheid tot verlenging van twee jaar. In die periode werd je dan wel geacht te promoveren. In mijn derde jaar werd het dus tijd om een promotieonderwerp te kiezen. Het eerste onderwerp ging over hogere orde asymptotische ontwikkelingen voor functies van uniforme spacings. Uniforme spacings worden gedefinieerd als de verschillen van opeenvolgende order statistics uit een steekproef van n onderling onafhankelijke $(0, 1)$ stochastische variabelen. In wiskundige notatie: noteer met U_1, U_2, \dots, U_n de steekproef uit een $(0, 1)$ -verdeling en $U_{1:n}, U_{2:n}, \dots, U_{n:n}$ de order statistics met $U_{0:n} = 0$ en $U_{n+1:n} = 1$. Uniforme spacings zijn dan gedefinieerd als $D_{i:n} = U_{i:n} - U_{i-1:n}$ voor $i = 1, 2, \dots, n+1$. Samen met Chris Klaassen en Roelof Helmers werden sommen van functies van deze uniforme spacings bestudeerd. Het onderzoek resulteerde in 5 artikelen (zie ondermeer



Op 3 januari 2018 raasde een storm over Zeeland. De kering werd gesloten. Foto: Ben Biondina | www.beeldbank.zeeland.nl

Does & Klaassen, 1984 en Does e.a., 1987). Dit onderwerp werd niet mijn promotieonderwerp. Met Willem van Zwet als promotor werden het toch hogere orde asymptotische ontwikkelingen voor een klasse rangtoetsen. Voor het een-steekproefprobleem was dat al gedaan door Wim Albers en het twee-steekproevenprobleem was door Van Zwet en Peter Bickel opgelost. Bleef over de grotere klasse van lineaire rangtoetsen met het twee-steekproevenprobleem en Spearman's rangcorrelatie coëfficiënt als bijzondere gevallen. Het werd een technisch proefschrift met lange en soms ook ingewikkelde formules om Berry-Esseen-stellingen te bewijzen en Edgeworth-ontwikkelingen af te leiden. De complexiteit werd vergroot door onbegrensde score voortbrengende functies toe te laten, zodat de normalscores, die bijvoorbeeld bij de Van der Waerden-toets worden gebruikt, binnen de resultaten vielen. Het onderzoek leidde tot zes artikelen, waaronder in de *Annals of Probability* en de *Annals of Statistics* (zie Does, 1982 en 1983). In 1982 verdedigde ik mijn proefschrift getiteld *Higher Order Asymptotics for Simple Linear Rank Statistics* aan de Universiteit Leiden.

Medische informatica en statistiek

Nu ik een ruggengraat had als mathematisch statisticus, vond ik het tijd om een carrièreswitch te maken naar de toepassingen. Begin tachtiger jaren lagen de banen echter niet voor het oprapen. Gelukkig was er in het zuiden van het land een nieuwe universiteit opgericht en op 1 mei 1981 ben ik daar begonnen als wetenschappelijk medewerker eerste klasse. De Universiteit Maastricht had een matrixstructuur met capaciteitsgroepen die het onderwijs en onderzoek verzorgden. Ik werd de eerste medewerker van de capaciteitsgroep Medische Informatica en Statistiek. Het onderwijssysteem in Maastricht was probleem-

gestuurd en het betekende dat je als tutor optrad voor kleine groepen studenten (maximaal 12) bij bijvoorbeeld het blok Wetenschapstheorie en Methodologie. Voor het statistiekonderwijs was er een uitzondering gemaakt. Studenten mochten hoorcolleges volgen en vervolgens praktijklessen volgen in kleine groepen. In mijn periode op de Universiteit van Maastricht steeg het aantal studenten Gezondheidswetenschappen heel snel (van 80 in 1981 naar bijna 1000 in 1989). Ook het aantal collega's groeide mee van 1 naar 40, maar de onderwijslast was ongekend hoog.

Mijn eerste onderzoeksonderwerp in Maastricht kwam van een immunoloog die geïnteresseerd was in de bepaling van de relatieve frequentie van een bepaald type cel in een populatie van cellen. Ook nu nog een heel actueel onderwerp bij de bestrijding van het COVID-19-virus, waar de zogenaamde T-cellen van belang zijn voor de immuunrespons. In een laboratoriumopstelling kan worden vastgesteld of het type cel al dan niet aanwezig is. Door slim een verdunningsreeks te ontwerpen waarbij in het begin een hoge score op aanwezigheid wordt gehaald en naar mate de verdunningen toenemen lagere scores op aanwezigheid worden gehaald, kan deze relatieve frequentie van de type cel bepaald worden. Zowel de proefopzet als de statistische analysemethoden werden door Wim Albers en onze eerste promovendus (Leo Strijbosch) bestudeerd. Het leidde tot artikelen in onder meer *Biometrics* en *Statistics in Medicine* (zie bijvoorbeeld Does e.a., 1988), maar ook in de medische tijdschriften zoals *Transplantation* en *Journal of Immunological Methods*. Een tweede (Tjaart Imbos) en derde (Frans Tan) promovendus deden onderzoek naar een stochastisch groeiemodel toegepast bij herhaald toetsen van academische kennis. Bij de Medische Faculteit deden ze dat met de voortgangstoets; een instrument om het cognitieve einddoel van de medische studie te meten. Vier keer per

jaar werd een voortgangstoets bestaande uit 250 juist-onjuist vragen afgenomen bij alle studenten uit het eerste tot en met het zesde jaar. Een student, die het onderwijs volledig volgt, neemt dus tenminste 24 keer deel aan een voortgangstoets, zodat vastgesteld kan worden hoe de groei in kennis van individuele studenten verloopt. Een van de artikelen in deze onderzoekslijn verscheen in *Psychometrika* (zie Albers e.a., 1989).

Naast onderwijs en onderzoek heb ik mij bij de Universiteit van Maastricht ook beziggehouden met bestuurlijke taken zoals voorzitter van de examencommissie, lid van de faculteitsraad, hoofd van de afdeling, en mede-organisator van twee internationale congressen (the Satellite Meeting on Mathematical Statistics and Probability to the ISI Session in 1985 en the Tenth International Meeting on Clinical Biostatistics in 1989).

In november 1989 vond ik het tijd om wederom een carrièreswitch te maken. Ik koos voor het bedrijfsleven en Jos de Kroon van het CQM, een adviesbureau binnen Philips, stelde mij in de gelegenheid om een echt beroep te leren: adviseur. In deel 2 van dit persoonlijke verslag zal de periode 1990-2021 besproken worden.

LITERATUUR

- Albers, W., Does, R.J.M.M., Imbos, Tj. & Janssen, M.P.E. (1989). A stochastic model applied to repeated tests of academic knowledge, *Psychometrika* 54, 451-466.
- Bethlehem, J.G., Does, R.J.M.M. & Gill, R.D. (1977). *Verdelingsvrije methoden bij censurering*, Rapport SN 6/77, Amsterdam: Mathematisch Centrum.
- Bethlehem, J.G., Does, R.J.M.M. & Helmers, R. (1978). *Statistische analyse van verdelingen van golfhoogten en waterstanden bij de Oosterschelde*, Rapport SD 113/78, Amsterdam: Mathematisch Centrum.
- Does, R.J.M.M. (1982). Berry-Esseen theorems for simple linear rank statistics under the null hypothesis, *Annals of Probability* 10, 982-991.
- Does, R.J.M.M. (1983). An Edgeworth expansion for simple linear rank statistics under the null hypothesis, *Annals of Statistics* 11, 607-624.
- Does, R.J.M.M., Helmers, R. & Klaassen, C.A.J. (1987). On the Edgeworth expansion for the sum of a function of uniform spacings, *Journal of Statistical Planning and Inference* 17, 149-157.
- Does, R.J.M.M. & Klaassen, C.A.J. (1984). The Berry-Esseen theorem for functions of uniform spacings, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 65, 461-471.
- Does, R.J.M.M., Strijbosch, L.W.G. & Albers, W. (1988). Using jackknife methods for estimating the parameter in dilution series, *Biometrics* 44, 1093-1102.

Ronald Does is emeritus hoogleraar Industriële Statistiek aan de Universiteit van Amsterdam.
E-mail: r.j.m.m.does@uva.nl



Beter inzicht in energie-impact

De trage totstandkoming van een boycot op olie en gas uit Rusland vanwege de inval in Oekraïne maakt pijnlijk duidelijk hoe afhankelijk Europa is van fossiele brandstoffen. Naast deze directe afhankelijkheid zijn door te toegenomen economische onzekerheid de olie- en gasprijzen naar ongekende hoogte gestegen met een direct effect op de inflatie. In de media wordt druk gediscussieerd over het effect van deze ontwikkelingen op onze economie en de vraag of we een versnelling van de transitie naar duurzamere energieopwekking gaan zien. Het aandeel van energie in de economische prestaties van ons land is volgens het CBS 3,5% van het bruto binnenlands product (BBP)¹, daarmee lijkt de invloed van wijzigingen in de energie beperkt. In hun laatst verschenen rapport² onderzoeken economen van het IPCC onder andere de impact van veranderingen in energiegebruik. Zij komen tot de conclusie dat de impact van maatregelen om de gevolgen van klimaatverandering te beperken, zoals de reductie van energiegebruik, minder dan 0,1% van de jaarlijkse groei van het BBP zal zijn. Daarmee is een incentive om nu te verduurzamen nauwelijks aanwezig. Gezien de nu al merkbare invloed van stijgende energieprijzen heb ik zo mijn twijfels of de inschatting van die beperkte invloed wel betrouwbaar is.

Energie als productiefactor

Om het effect van veranderingen in de beschikbare energie op de economie te bepalen wordt vaak gebruik gemaakt van een productiefunctie, bijvoorbeeld de Cobb-Douglas productiefunctie, bekend uit de economieboeken.

$$Y = A \cdot L^{1-\alpha} \cdot K^\alpha$$

Deze functie wordt door neoklassieke economen veel gebruikt om de relatie weer te geven tussen input, in dit geval fysiek kapitaal (K) en arbeid (L), en output (Y of wel BBP). De A-factor vertegenwoordigt technologie in het model en kan worden gebruikt om verbeteringen in de efficiency van kapitaal en arbeid te modelleren. De parameter α meet de verandering in output bij een verandering van kapitaal en is dus als output elasticiteit te interpreteren. Met $\alpha = 0,30$ zal de output met 0,3% toenemen als kapitaal met 1% toeneemt.

Wat direct opvalt, en dat geldt voor veel van de neoklassieke IPCC modellen, is dat energie als productiefactor ontbreekt. Een manier om energie alsnog op te nemen is door het analoog aan kapitaal en arbeid aan het model toe te voegen³

$$Y = A \cdot L^{1-\alpha-\mu} \cdot K^\alpha \cdot E^\mu$$

De parameter μ van de energiefactor is voor Nederland gelijk aan het aandeel van energie in het BBP. Een af-

name van energie met 50% leidt dan volgens dit model tot een afname van het BBP met 2,05%. Niet direct iets om je zorgen om te maken? In het bovenstaande model wordt de aanname gedaan dat veranderingen in energie beschikbaarheid geen invloed hebben op de prestaties van kapitaal en arbeid en dat is een te simpele weergave van de werkelijkheid. Machines hebben nu eenmaal energie nodig om te kunnen produceren. Energie is dus geen productiemiddel maar een voorwaarde voor kapitaal en mensen om te kunnen produceren.

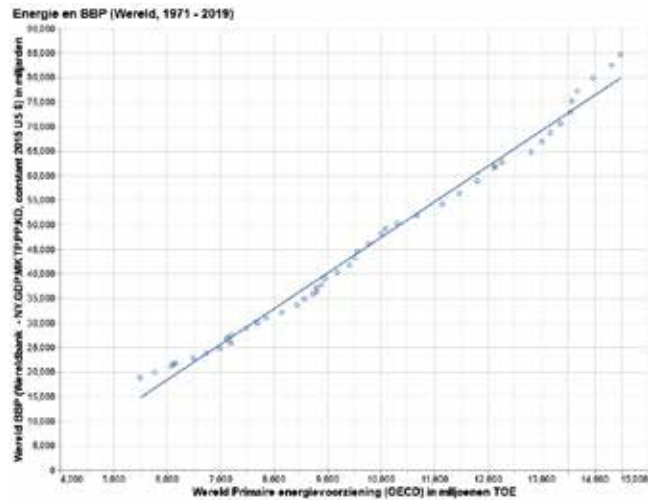
Energie is niet marginaal

Om energie beter in de productiefunctie te verwerken kunnen arbeid en kapitaal als functie van energie worden uitgedrukt⁴

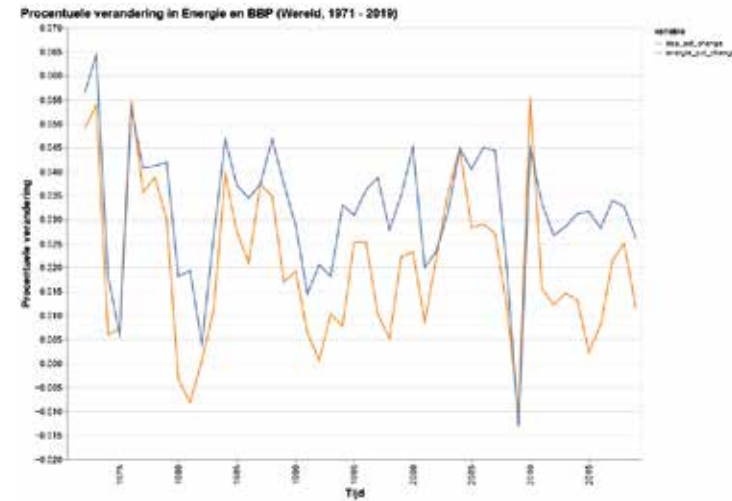
$$K(E) = K \cdot E_K \cdot e_K \text{ en } L(E) = L \cdot E_L \cdot e_L$$

met K en L de hoeveelheid kapitaal en arbeid, E_K en E_L de energie-input voor respectievelijk kapitaal en arbeid en e_K en e_L de energie-efficiëntie van beide productiefactoren. Voor arbeid geldt dat $E_L \cdot e_L = \Lambda$ al eeuwen constant is, ongeveer 100 Watt per uur per persoon. Door substitutie van de functies in de Cobb-Douglas productiefunctie volgt:

$$Y = A \cdot \Lambda^{1-\alpha} \cdot (K \cdot E_K \cdot e_K)^\alpha \cdot L^{1-\alpha}$$



Figuur 1. Energie en BBP⁶



Figuur 2. Verandering in energie en BBP

De verandering van energie input heeft nu direct effect op de bijdrage die kapitaal levert aan het BBP. De mate waarin wordt bepaald door de waarde van α . Empirisch onderzoek van Mankiw⁵ suggereert 0,8 als waarde voor α , waardoor een 50% reductie van energie leidt tot een daling van het BBP met 43%, een bijna 1:1 impact op het BBP. Een relatie die door data uit de praktijk lijkt te worden bevestigd. Als Primaire energievoorziening en BBP voor de wereld met elkaar worden vergeleken voor de periode vanaf 1971 tot 2019 (zie figuur 1) is een sterk lineaire verband zichtbaar. Merk op dat voor afzonderlijke landen deze relatie anders zal zijn vanwege verschillen in efficiëntie waarmee energie wordt opgewekt. Ook als we kijken naar de jaar op jaar veranderingen in energie en het BBP voor de wereld (zie figuur 2), zien we een sterke samenhang (correlatie 0,84)

Uit empirisch onderzoek⁷ naar de relatie tussen BBP en energie volgt dat er over de afgelopen 50 jaar een nagenoeg constante elasticiteit wordt gevonden van 0,8 tussen BBP en Energie. Dat wil zeggen, een 1% groei van het BBP leidt tot een 0,8% groei van energie (of andersom). Uitgaande van deze empirische elasticiteit, zal een reductie van energiegebruik een direct en substantieel effect hebben op de economie.

Betere modellen leiden tot reëler beeld

Hoewel energieprijzen de laatste tijd hoog in de belangstelling staan en steeds duidelijker wordt dat de omslag van fossiel naar duurzamere energieopwekking niet eenvoudig te maken is, is er slechts een beperkte bereidheid om nu maatregelen te nemen om de verduurzaming van de economie te versnellen. Ik wijt dat aan een ver-

keerd beeld dat beleidsmakers hebben van de impact van energie op onze economie. De door IPCC-economen voorgerekende beperkte impact van wijzigingen in de energiehuishouding op de economie leidt tot een laag urgentiegevoel en daardoor blijven concrete plannen uit. Gevolg is dat de klimaatdoelstellingen van 2030 en 2050 steeds lastiger, zo niet onmogelijk, te halen zullen zijn. Data en betere economische modellen, waarin energie expliciet wordt gemodelleerd, leveren een reëler beeld van de consequenties van veranderingen. Dit zal het urgentiegevoel aanwakkeren en naar ik hoop de bereidheid tot handelen vergroten.

JOHN POPPELAARS, Doing the Math
E-mail: john@doingthelmath.nl

NOTEN

1. CBS. *Economische ontwikkeling van de energievoorziening*. (2020). <https://www.cbs.nl/nl-nl/longread/diversen/2020/economische-ontwikkeling-van-de-energievoorziening/2-nederlandse-energievoorziening-economisch-verkend>
2. IPCC (2022), <https://www.ipcc.ch/report/ar6/wg2/>
3. Engström, G., & Gars, J. (2016). Climatic Tipping Points and Optimal Fossil-Fuel Use. *Environmental and Resource Economics*, 65, 541–571.
4. Keen, S., Ayres, R., & Standish, R. (2019). A Note on the Role of Energy in Production. *Ecological Economics*, 157, 40–46
5. Mankiw, N., Phelps, E., & Romer, P. (1995). The Growth of Nations. In *Brookings Papers on Economic Activity*, Vol. 1995, No. 1, 25th Anniversary Issue. pp. 275–326.
6. *Primaire energie voorziening OECD*. <https://data.oecd.org/energy/primary-energy-supply.htm> en *Wereld BBP*, Wereldbank, <https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.MKTP.PP.KD&country=#>
7. Stern, D. I. (2011). *The Role of Energy in Economic Growth*. Crawford School Centre for Climate Economics & Policy (Paper No. 3.10)



Stiglers wet, het Mattheüseffect of selectiebias?

RICHARD STARMANS

In 2021 werd aan de van oorsprong Nederlandse econometrist Guido Imbens (1963) en de Amerikaans-Israëliëse econoom Joshua Angrist (1960) The Nobel Memorial Prize in Economic Sciences toegekend 'for their methodological contributions to the analysis of causal relationships'. Het was niet de eerste maal dat onderzoek naar formele, probabilistische benaderingen van oorzakelijkheid met een belangrijke prijs werd gehonoreerd. Zo ontving de Amerikaanse informaticus en AI-wetenschapper Judea Pearl (1936) al in 2011 de A.M. Turing Award voor zijn bijdragen aan de AI 'through the development of a calculus for probabilistic and causal reasoning'. Onlangs viel Pearl wederom in de prijzen, toen hem de BBVA Foundation Frontiers of Knowledge Award in Information and Communication Technologies 2021 werd toegekend, volgens de selectiecommissie 'for bringing a modern foundation to artificial intelligence', meer in het bijzonder betreffende 'probabilistic reasoning and the inference of causal relationships'. De BBVA Foundation 'expresses the corporate social responsibility of the BBVA Group', een financiële instelling die in 25 landen actief is. De stichting werd opgericht in 2008 en de Foundation Frontiers of Knowledge Awards is inmiddels aan haar veertiende editie toe. Een en ander geeft aanleiding tot een drietal overwegingen.

Causaliteit in de mode

Allereerst mag duidelijk zijn dat onderzoek naar causaliteit is opgeschoven naar het zenit van het wetenschappelijk onderzoek en pogingen om het begrip uit het wetenschappelijke discours te elimineren bleken uiteindelijk weinig succesvol. Lange tijd gold causaliteit als een weerbarstig en obscuur begrip, kende een moeizame genealogie met volop controversen, maar geen bruikbare formaliseringen en weinig vooruitgang. Velen stelden dan ook dat in een wetenschappelijk wereldbeeld geen plaats kan zijn voor zulk een archaïsche notie. Onder de critici schaalden zich in het verleden beroemde denkers als Ernst Mach, Bertrand Russell, Karl Pearson en Ludwig Wittgenstein en meer recentelijk Paul Churchland, maar ook diverse boegbeelden van *data science*.

Daarnaast moet worden opgemerkt dat de nog prille appreciatie vooral afkomstig is uit de hoek van informatica, AI en *cognitive science* en duidelijk minder uit de theoretische, wiskundige statistiek. Dat blijkt onder meer wanneer we onder de loep nemen door wie Pearl werd genomineerd. De lijst bevat onder meer de illustere namen van psycholoog en Nobelprijs-laureaat voor de economie Daniel Kahneman (1934), AI-wetenschapper Stuart Russell (1962), bekend van de moderne klas-

sieker *Artificial Intelligence; a modern approach*, maar ook die van *The Master Algorithm*-auteur Pedro Domingos (1965) en Vinton Gray Cerf (1943), ook wel bekend als Google's 'Chief Internet evangelist'. Dat alles wekt misschien weinig bevreemding, gelet op de langdurige moeizame samenspraak tussen statistiek en causaliteit, het verzuilde landschap van het huidige causaliteitsonderzoek en vooral Pearls haat-liefde verhouding met de statistiek. Men denke bij voorbeeld aan de Pearl-Imbens controverse en aan *The Book of Why* uit 2018, dat een bijna driehonderd pagina's tellende aanklacht tegen de statistiek bevat, waarbij Pearl in 'j'accuse-stijl' van leer trekt tegen de wijze waarop de statistiek zijns inziens meer dan 100 jaar lang vooruitgang in de wetenschap heeft tegengewerkt door causaliteit te veronachtzamen (Starmans, 2022a).

Een derde overweging overstijgt het thema van de causaliteit en betreft de status en het belang van wetenschappelijke prijzen in het algemeen. Voor sommigen bevestigen prestigieuze prijzen het sacrosancte karakter van een theorie, resultaat of discipline; zij vormen een keurmerk, een bekrachtiging van kwaliteit, van waarheid. Een lijst met laureaten is als het ware een staalkaart of synopsis van state-of-the-art kennis, de neerslag van een objectieve wijze van canonvorming. Voor anderen hebben zij slechts marginale betekenis; zij vormen een randverschijnsel, een charmant en luchthartig ritueel, dat onvermijdelijk enige willekeur en subjectiviteit met zich meebrengt en soms dan ook tot debat of polemiek kan leiden, maar dat de kern van wetenschap en kennis niet raakt of beïnvloedt. Kortom, niet iets waarmee serieuze wetenschapshistorici zich dienen bezig te houden. Toch is bij dit alles ook een meer kritische, principiële houding denkbaar. Deze lijkt vandaag de dag sterk aan belang te winnen in het licht van veranderende opvattingen van belonen en waarderen aan de universiteit en van groeiende (politieke) belangstelling voor de wisselwerking tussen wetenschap en samenleving, inclusief thema's als diversiteit, inclusiviteit, identiteit, bias en ongelijkheid. Het mechanisme van prijzentoekenning weerspiegelt bestaande machtsstructuren en traditionele vormen van 'institutionalisering van de kennis' en wie deze structuren en tradities wil begrijpen, blootleggen of veranderen, kan er moeilijk omheen.

Merton en Stigler

Dit alles klinkt wellicht enigszins zwaar op de hand en modieus, maar het is allerminst nieuw en zelfs geworteld in een respectabele traditie, die voor een aanzienlijk deel is terug te voeren tot het werk van de toonaangevende wetenschapshistoricus Robert K. Merton (1910–2003). Deze geniet in de wetenschapsgeschiedenis vooral bekendheid vanwege de veel bediscussieerde Merton-these, waarin een positief verband wordt gelegd tussen protestants piëtisme en arbeidsethos enerzijds en de opmars van de experimentele wetenschap ten tijde van de Wetenschappelijke Revolutie anderzijds. In klassieke publicaties als *Science, Technology and Society in Seventeenth Century England* (1938) en *The Sociology of Science* (1973) kiest Merton een externalistische benadering. Hij beschouwt de wetenschap primair als sociaal systeem en onderzoekt de wisselwerking tussen het wetenschappelijk bedrijf, cultuur en samenleving; de stakeholders, de instituties, de waarden. Daarbij besteedt hij vooral aandacht aan receptie, waardering en beloningssystemen en de machtsverhoudingen die hieraan ten grondslag liggen. En, aan de consequenties ervan voor de acceptatie van nieuwe of alternatieve theorieën, canonvorming, de vooruitgang van de kennis. Aan de hand van uiteenlopende historische studies ontrafelt hij misverstanden en anomalieën uit de wetenschapsgeschiedenis op een voor de onbevangen sciëntist vaak ontvullende wijze. Zo toont hij aan dat sommige fundamentele concepten en ideeën vrijwel unaniem worden geaccepteerd en overgenomen, terwijl de uitvinder ervan wordt vergeten of bewust en aantoonbaar wordt genegeerd of uitgesloten.

Tegelijkertijd vindt dikwijls een soort epische verdichting plaats. In zijn artikel *The Matthew Effect in Science: the reward and communication systems of science are considered* uit 1968 betoogt Merton dat beroemde wetenschappers vaak ten onrechte krediet krijgen voor hun bijdragen, terwijl de lagere Olympiërs het omgekeerde ervaren. Met de term 'Matthew-effect' verwijst hij openlijk naar een passage uit het Nieuwe Testament (Mattheüs 25:14-30), door economen nogal eens geciteerd ter illustratie van het verschijnsel dat de rijken steeds rijker en de armen steeds armer worden. Merton veraanschouwelijkt daarbij niet alleen dat prioriteitsgeschillen tussen wetenschappers op twijfelachtige wijze worden beslecht, maar gaat ook uitvoerig in

op het verschijnsel van de eponymie; stellingen, wetten, theorieën of methoden worden merkwaardigerwijs dikwijls getooid met de naam van een wetenschapper die er niets of vrijwel niets mee van doen heeft. Toen de historicus van de statistiek Stephen Stigler (1941) dit verschijnsel met een knipoog en een kwinkslag typeerde als 'Stiglers Wet van de Eponymie', haastte hij zich dan ook te benadrukken dat dit inzicht volledig is te herleiden tot Merton. Deze geeft allerlei 'gradaties' en modaliteiten van eponymie, en laat onder meer zien dat dikwijls een tijdgenoot meer aanspraak kan maken op een resultaat dan de vermeende ontdekker, of dat er juist sprake is van onafhankelijke, soortgelijke ontdekkingen, die min of meer simultaan plaatsvinden. (Starmans, 2012) Hoe dan ook, velen zullen de ironie en zelfspot van Stigler waarderen, maar de ondertoon is ernstig en dat geldt a fortiori voor Mertons werk. Voor vele wetenschapshistorici was zijn werk een brug te ver, een ondermijning van de historiografie en hij kreeg niet zelden het verwijt de deur open te zetten voor scepticisme, relativisme of zelfs wantrouwen tegen de rationaliteit van de wetenschap. Het is saillant dat de bij het grote publiek veel bekendere filosofen als Thomas Kuhn en Paul Feyerabend in de jaren 60 en 70 in dit opzicht in Mertons voetsporen traden, zeker ook omdat zij zich ter onderbouwing van hun bevindingen eveneens bedienden van gedetailleerde historische casestudies.

Waarheid als constructie

De ironie wil dat Merton nota bene van tijdgenoten en latere generaties de kritiek kreeg in zijn analyses niet genoeg te gaan. Voor denkers als de socioloog/techniekfilosoof Bruno Latour (1947) en de fysicus/socioloog Andrew Pickering (1948) wordt de kennis zelf als het ware een sociale constructie, er is geen objectieve werkelijkheid, wetenschappelijke kennis is volledig cultureel bepaald en wetenschappelijke feiten zijn primair een sociaal voortbrengsel of produkt. Pickering zette met zijn *Constructing Quarks: A Sociological History of Particle Physics* uit 1984 een belangrijke stap en geldt als een van de grondleggers van het sociaal-constructivisme. Bruno Latour werd bekend met vroege, welhaast antropologische studies als *Laboratory Life* (1979) en *Science in Action* (1986), maar zou een steeds strengere socio-technische invalshoek kiezen,

waarbij claims van objectiviteit worden losgelaten, onderzoekers agnostisch ten opzichte van elke waarheidsclaim dienen te zijn en een methodologie van scepticisme en het ontmantelen van vooroordelen voorop staat. Vanuit zijn actor-network theorie laat hij zien dat het succes van wetenschappelijke praktijken ten opzichte van andere sociale praktijken vooral in de schaalbaarheid van hun netwerk ligt. Deze hier enigszins gechargeerd geschetste benadering mag een antidotum lijken tegen een naïef geloof in objectiviteit en waardenvrijheid van wetenschap, maar kan ook tot excessen leiden. Dat laatste is althans de visie van de Amerikaanse natuurkundige Alan Sokal (1955), die in de jaren 80 sympathiseerde met het sandinistische, socialistische regime in Nicaragua en ook in het land ging werken. Toen hij in de jaren 90 terugkeerde naar de VS merkte hij dat de universiteiten in de greep waren gekomen van een merkwaardige links-liberale ideologie, die hij met het postmodernisme associeerde. Volgens hem heerste met name bij de humaniora en de sociale wetenschappen een cultureel-relativisme, waarbij wetenschap slechts een narratief is te midden van vele andere, waarbij elke onderdrukte discipline of groep haar eigen waarheid kan claimen en politiek activisme het zoeken naar waarheid heeft vervangen. Toen hij in 1996 als pastiche een nonsens-artikel vol wartaal aanbod aan het tijdschrift *Social Text* en dit zonder bezwaar werd gepubliceerd, was de Sokal-affaire geboren. In zijn *Impostures Intellectuelles* uit 1997 fileerde Sokal boegbeelden van de Franse filosofie, waaronder Latour, die hij allen verweet terug te grijpen op resultaten uit de natuurwetenschappen, zonder hier ook maar iets van te begrijpen. De kwestie zou leiden tot de notoire *science wars*, die gelet op het huidige verharde klimaat aan de Amerikaanse universiteiten nog steeds voortduren (Starmans, 2022b).

Het moge duidelijk zijn dat dit korte essay geen verstrekkende conclusies of stellingnames toelaat. Feit is dat veranderende opvattingen over de wisselwerking tussen wetenschap en samenleving een evolutie van wetenschappelijke praktijken zichtbaar maken: onderzoeksagenda's, subsidieprogramma's, canonvorming, benoemingscriteria, samenstelling editorial boards, belonen en beoordelen. Ook de mechanismen voor prijzentoekenning ontkomen hier niet aan. Wie meent dat het met het Mattheüs-effect en verwante thema's wel meevalt, zal waarderen dat vanzelfsprekendheden binnen de wetenschap onder de loep worden genomen, al was het alleen

maar vanuit methodologisch oogpunt om selectiebias en verwante vormen van misrepresentatie van een populatie of domein te verminderen.

Epiloog

Beperken we ons even tot de statistiek, dan is genoegzaam bekend dat fundamenteel statistisch onderzoek dikwijls moeizaam of niet met belangrijke prijzen gehonoreerd wordt, niet alleen binnen de wiskunde, maar ook wanneer het gaat om een ruimer, interdisciplinair domein. Daarnaast is het een feit dat statistische bijdragen sedert de opkomst van het vak in de tweede helft van de 19e eeuw er lange tijd bekaaid vanaf kwamen in de traditionele wetenschapshistorische en -filosofische literatuur. In de canonvorming van de ideeëngeschiedenis bleven zij daarmee stelselmatig onderbelicht. Daarin kwam ongeveer veertig jaar geleden enige verandering, maar dan toch vooral door het werk van gespecialiseerde historici van de statistiek, onder wie voornoemde Stephen Stigler.

Tot slot nog even dit. In 1922, precies 100 jaar geleden ontving Albert Einstein de Nobelprijs voor Natuurkunde, zoals bekend niet voor de relativiteitstheorie en merkwaardigerwijs ook nog eens met terugwerkende kracht voor het jaar 1921. Wie nog in de veronderstelling mocht verkeren dat de toekenning van 's werelds meest prestigieuze prijs als vanzelfsprekend transparant, rationeel en eerlijk verloopt, wordt ruw wakker geschud in een tweetal recente, historische artikelen in het *Nederlands Tijdschrift*

voor *Natuurkunde* (Van Lunteren, 2021; Kox, 2022). Niets menselijks bleek de commissieleden vreemd; persoonlijke voorkeuren en antipathieën, oude rekeningen die nog moesten worden vereffend, verborgen agenda's en soms ronduit bizarre compromissen speelden een niet te onderschatten rol. Zelfs Einstein kreeg ermee te maken. Hoe dan ook, bij een al te idyllische voorstelling van de sacrale status en betekenis van wetenschappelijke prijzen kunnen gedetailleerde historische case-studies soms een welkome nuancering bieden.

LITERATUUR

- Kox, A. (2022). Terug in de tijd. *Nederlands Tijdschrift voor Natuurkunde*, 1, 31–32.
- Lunteren, F. van. (2021). De wrok van Arrhenius; De beginjaren van de Nobelprijzen voor de Natuur- en Scheikunde, *Nederlands Tijdschrift voor Natuurkunde*, 1, 14–17.
- Starmans, R. J. C. M. (2012). Betwiste paradigma's en vermeende revoluties; De moeizame ironie van Stiglers Wet. *Filosofie-Tijdschrift*, 22(2).
- Starmans, R. J. C. M. (2022a). De reikwijdte van de counterfactual; Over causaliteit, potential outcomes en grafische modellen. *STAtOR*, 23(1).
- Starmans, R. J. C. M. (2022b). De Sokal-affaire in retrospectief; Over parodieën, crises en wantrouwen in wetenschap. *Filosofie-Tijdschrift*, 32(1).

RICHARD STARMANS is verbonden aan de Faculteit Bèta-wetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht en aan Tilburg University. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica.

E-mail: starmans@cs.uu.nl

Good news from Young Statisticians



After another long lockdown-winter, the Young Statisticians section was finally able to organize its first event of 2022, which took place on April 7th: a Statistics Café on causality. Oisín Ryan, Bart Eggen and Jesse Krijthe each gave interesting talks on different aspects of this central topic within statistics.

We aim to soon plan a company visit to an organization that uses statistics and data science in daily practice. (If you have any suggestions for this, don't hesitate to contact us!) To keep informed on our future events, or to subscribe to our newsletter, make sure to visit our website: vvsor.nl/young-statisticians.

PEILINGPRAKTIJKEN

Rommel niet met de assen van een grafiek



Om de omvang van allerlei verschijnselen uit een grafiek te kunnen aflezen, heb je assen nodig. Die assen horen voorzien te zijn van maatstreepjes en bij die maatstreepjes horen waarden te staan. Bij het maken van een grafiek moet je die assen op een correcte manier tekenen. Doe je dat niet dan leidt de grafiek al gauw tot verkeerde interpretatie.

Een veel voorkomende fout is de as niet laten beginnen bij 0. Als de as een omvang, hoeveelheid, waarde, duur, aantal of percentage moet aanduiden, dan is er een natuurlijk nulpunt. De as moet dan bij dit nulpunt beginnen. Als dat niet het geval is, dan lijken allerlei verschijnselen groter dan ze in werkelijkheid zijn. Er is dus sprake van overdrijving. De grafiek is suggestief. Aan de hand van een paar voorbeelden laten we zien hoe een grafiek je op het verkeerde been kan zetten.

De nieuwe Amerikaanse zorgverzekering Obamacare

Een eerste voorbeeld van gerommel met de as komt van Fox News. Dat is een van de grote nieuwszenders in de

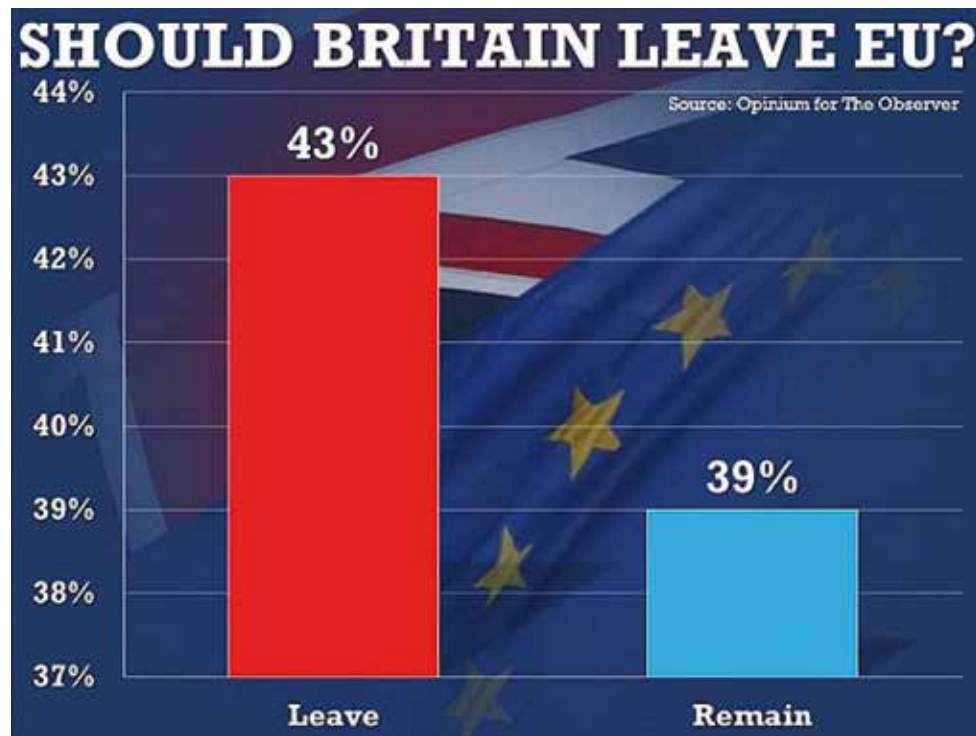
Verenigde Staten. De zender is de stem van de rechtse, conservatieve Amerikanen. In de berichtgeving liet Fox geen mogelijkheid onbenut om de regering van president Barack Obama in een kwaad daglicht te zetten.

Het was een van de belangrijke doelen van de Amerikaanse president Barack Obama om het systeem van gezondheidsverzekeringen te hervormen. Daarvoor voerde hij een nieuwe wet in: de *Patient Protection and Affordable Care Act*. De wet verplichtte iedere Amerikaan een zorgverzekering af te sluiten. En de zorgverzekeraars mochten niemand weigeren. In de volksmond werd deze wet *Obamacare* genoemd. De Republikeinen waren fel tegen de wet. Daarom probeerde Fox News de invoering van Obamacare af te schilderen als een volledige flop. De regering Obama mikte erop dat op 31 maart 2014 zo'n 7 miljoen Amerikanen zich moesten hebben aangemeld voor de nieuwe zorgverzekering. Dus probeerde Fox News op 27 maart 2014 met de linker grafiek in figuur 1 aan te geven dat dit doel bij lange na niet was gehaald.

Het verschil tussen de twee staven in het linker staafdiagram is inderdaad bijzonder groot. De grafiek suggereert dat het aantal aanmeldingen zelfs minder is dan de helft van de doelstelling. Maar een nadere blik



Figuur 1. Fox News rommelt met een grafiek over Obamacare. Bron: Fox News, 2014



Figuur 2. De Daily Mail rommelt met een grafiek over een Brexit-peiling. Bron: Daily Mail, 2016

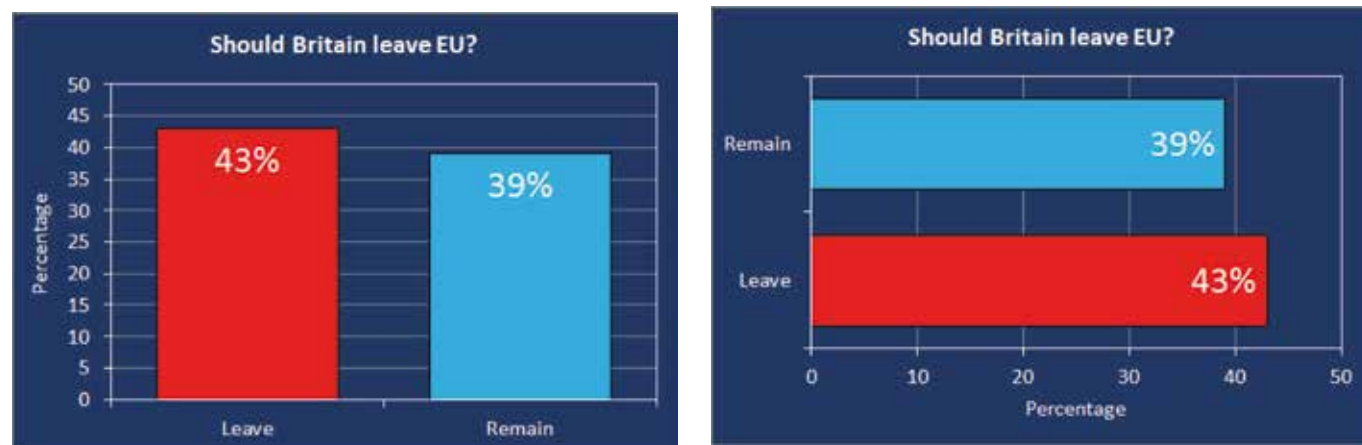
op de cijfers leert dat in feite het verschil helemaal niet zo groot is: er zijn immers al 6 miljoen aanmeldingen en dat is maar 1 miljoen minder dan de doelstelling van ruim 7 miljoen. De staven weerspiegelen dus niet de werkelijke cijfers. Dat komt omdat de verticale as niet bij 0 begint. Dat is niet eenvoudig vast te stellen omdat die verticale as helemaal geen schaalverdeling heeft gekregen. Daarmee is dit een bijzonder slechte grafiek. Als je niet goed oplet, wordt je al snel op het verkeerde been gezet.

Er was veel kritiek van deskundigen op de grafiek van Fox News. Dat leidde ertoe dat Fox News de grafiek een paar dagen later (op 1 april 2014) corrigeerde. De rechter

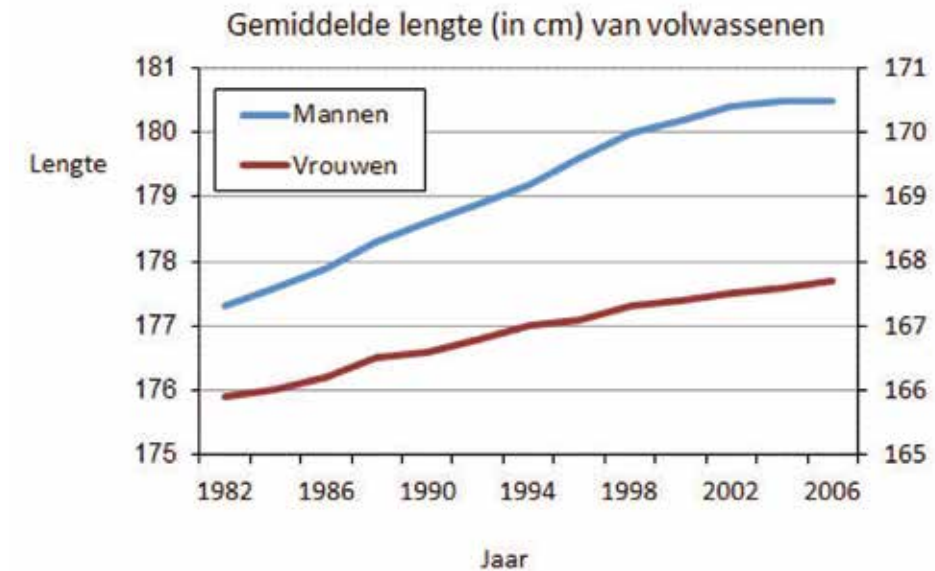
grafiek in figuur 1 is de gecorrigeerde versie van de grafiek. De verticale as heeft nu een keurige schaalverdeling. En die schaalverdeling begint, zoals het hoort, bij 0. Het verschil tussen beide staven is nu veel minder groot. De grafiek geeft een veel realistischer beeld van de situatie.

Willen de Britten uit de EU?

Een tweede voorbeeld van het manipuleren van de as van een grafiek komt van een Britse peiling over *Brexit*, het verlaten van de Europese Unie (EU) door het Verenigd Koninkrijk (VK). De grafiek in figuur 2 is afkomstig uit de



Figuur 3. Twee versies van de gerepareerde grafiek over de Brexit-peiling



Figuur 4. Het CBS rommelt met de assen van een grafiek met de lengte van mannen en vrouwen. Bron: CBS-Webmagazine, 17 januari 2008

Britse krant *Daily Mail* van 3 april 2016. De boodschap lijkt duidelijk: een groot deel van de bevolking (de rode staaf) wil dat het VK de EU verlaat (*Brexit*). Slechts een klein deel van de bevolking (de blauwe staaf) wil dat het VK in de EU blijft.

Helaas, er is wat mis met de grafiek: de verticale as begint niet met 0, maar bij 37%. Door deze truc lijkt het verschil tussen 'Leave' and 'Remain' veel groter dan het in werkelijkheid is. De krant overdrijft dus het verschil.

Figuur 3 bevat twee gerepareerde versies van het staafdiagram in figuur 2. De belangrijkste wijziging is dat de as nu bij 0 begint. Daardoor is het verschil tussen 'Leave' en 'Remain' veel kleiner. Dat klopt wel, want het verschil is maar vier procentpunten.

Een tweede wijziging is dat de rechter grafiek in figuur 3 gedraaid is zodat de staven nu horizontaal zijn getekend in plaats van verticaal. Een grafiek als deze biedt meer ruimte voor (horizontale) teksten en bevordert daarmee de leesbaarheid. De rechter grafiek heeft mijn voorkeur.

Zijn mannen veel langer dan vrouwen?

Ten slotte nog een derde voorbeeld van gerommel met de as. Het gaat om een grafiek die in 2008 door het Centraal Bureau voor de Statistiek (CBS) is gemaakt. In figuur 4 is deze grafiek gereproduceerd.

Een eerste blik op deze grafiek roept de indruk op van een sterke toename van de gemiddelde lengte. Dat geldt vooral voor mannen (de blauwe lijn) en in iets mindere mate voor vrouwen (de rode lijn). Verder wekt de grafiek

de suggestie dat vrouwen gemiddeld maar half zo lang zijn als mannen.

Er zijn twee problemen met de grafiek. In de eerste plaats zijn er twee verticale assen. De linker as is voor de mannen en de rechter as moet je gebruiken voor de vrouwen. De twee assen lopen niet synchroon. Daardoor kun je mannen en vrouwen niet met elkaar vergelijken. In de tweede plaats beginnen beide assen niet bij het natuurlijke nulpunt. De linker verticale as begint bij 175 en de rechter bij 165.

Als je dezelfde as zou gebruiken voor mannen en vrouwen en bovendien die as bij 0 zou laten beginnen, dan krijg je een veel rustiger beeld van de ontwikkeling van de gemiddelde lengte van mannen en vrouwen. De groei gaat minder hard en de verschillen tussen mannen en vrouwen zijn minder groot.

Conclusie

Je ziet regelmatig grafieken langskomen waarin gemiddeld is met de verticale as. Zulke grafieken roepen meestal een verkeerd beeld op van de situatie. Daarom is het goed om altijd eerst even naar de as van een grafiek te kijken voordat je hem gaat interpreteren.

JELKE BETHLEHEM werkte bij het CBS en is emeritus hoogleraar aan de Universiteit Leiden. Hij is een expert op het gebied van steekproeven, vragenlijsten en weergave van onderzoeksresultaten. Deze onderwerpen behandelt hij regelmatig in zijn blog. E-mail: mail@jelkebethlehem.nl



Jelke Bethlehem. (2022). *Het Grafiekenboek*. Amsterdam University Press. Isbn 978 94 637 2098 4, € 29,99, 222 blz.

Onze columnist Jelke Bethlehem heeft een zeer brede belangstelling, maar volgers van zijn blog¹ weten dat grafieken zijn bijzondere aandacht hebben. Ook zijn column in dit nummer gaat over grafieken. Onlangs verscheen van hem *Het Grafiekenboek* waarin hij een systematisch overzicht geeft van wat er allemaal bij het maken van een goede grafiek komt kijken. Redacteur Gerrit Stemerding heeft met de auteur gesproken. De toon is informeel, tenslotte komen zij elkaar al sinds 1972 tegen in allerlei besturen, commissies en werkgroepen.

Grafieken kunnen een zegen zijn, maar ook een ramp

GERRIT STEMERDINK

Jelke, allereerst mijn complimenten voor de zeer uitgebreide en systematische presentatie in het boek. Was er een aanleiding om, na de vele blogs hierover, alles eens samen te vatten?

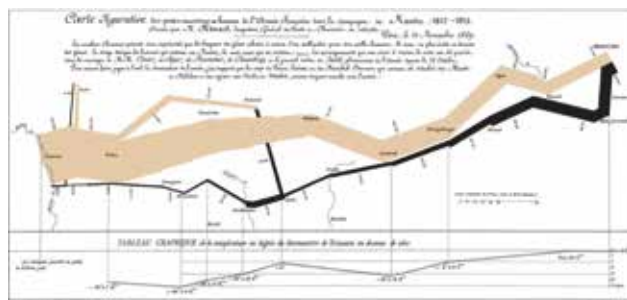
Jelke Bethlehem. Bedankt voor de aardige woorden. Ik ben inderdaad al een tijd met grafieken bezig. Ik kan me herinneren dat ik in 1984 een kritische notitie schreef over de kleurgrafieken in het *Statistisch Zakboek* van het CBS. In die tijd begon deze bestseller van het CBS altijd met zestien ongenummerde bladzijden met grafieken in kleur. In die notitie ging het over het gebruik van kleur, de vormgeving van grafieken en de boodschap die je ermee kunt overbrengen.

Aanvankelijk was het maken van grafieken ambachtelijk werk, de eerste grafieken waren ware kunstwerken.

Denk maar eens aan de grafiek over de veldtocht van Napoleon naar Rusland die in 1869 werd gemaakt door de Franse ingenieur Joseph Minard. Volgens sommige statistici is dit een van de mooiste grafieken ooit.

Met de opkomst van de computer werd het steeds eenvoudiger om grafieken te maken. Het gebruik van grafieken nam een grote vlucht met de opkomst van de PC met grafische schermen. Er kwamen dus steeds meer grafieken, maar helaas waren niet al die grafieken goede grafieken. Met een foute grafiek kun je vrij eenvoudig gebruikers op het verkeerde been zetten. Ontwerpers van grafieken doen dit meestal niet bewust. Problemen ontstaan soms doordat ontwerpers van grafieken zich meer bezighouden met een leuke vormgeving dan met het overbrengen van de correcte statistische boodschap in de gegevens. Dus was het tijd om alles over grafieken nog eens op een rij te zetten. En dat is *Het Grafiekenboek*.

Een bijzonder aspect van je boek is dat je bij iedere soort grafiek allereerst een voorbeeld geeft van hoe het niet zou moeten. Vervolgens leg je uit wat er precies fout is en hoe het beter kan. Daarin verschilt je boek van de gebruikelijke wijze om leerstof te presenteren. Heeft het je veel moeite gekost om al die foute voorbeelden te vinden?



Nee, het is niet zo moeilijk om foute voorbeelden van grafieken te vinden. Grote organisaties zoals het CBS en het SCP publiceren meestal wel goede grafieken, al ken ik ook wel slechte grafieken van Eurostat, het statistische bureau van de Europese Unie. En de grotere en betere marktonderzoeksbureaus produceren meestal ook wel goede grafieken. Maar in de media (tv, sociale media) kom je veel grafieken tegen die niet door de beugel kunnen. Een slecht voorbeeld is de Amerikaanse nieuwszender Fox News. Die probeert grafieken zodanig te manipuleren dat de Democraten en Democratische presidenten (Obama, Biden) het slecht lijken te doen. Bij Fox News lijkt het publiceren van slechte grafieken boos opzet, maar bedenk dat slechte grafieken ook per ongeluk kunnen ontstaan door onervarenheid van de ontwerpers. Het zijn ook de grafieken in sommige Engelse kranten die de toets der kritiek niet kunnen doorstaan. Zo stond ooit in de Daily Mail een suggestieve grafiek die zou moeten aangeven dat een heel grote meerderheid van de Britten voor Brexit was. En dat terwijl de percentages voor- en tegenstanders in werkelijkheid vlak bij elkaar lagen.

Wij zijn beide begonnen in een tijd dat de regeldrukker het gebruikelijke uitvoermedium was. Maar ook toen werden er al, zij het primitieve, grafieken geproduceerd. Meestal waren dat kolomdiagrammen opgebouwd uit hoofdletters X of iets dergelijks, grafieken voor publicaties werden door een tekenaar gemaakt. Met de komst van de laserprinters, midden jaren 80, kreeg ook de gebruiker de mogelijkheid zelf betere grafieken te maken. Heb jij het idee dat toen pas goed de wildgroei aan 'foute' grafieken ontstond?

Inderdaad, aanvankelijk waren veel grafieken handwerk. De eerste MS-DOS computers konden immers geen plaatjes op het scherm zetten. Dat werd beter toen er plotters op de markt kwamen. Dat waren wel grote, trage en dure apparaten. Het gebruik van de grafieken nam snel toe naarmate er meer software op de markt kwam waarmee je grafieken kon maken. Denk aan software zoals MS Excel en SPSS. En het gebruik van grafieken nam nog meer toe met de opkomst van sociale media. Dat is niet verbazingwekkend, want met een goede grafiek kun je bij uitstek de statistische boodschap in de data goed overbrengen aan het grote publiek.

In de tijd dat we samen actief waren in de Sectie Statistische Programmatuur van de VVSOR maakten de leden van die sectie zich regelmatig zorgen over het mogelijk onjuiste gebruik van geavanceerde programmatuur door

leken. 'Statistische Programmatuur: een Zegen of een Ramp?' was de vraag die we ons stelden, het was zelfs een keer het motto voor onze jaarvergadering. Vandaag de dag kunnen grafieken óók heel gemakkelijk worden gemaakt in programma's als SPSS en Excel. Ben je het met me eens dat we ons deze vraag opnieuw moeten stellen, maar dan nu toegespitst op grafieken? Was de toename van dit soort mogelijkheden wellicht mede aanleiding voor je boek?

Grafieken kunnen een zegen zijn, maar ook een ramp. Het zou goed zijn als onderzoekers zich meer bewust waren van de gevaren van slechte grafieken. Ze zouden goed moeten nadenken over hun keuze voor een type grafiek en de manier waarop ze hem vormgeven. Het is nu te makkelijk om een foute grafiek te maken. Maar ik ben zeker niet tegen het gebruik van grafieken. Een goede grafiek is een heel krachtig hulpmiddel om je gegevens beter te leren begrijpen je resultaten te tonen aan een groot publiek. Dus ik zou je uitspraak eigenlijk moeten nuanceren 'Goede grafieken zijn een zegen en slechte grafieken zijn een ramp'. Dit is de boodschap die ik uit wil dragen in mijn boek.

Misschien is het goed nog even te melden dat er ook nog andere zaken zijn in de statistiek waarover we ons bezorgd kunnen maken. En dan heb ik het over het gebruik van peilingen. Daarbij verzamel je data via een steekproef uit de bevolking. Als die steekproef niet netjes (aselect) trekt, loop je een levensgroot risico om verkeerde conclusies uit een onderzoek te trekken. Je steekproef is dan immers niet representatief. En slechte vraagstelling helpt ook niet mee. Kortom, nog meer leed. Meer over peilingen (polls) kun je vinden in mijn boek *Understanding Public Opinion Polls*² uit 2018.

Op het gevaar af dat dit als 'vriendendienst' wordt gezien wil ik toch dit boek bij de lezers van STATOR van harte aanbevelen. Het heeft volgens mij de potentie net zo'n klassieker te worden als Visual Revelations van Howard Wainer³ dat 25 jaar geleden verscheen, of The Visual Display of Quantitative Information van Edward Tufte⁴ van bijna 40 jaar geleden. Misschien moet je maar eens een Engelstalige versie overwegen.

NOTEN

1. <https://peilingpraktijken.nl/weblog>
2. Bethlehem, J. (2018). *Understanding Public Opinion Polls*. CRC Press.
3. Wainer, H. (1997). *Visual Revelations*. Copernicus Springer-Verlag.
4. Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press.

Lof der luiheid

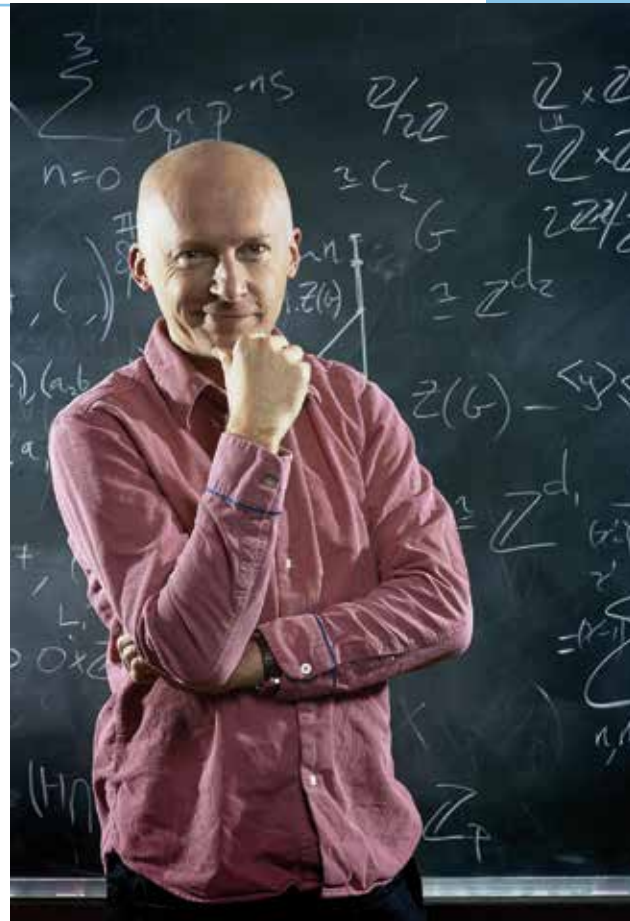
Wiskundigen zijn lui! Dat is geen bedenkensel van mijzelf, hoewel ik het er mee eens ben. Ik ontleen het aan Marcus du Sautoy, hoogleraar aan de universiteit van Oxford. Hij heeft een aantal boeken geschreven om wiskunde te populariseren. Je zou hem de Britse Ionica Smeets kunnen noemen; hun leeropdracht is ook grotendeels gelijklopend.

Ik ben een enorme fan van hem – overigens ook van haar – en heb veel van zijn boeken met groot plezier gelezen. Maar vooral heb ik genoten van zijn zeer recente (oktober 2021) boek *Thinking Better*.

Kern van zijn betoog is dat een wiskundige in principe lui is, het ultieme streven is zo kort mogelijke bewijzen en het vermijden van overbodig rekenwerk door wiskundig te denken. Een bekend voorbeeld is de anekdote over de schooljongen Gauß die de cijfers van 1 tot 100 moest optellen, hij stelde dat 1 plus 100 hetzelfde was als 2 plus 99 en dat 50 maal herhaald. Het tot verbijstering van zijn leraar ogenblikkelijk gegeven antwoord was dus 5050: 50 maal 101.

Christiaan Huijgens is een ook goed voorbeeld van wiskundig denken. In *Van Reekening in Speelen van Geluck** behandelt hij in een aantal opvolgende stellingen problemen uit de kansrekening. Beginnend met een eenvoudig probleem gaat hij steeds een stapje verder en brengt ieder nieuw probleem terug op een eerder bewezen stelling. Een wiskundige techniek die ons maar al te zeer vertrouwd is.

Wat die luiheid van wiskundigen betreft – zelf noemen we dat liever efficiëntie – is er een mooie grap om dat te illustreren. Die grap gaat over het verschil tussen een wiskundige en een natuurkundige. Beide worden onderworpen aan een identieke proef: ze komen in een ruimte met een kraan, een gaspits en een lege ketel. De opdracht is een ketel kokend water te maken. Beide vullen de ketel onder de kraan, zetten hem op de gaspits, steken die aan en wachten tot het water kookt: geen verschil tussen



Marcus du Sautoy

de beide wetenschappers. Dan komt het tweede deel van de proef, ze komen weer in die ruimte, maar nu is de ketel niet meer leeg, maar al gevuld met koud water. De natuurkundige zet hem op de gaspits, steekt die aan en wacht tot het water kookt. De wiskundige echter giet de ketel leeg en zegt 'nu is het teruggebracht tot het vorige probleem en dat is al opgelost'.

Overigens valt het met die luiheid wel mee, je moet vaak hard werken om je lui te mogen noemen.

GERRIT STEMERDINK is eindredacteur van STATOR.
E-mail: gjstemerdink@hotmail.com

* *Reekening in Speelen van Geluck*. Oorspronkelijk brieven van Christiaan Huijgens aan zijn leermeester, de Leidse hoogleraar Van Schooten, door deze als aanhangsel in een boek opgenomen. Brieven van 1656, boek van 1657 (Latijn) en 1660 (Nederlands).

LITERATUUR

Du Sautoy, M. (2021). *Thinking Better; The art of the shortcut in math and life*. Basic Books, ISBN 9780008393922.

An Italian CSI drama

Social media, a broken legal system, and Micky Mouse statistics

Foto: stux via Pixabay

This article is an adjusted version of a blog of Richard Gill on a case of Daniela Poggiali, an Italian nurse suspected of having killed around ninety patients. When Gill read about this in the newspapers he offered her lawyer support on the statistics of the case.

RICHARD GILL

The title of this article – 'Social media, a broken legal system, and Micky Mouse statistics' – might refer to the very, very famous trials of Amanda Knox in the case of the murder of Meredith Kercher. However, I am writing about a case that is much less known outside of Italy (neither victim nor alleged murderer was a rich American girl). This is the case of Daniela Poggiali, a nurse suspected by the media and accused by prosecution experts of having killed around 90 patients in a two-year killing spree terminated by her arrest in April 2014. She has just been exonerated after a total of three years in prison with a life sentence as well some months of pre-trial detention. This case revolved around statistics of an increased death rate

during the shifts of a colourful nurse. I was a scientific expert for the defence, working with an Italian colleague, Julia Mortera (Università degli Studi Roma Tre), later assisted by her colleague Francesco Dotto.

Piet Groeneboom and I worked together on the statistics of the case of Lucia de Berk, see our paper in *Chance* (Gill, Groeneboom, and Jong 2018). In fact, it was remarkable that the statistical community in the Netherlands got so involved in that case. A Fokke and Sukke cartoon entitled "Fokke and Sukke know it intuitively" had the exchange "The probability that almost all professors of statistics are in agreement ... is obviously very small indeed".

Indeed, it wasn't. That was one of the high points of my career. Another was Lucia's final acquittal in 2011, at which the judges took the trouble to say out loud, in public, that the nurses had fought heroically for the lives of their patients; lives squandered, they added, by their doctors' medical errors.

At that point, I felt we had learnt how to fight miscarriages of justice like that, of which I rapidly became involved in several. So far, however, with rather depressing results. Till a couple of months ago. This story will not have much to do with mathematics. It will have to do with simple descriptive statistics, and I will also mention the phrases "*p*-value" and "Bayes' rule" a few times.* I think it is important for mathematicians in general to know more about what statisticians can do – not so much through using deep and exciting mathematics, though that does happen too, of course – but because one of the skills of a professional statistician is the abstraction of messy real-world problems involving chance and data. It's not for everybody. Many mathematical statisticians prefer to prove theorems, just like any other mathematician. In fact, I often do prefer to do that myself, but I like more being able to alternate between the two modes of activity, and I do like sticking my nose into other people's business, and learning about what goes on in, for instance, law, medicine, or anything else. Each of the two activity modes is a nice therapy for the frustrations which inevitably come with the other.

The case of Daniela Poggiali

The Daniela Poggiali case began, for me, soon after the 8th of April, 2014, when it was first reported in international news media. A nurse at the *Umberto I* hospital in

the small town of Lugo, not far from Ravenna, had been arrested and was being investigated for serial murder. She had had photos of herself taken laughing, close to the body of a deceased patient, and these "selfies" were soon plastered over the front pages of tabloid media. Pretty soon, they arrived in *The Guardian* and *The New York Times*. The newspapers sometimes suggested she had killed 93 patients, sometimes 31, sometimes it was other large numbers. It was suspected that she had used Potassium Chloride on some of those patients. An ideal murder weapon for a killer nurse since easily available in a hospital, easy to give to a patient who is already hooked up to an IV drip, kills rapidly (cardiac arrest – it is used in America for executions), and after a short time hard to detect. After death, it redistributes itself throughout the body where it becomes indistinguishable from a normal concentration of Potassium.

Many features of the case reminded me strongly of the case of Lucia de Berk in the Netherlands. In fact, it seemed very fishy indeed. I found the name of Daniela's lawyer in the online Italian newspapers, Google found me an email address, and I sent a message offering support on the statistics of the case. I also got an Italian statistician colleague and good friend, Julia Mortera, interested. Daniela's lawyer was grateful for our offer of help. The case largely hinged on a statistical analysis of the coincidence between deaths on a hospital ward and Daniela's shifts there. We were emailed pdfs of scanned pages of a faxed report of around 50 pages containing results of statistical analyses of times of shifts of all the nurses working on the ward, and times of admission and discharge (or death) of all patients, during much of the period 2012–2014. There were a further 50 pages (also scanned and faxed) of appendices containing print-outs of the raw data submitted by hospital administrators to police investigators. Two huge messy Excel spreadsheets.

The authors of the report were Prof. Franco Tagliaro (Univ. Verona) and Prof. Rocco Micciolo (Univ. Trento). The two are respectively a pathologist/toxicologist and an epidemiologist. The epidemiologist Micciolo is a professor in a social science department, and member of an interfaculty collaboration for the health sciences. We found out that the senior and more influential author Tagliaro had published many papers on toxicology in the forensic science literature, usually based on empirical studies using data sets provided by forensic institutes. Occasionally, his friend Micciolo turned up in the list of authors and had supplied statistical analyses. Micciolo describes himself as a biostatistician. He has written Italian language textbooks on exploratory data-analysis with the statistical



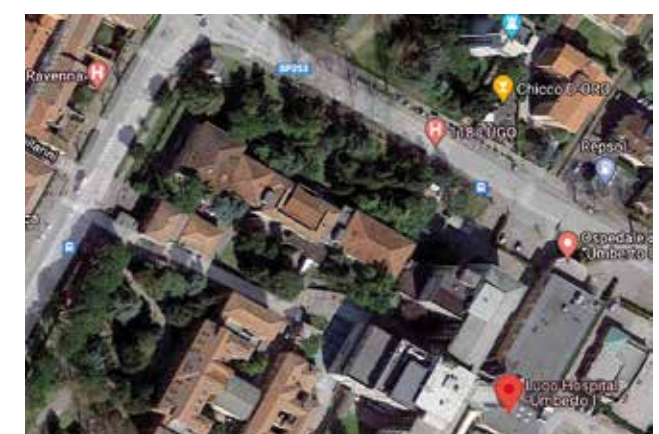
www.fokke.nl

package "R" and is frequently the statistician-coauthor of papers written by scientists from his university in many different fields including medicine and psychology. They both had decent H-indices, their publications were in decent journals, their work was mainstream, useful, "normal science". They were not amateurs. Or were they?

Daniela Poggiali worked on a very large ward with very many very old patients, many suffering terminal illnesses. Ages ranged from 50 up to 105, mostly around ninety. The ward had about 60 beds and was usually quite fully occupied. Patients tended to stay one to two weeks in the hospital, admitted to the hospital for reasons of acute illness. There was on average one death every day; some days none, some days up to four. Most patients were discharged after several weeks in hospital to go home or to a nursing home. It was an ordinary "medium care" nursing department (i.e., not an Intensive Care unit).

Some very simple statistics showed that the death rate on days when Poggiali worked was much higher than on days when she did not work. A more refined analysis compared the rate of deaths during the hours she worked with the rate of deaths during the hours she was not at work. Again, her presence "caused" a huge excess, statistically highly significant. A yet more refined analysis compared the rate of deaths while she was at work in the sectors where she was working with the rate in the opposite sectors. What does this mean? The ward was large and spread over two long wings of one floor of a large building, "Blocco B", probably built in the sixties.

Between the two wings were central "supporting facilities" and a main stairwell above the main entrance. Each wing consisted of many rooms (each room with several beds), with one long corridor through the whole building, see the floor plan. Sector A and B rooms were in one wing, first A and then B as you went down the



The long building at the top: 'Block B' of Umberto I hospital, Lugo

corridor from the central part of the floor. Sector C and Sector D rooms were in the other wing, opposite to one another on each side of the corridor. Each nurse was usually detailed in her shifts to one sector, or occasionally to the two sectors in one wing. While working in one sector, a nurse could theoretically easily slip into a room in the adjacent sector. Anyway, the nurses often helped one another, so they often could be found in the "wrong sector", but not often in the "wrong wing".

Tagliaro and Micciolo (in the sequel: TM) went on to look at the death rates while Daniela was at work in different periods. They noticed that it was higher in 2013 than in 2012, even higher in the first quarter of 2014, then – after Daniela had been fired – it was much, much less. They conjectured that she was killing more and more patients as time went by, till the killing stopped dead on her suspension and arrest.

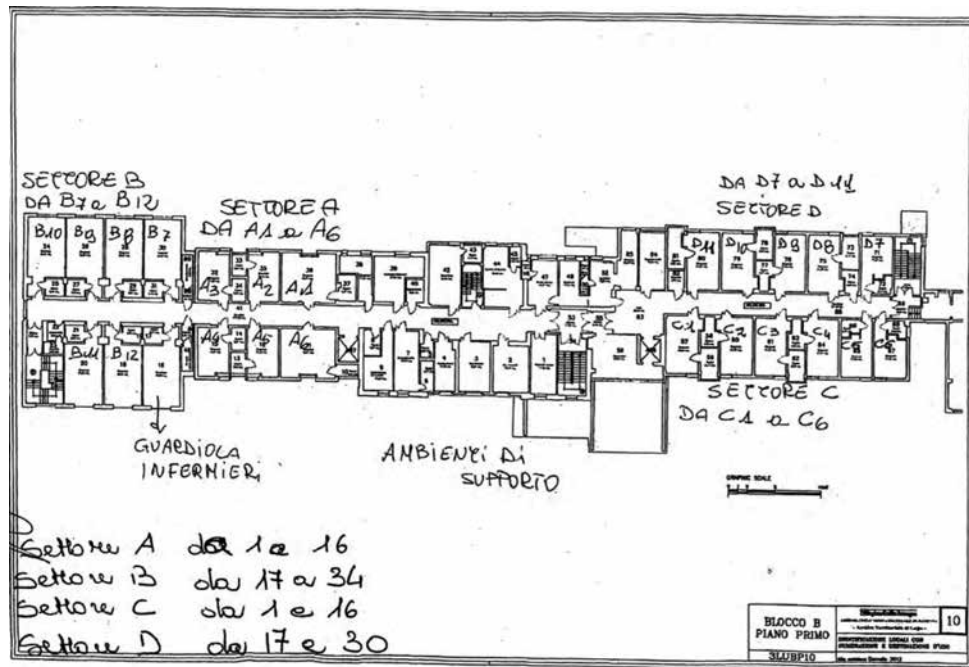
TM certainly knew that in theory, other factors might be the cause of an increased death rate on Poggiali's shifts. They were proud of their innovative approach of relating each death which occurred while Daniela was at work to the question of whether it occurred in Daniela's wing or in the other. They wrote that in this way they had controlled for confounders, taking each death to provide its own "control". (Similarly in the case of Lucia de B., statistician Henk Elffers had come up with an innovative approach. In principle it was not a bad idea though all it showed was that nurses are different). TM did not control for any other confounding factors at all. In their explanation of their findings to the court they repeatedly stated categorically that the *association* they had found must be *causal*, and Daniela's presence was the cause. Add to this that their clumsy explanation of *p*-values might have mislead lawyers, journalists and the public. In such a case, a *p*-value is the probability of what you see (more precisely, of *at least* what you see), assuming pure chance. That is not the same as the probability that pure chance was the cause of what you see – the fallacy of the transposed conditional, also known as "the prosecutor's fallacy".

Exercise to the reader: when is this fallacy not a fallacy? Hint: revise your knowledge of *Bayes' rule*: posterior odds equals prior odds time likelihood ratio.

$$\frac{\Pr(H_p | Evidence)}{\Pr(H_d | Evidence)} = \frac{\Pr(Evidence | H_p)}{\Pr(Evidence | H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}$$

Posterior Odds Likelihood Ratio Prior Odds

Bayes rule in odds form. *p* and *d* stand for 'prosecution' and 'defence' respectively, *H* stands for 'Hypothesis'



Plan of ward (from TM's report)

We asked Tagliaro and Micciolo for the original Excel spreadsheets and for the “R” scripts (R Core Team 2013) they had used to process the data. They declined to give them to us, saying this would not be proper since they were confidential. We asked Daniela’s lawyer to ask the court to ask for those computer files on our behalf. The court declined to satisfy our request. We were finally sent just the Excel files by the hospital administration, a week before we were called to give evidence. Fortunately, with a combination of OCR and a lot of painstaking handwork, a wealthy friend of Daniela’s lawyer had already managed to help us get the data files reconstructed. We performed a lot of analyses with the help of a succession of students because extracting what we needed from those spreadsheets was an extraordinarily challenging issue. One kept finding anomalies that had to be fixed in one way or another. Even when we had “clean” spreadsheets, it still was a mess.

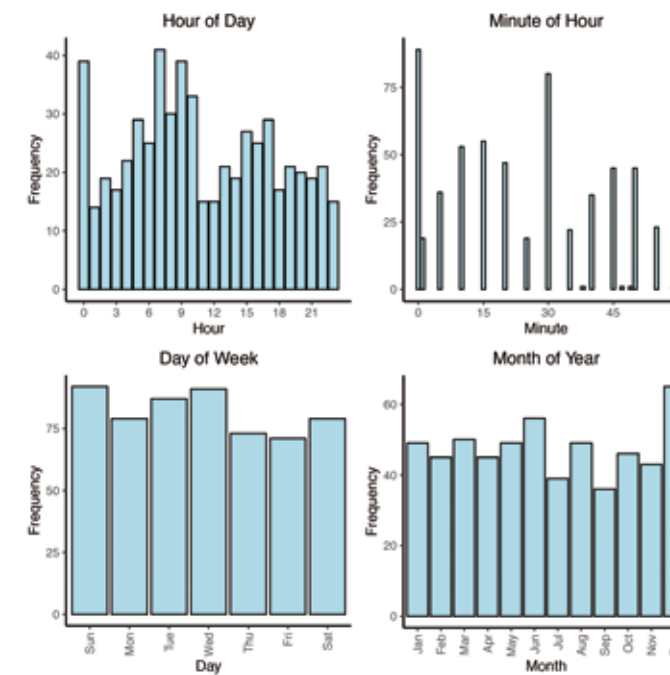
Garbage in, garbage out?

Next, we started looking for confounding factors that might explain the difference between Daniela and her colleagues, which certainly was striking and real. But was it perhaps entirely innocent?

First of all, simple histograms showed that death rates on that ward varied strongly by month, with big peaks in June and again in January. That is what one should ex-

pect. The humid heat and air pollution in the summer; or the damp and cold and the air pollution in the winter, exacerbated by winter flu epidemics. Perhaps Daniela worked more at bad times than at good times? No. It was clear that sectors A+B were different from C+D. Death rates were different but also the number of beds in each wing was different. Perhaps Daniela was allocated more often to “the more difficult” sections? It was not so clear. Tagliaro and Micciolo computed death rates for the whole ward, or for each wing of the ward, but never took account of the number of patients in each wing nor of the severity of their illnesses.

Most interesting of all was what we found when we looked at the hour of the time of death of patients who died, and the minute of the time of death of patients who died. Patients tended to die at times which were whole hours, “half past” was also quite popular. There was however also a huge peak of deaths between midnight and five minutes past midnight! There were fewer deaths in a couple of hours soon after lunch time. There were large peaks of deaths around the time of handover between shifts: 7:00 in the morning, 2:00 in the afternoon, 9:00 in the evening. The death rate is higher in the morning than in the afternoon and higher in the afternoon than at night. When you’re dying (but not in intensive care, when it is very difficult to die at all) you do not die in your sleep at night. You die in the early morning as your vital organs start waking up for the day. Now, also not surprisingly, the number of nurses on a ward is largest in the morning



Minute, hour, weekday, month of deaths (Dotto, Gill and Mortera, 2022)

when there is a huge amount of work to do; it’s much less in the afternoon and evening; and it’s even less at night. This means that a full-time nurse typically spends more time in the hospital during morning shifts than during afternoon shifts, and more time during afternoon shifts than during night shifts. The death rate shows the same pattern. Therefore, for every typical full-time nurse, the death rate while they are at work tends to be higher than when they are not at work!

Nurses aren’t authorized to officially register times of death. Only a doctor is authorized to do that. He or she is supposed to write down the time at which they have determined the patient is no longer alive. It seems that they often round that time to whole or half hours. The peak just after midnight is hard to explain. The *date* of death has enormous financial and legal consequences. The peak suggests that those deaths may have occurred anywhere in a huge time window. Whether or not doctors come to the wards on the dot at midnight and fill in forms for any patients who have died in the few hours before is hard to believe.

What is now clear is that it is mainly around the hand-over between shifts that deaths get “processed”. Quite a few times of death are so hard to know that they are shunted to five minutes past midnight; many others are located in the hand-over period but might well have occurred earlier.

Some nurses tend to work longer shifts than others. Some conscientiously clock in as early as they are allowed, before their shift starts, and clock out as late as

they can after their shift ends. Daniela was such a nurse. Her shifts were indeed statistically significantly longer than those of any of her colleagues. She very often stayed on duty several hours after the official end of the official ten-minute overlap between shifts. There was often a lot to do – one can imagine often involving taking care of the recently deceased. Not the nicest part of the job. Daniela was well known to be a rather conscientious and very hard worker, with a fiery temper, known to play pranks on colleagues or to loudly disagree with doctors for whom she had a healthy disrespect.

Incidentally, the rate of admissions to *Umberto I* hospital tumbled down after the news broke of a serial killer – and the news broke the day after the last day the serial killer was at work, together with the publication of the lurid “selfie”. The rate of deaths was slowly increasing over the two years up to then, as was in fact also the rate of admissions and the occupancy of the ward. A hospital getting slowly more stressed? Taking on more work?

No correlation without causation

If one finds a correlation between X and Y, it is a sound principle to suppose that it has a causal explanation. Maybe X causes Y, maybe Y causes X, ... and maybe W causes both X and Y, or maybe X and Y both cause Z and there has been a selection on the basis of Z. In the case of Lucia de B., her association between inexplicable incidents

and her presence on the ward was caused by her, since the definition of “unexpected and inexplicable incident” included her being there. She was already known to be a weird person and it was already clear that there were more deaths than usual on her ward. The actual reason for that was a change of hospital policy, moving patients faster from intensive care to medium care so that they could die at home, rather than in the hospital. If she was not present, then the medical experts always could come up with an explanation for why that death though perhaps a bit surprising at that moment, was expected to occur soon anyway. But if Lucia was there then they were inclined to believe in foul play because after all there were so many incidents in her shifts.

Julia and I are certain that the difference between Daniela’s death rates and those of other nurses is to a huge extent explainable by the anomalies in the data which we had discovered and by her long working hours.

Some residual difference could be due to the fact that a conscientious nurse actually notices when patients have died, while a lazy nurse keeps a low profile and leaves it to her colleagues to notice, at hand-over. We have been busy fitting sophisticated regression models to the data but this work will be reported in a specialist journal (Dotto, Gill and Mortera, 2022). It does not tell us more than what I have already said. Daniela is different from the other nurses. All the nurses are different. She is extreme in a number of ways: most hours worked, longest shifts worked. We have no idea how the hospital allocated nurses to sectors and patients to sectors. We probably won’t get to know the answer to that, ever. The medical world does not put out its dirty washing for everyone to see.

We wrote a report and gave evidence in person in Ravenna in early 2015. I did not have time to see the wonderful Byzantine mosaics though I was treated to some wonderful meals. I think my department paid for my air ticket. Julia and I worked “*pro deo*”. In our opinion, we totally shredded the statistical work of Tagliaro and Micciolo. The court however did not agree. “The statistical experts for the defence only offered a theoretical discourse while those of the prosecution had scientifically established hard facts”. In retrospect, we should have used stronger language in our report. TM stated that they had definitively proven that *Daniela’s presence caused 90 or so extra deaths*. They stated that this number could definitely not be explained as a chance fluctuation. They stated that, of course, the statistics did not *prove* that she had deliberately *murdered* those patients. We, on the other hand,

had used careful scientific language. One begins to understand how it is that experts like Tagliaro and Micciolo are in such high demand by public prosecutors.

There was also toxicological evidence concerning one of the patients and involving KCl, but we were not involved in that. There was also the “selfie”, there was character evidence. There were allegations of thefts of patients’ personal jewellery. It all added up. Daniela was convicted of just one murder. The statistical evidence provided her motive: she just loved killing people, especially people she didn’t like. No doubt a forensic psychologist also explained how her personality fitted so well to the actions she was alleged to have done.

Rapidly, the public prosecution started another case based largely on the same or similar evidence but now concerning another patient, with whom Daniela had had a shouting match, five years earlier. In fact, this activity was probably triggered by families of other patients starting civil cases against the hospital. It would also clearly be in the interest of the hospital authorities to get new criminal proceedings against Daniela started. However, Daniela’s lawyers appealed against her first conviction. It was successfully overturned. But then the court of cassation overturned the acquittal. Meantime, the second case led to a conviction, then acquittal on appeal, then cassation. All this time Daniela was in jail. Cassations of cassations meant that Daniela had to be tried again, by yet another appeal court, for the two alleged murders. Julia and I and her young colleague Francesco Dotto got to work again, improving our arguments and our graphics and our formulations of our findings.

Statistical modelling of a problem in toxicology

At some point, triggered by some discussions with the defence experts on toxicology and pathology, Julia took a glance at Tagliaro’s quite separate report on the toxicological evidence. This led to a breakthrough, as I will now explain.

Tagliaro knew the post-mortem “vitreous humour” potassium concentration of the last patient, a woman who had died on Daniela’s last day. That death had somehow surprised the hospital doctors, or rather, as it later transpired, it didn’t surprise them at all: they had already for three months been looking at the death rates while Daniela was on duty and essentially building up a dossier against her, just waiting for a suitable “last straw”! More-

over, they already had their minds on KCl, since some had gone missing and then turned up in the wrong place. Finally, Daniela had complained to her colleagues about the really irritating behaviour of that last patient, 73-year-old Rosa Calderoni.

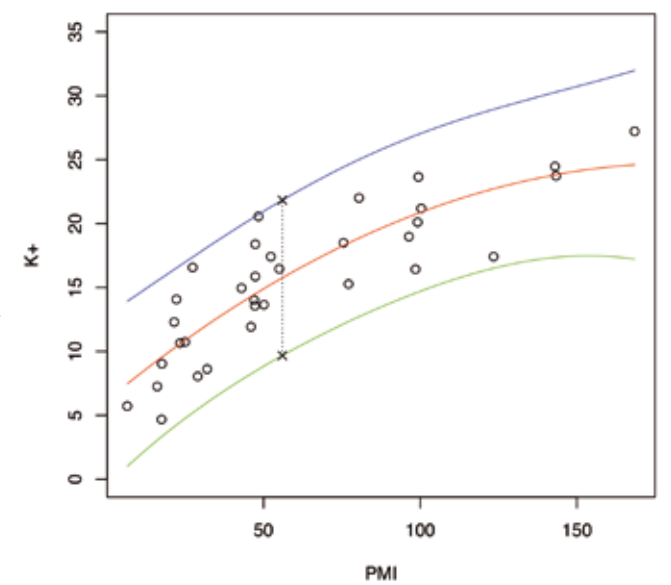
“Vitreous humour” is the transparent, colourless, gelatinous mass which fills your eyeballs. While you are alive it has a relatively low concentration of potassium. After death, cell walls break down, and potassium concentration throughout the body equalises. Tagliaro had published papers in which he studied the hourly rate of increase in the concentration, using measurements on the bodies of persons who had died at a known time of causes unrelated to potassium chloride poisoning. He even had some fresh corpses on which he could make repeated measurements. His motivation was to use this concentration as a tool to determine the PMI (post-mortem interval) in cases when we have a body and a post-mortem examination but no time of death. In (Pigaiani et al. 2020) (without Micciolo’s aid) he did a regression analysis, plotting a straight line through a cloud of points (y = concentration, x = time since death). He had about 60 observations, mostly men, mostly rather young. In a second paper (Palacio et al. 2021), now with Micciolo, he fitted a parabola, and moreover noted that there was an effect of age and of sex. The authors also observed the huge variation around that fitted straight line and concluded that the method was not reliable enough for use in determining the PMI. But this did not deter Tagliaro, when writing his toxicological report on Rosa Calderoni! He knew the potassium concentration at the time of post-mortem, he knew exactly when she died, he had a number for the natural increase per hour after death from his first, linear, regression model. With this, he calculated the concentration at death. Lo and behold: it was a concentration which would have been fatal. He had proved that she had died of potassium chloride poisoning.

Julia and Francesco used the model of the second paper and found out that if you would assume a normal concentration at the time of death, and take account of the variability of the measurements and of the uncertainty in the value of the slope, then the concentration observed at the time of post mortem was maybe above average, but not surprisingly large at all (Dotto, Gill and Mortera, 2022).

Daniela Poggiali became a free woman. I wish her a big compensation and a long and happy life. She’s quite a character.

Aside from the “couleur locale” of an Italian case, this

case had incredibly much similarity with the case of Lucia de Berk. It has many similarities with quite a few other contested serial killer nurse cases, in various countries. According to a *Netflix* series, in which a whole episode is devoted to Daniela, these horrific cases occur all the time. They are studied by criminologists and forensic psychologists who have compiled a list of “red flags” intended to help warn hospital authorities. The scientific term here is “health care serial killer”, or HCSK. One of the HCSK red flags is that you have psychiatric problems. Another is that your colleagues think you are really weird. Especially when your colleagues call you an angel of death, that’s a major red flag. The list goes on. These lists are developed in scientific publications in important mainstream journals, and the results are presented in textbooks used in university criminology teaching programs. Of course, you can only scientifically study convicted HCSKs. Your sources of data are newspaper reports, judges’ summings up, the prosecution’s final summary of the case. It is clear that these red flags are the things that convince judges and jurors to deliver a guilty verdict. These are the features that will first make you a suspect, which police investigators will look for, and which will convince the court and the public of your guilt. Amusingly, one of the side effects of the case of Lucia de Berk was contributing a number of entries to this list, for instance the Stephen



Prediction of vitreous humour K⁺ concentration 56 hours after death without K⁺ poisoning (Dotto, Gill and Mortera, 2022)

King horror murder novels she had at home and which were even alleged to have been stolen from the library. Her conviction for theft of several items still stands. As does Daniela's: this means that Daniela is not eligible for compensation. In neither case was there any real proof of thefts.

Embarrassingly, the case of Lucia de B. case had to be removed from the collections of "known" HCSK cases after 2011, and the criminologists and forensic psychologists also now mention that statistical evidence of many deaths during the shifts of a nurse is not actually a very good red flag. They have learnt something, too. However, their lists still include many disputed cases.

The moral of the case

Interesting is also the incidence of these cases: less than 1 in a million nurses killing multiple patients per year, according to these researchers (Forrest 1995; Yardley and Wilson 2016). These are researchers who have the phenomenon of HCSKs as their life work, giving them opportunities to write lurid books on serial murder, appear in TV panels and TV documentaries explaining the terrible psychology of these modern-day witches, and to take the stand as prosecution witnesses. Now, that "base rate" is actually rather important, even if only known very roughly. It means that such crimes are very, very unusual. In the Netherlands, one might expect a handful of cases per century; maybe on average 100 deaths in a century. There are actually only about 100 murders altogether in the Netherlands per year. On the other hand, more than 1000 deaths every year are due to medical errors. That means that evidence against a nurse suspected of being a HCSK should be very strong indeed before it could convince a rational person that they have a new HCSK on their hands. Lawyers, judges, journalists and the public are unfortunately perhaps not rational persons. They are certainly not good with probability, and not good with Bayes' rule. (It is not allowed to be used in a UK criminal court, because judges have ruled that jurors cannot possibly understand it).

I am still working on one UK case, Ben Geen (Fenton et al. 2021). I believe it is yet another example of a typical innocent HCSK scare in a failing hospital leading to a typical unsafe conviction based largely on the usual red flags and a little bit of bad luck. At least, I see no reason whatsoever to suppose that Ben Geen was guilty of the crimes for which he is sitting out a life sentence. Meanwhile, a

new case is starting up in the UK: *Lucy Letby!* I sincerely hope not to be involved with that one.

Time for a new generation of nosy statisticians to do some hard work. In order to help them, as well as to help lawyers, medical experts, hospital authorities, and the public, the Royal Statistical Society is publishing a 60 page handbook (Green et al., 2022) on how to recognise and avoid the statistical pitfalls which these cases contain.

* This article originally appeared in *Nieuw Archief voor de Wiskunde* and was written for a broad 'mathematical' audience. Obviously, the readers of *STATOR* do know a lot about p-values and Bayes theorem.

REFERENCES

- Dotto, F., Gill, R. D., & Mortera, J. (2022). Statistical Analyses in the case of an Italian nurse accused of murdering patients. *Law, Probability and Risk* (under revision, conditional acceptance). <https://arxiv.org/abs/2202.08895>
- Fenton, N., Gill, R. D., Lagnado, D., & Neil, M. (2021). Statistical Issues in Serial Killer Nurse Cases. <https://arxiv.org/abs/2106.00758>.
- Forrest, A. R. W. (1995). Nurses Who Systematically Harm Their Patients. *Medical Law International*, 1(4), 411–21. <https://doi.org/10.1177/096853329500100404>.
- Gill, R. D., Groeneboom, P., & de Jong, P. (2018). Elementary Statistics on Trial—the Case of Lucia de Berk. *CHANCE*, 31(4), 9–15. <https://doi.org/10.1080/09332480.2018.1549809>.
- Green, P., Gill, R. D., Mackenzie, N., Mortera, J., & Thompson, W. (2022). *Healthcare Serial Killer or Coincidence? Statistical Issues in Investigation of Suspected Medical Misconduct*. Royal Statistical Society, Statistics and the Law section (to appear, September 2022).
- Palacio, C., Rossella Gottardo, R., Cirielli, V., Musile, G., Agard, Y., Bortolotti, F., & Tagliaro, F. (2021). Simultaneous Analysis of Potassium and Ammonium Ions in the Vitreous Humour by Capillary Electrophoresis and Their Integrated Use to Infer the Post Mortem Interval (PMI). *Medicine, Science and the Law*, 61(1_suppl), 96–104.
- Pigaiani, N., Bertaso, A., De Palo, E.F., Bortolotti, F., & Tagliaro, F. (2020). Vitreous Humor Endogenous Compounds Analysis for Post-Mortem Forensic Investigation. *Forensic Science International*, 310, 110235.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Yardley, E., & Wilson, D. (2016). In Search of the 'Angels of Death': Conceptualising the Contemporary Nurse Healthcare Serial Killer. *Journal of Investigative Psychology and Offender Profiling*, 13(1), 39–55. <https://doi.org/https://doi.org/10.1002/jip.1434>.

RICHARD GILL is professor emeritus of Statistics at Leiden University. Apart from other distinctions he is an Honorary Member of the VVSOR.
E-mail: gill@math.leidenuniv.nl



Hij was bescheiden. Hij was aardig. Hij was een groot denker. Hij kon schrijven. En hij wist veel. Hij stierf op vrijdag 1 april 2022. Zijn naam was Gerhard Woeginger, een productieve computerwetenschapper.

Zijn leven en carrière kunnen kort als volgt worden samengevat. Hij werd geboren op 31 mei 1964 in Graz, studeerde aan de Technische Universiteit van Graz en promoveerde in 1991 bij Franz Rendl op een proefschrift getiteld 'Geometric clustering, Reconstruction and Embedding Problems: Combinatorial Properties and Algorithms'. Hij werd in 2001 hoogleraar aan de Universiteit Twente en trad in 2004 in dienst bij de TU/e als hoogleraar Combinatorial Optimization. In 2016 verhuisde hij naar RWTH Aken. Tijdens zijn tijd aan de TU/e begeleidde hij zo'n tien promovendi.

Zijn impact op het werkveld is niet zo gemakkelijk kort te beschrijven. Hij dacht mee op bijna elk gebied binnen de theoretische informatica. Social choice, bibliometrie, algoritmen (vooral online), benaderbaarheid, computationale geometrie en natuurlijk zijn grote passie: computationale complexiteit. Zijn talent om verbanden te zien tussen verschillende problemen was uitmuntend. Zijn vermogen om de essentie te distilleren en het vervolgens op te schrijven op een manier dat het allemaal heel natuurlijk leek, was griezelig. En zijn gedrevenheid en enthousiasme om makkelijk van moeilijk te onderscheiden, werkte aanstekelijk. De afdeling wiskunde en informatica van de TU/e is hem veel verschuldigd - we zijn hem erg dankbaar

voor zijn tijd op de afdeling. Zijn vriendelijkheid gecombineerd met een diepe wiskundige nieuwsgierigheid is een bron van inspiratie geweest voor iedereen om hem heen. Die geest is nog steeds onderdeel van de afdeling.

Hij zat in de programmacommissie van een enorm aantal conferenties, was programmavoorzitter van ESA1997, MAPSP2005, IPCO2011, EURO2009 en zat in het bestuur van een tiental tijdschriften waaronder *OR Letters*. Hij heeft de P-versus-NP-pagina opgezet en onderhouden, een vintage Gerhard-achtige reeks webpagina's die probeert de P=NP-kwestie op te lossen. Om te zeggen dat hij heeft bijgedragen aan de kerstpuzzel (adventskalender) is een understatement, hij zorgde eigenhandig voor het bestaan ervan. En er is nog veel, veel meer te zeggen.

Bovenal kon hij goed luisteren - hij kon nog chocola maken van je meest ongestructureerde woorden. En dan schreef hij een artikel, sneller dan het licht. We zullen zijn aanwezigheid op conferenties, zijn vragen bij presentaties en zijn bescheiden glimlach bij het ontdekken van resultaten missen. Hij kende de oorsprong van de term NP-complete, hij kende artikelen uit vervlogen tijden in obscure tijdschriften; maar hij wist ook hoe een dartbord in elkaar stak en hoe je voetballen moest.

We herinneren Gerhard als een vriendelijke collega met een formidabele drive voor de wetenschap. We zijn geschokt door zijn veel te vroege heengaan en onze gedachten zijn bij degenen die hij achterlaat.

FRITS SPIEKMA



Erelidmaatschap voor prof. dr. Jacqueline Meulman

Tijdens de Annual Meeting van de VVSOR op 17 maart 2022 heeft de Algemene Vergadering unaniem het voorstel van het bestuur aangenomen om Jacqueline Meulman te benoemen tot erelid van de vereniging. Dit vanwege haar grote verdiensten voor de statistische wetenschap én voor de vereniging.

Jacqueline Meulman is een van de meest vooraanstaande Nederlandse statistici. Ze is verbonden aan de Universiteit Leiden en Stanford University en heeft veel bijgedragen aan niet-lineaire multivariate analyse. Ze is lid van de Koninklijke Nederlandse Akademie van Wetenschappen én van de Koninklijke Hollandsche Maatschappij der Wetenschappen. Internationaal heeft ze onder meer een Lifetime Achievement Award van de Psychometric Society ontvangen.

In onze vereniging heeft ze een belangrijke rol gespeeld, vooral tijdens haar periode (2011 – 2017) als voorzitter. De sectie Data Science werd opgericht en ze slaagde er in de samenwerking met onze uitgeverij te vernieuwen en in ons voordeel te verbeteren. Daarnaast was ze erop gespitst de vereniging te verjongen. Door de sectie Young Statisticians op te richten gaf ze hiertoe een belangrijke aanzet. Dit resulteerde in een groot aantal jonge nieuwe leden.

De redactie van *STA&OR* feliciteert Jacqueline van harte met dit erelidmaatschap.

VVSOR Willem R. van Zwet Prijs voor Rianne de Heide

Eveneens tijdens de Annual Meeting 2022 werd aan Rianne de Heide de Willem R. van Zwet Prijs uitgereikt voor haar proefschrift *Bayesian Learning: Challenges, Limitations and Pragmatics*.

Het juryrapport vermeldt 'Dit jaar ontvingen we een klein aantal nominaties van hoge kwaliteit voor de van Zwet-prijs, die een verscheidenheid aan onderwerpen op het gebied van statistiek en operationeel onderzoek bestrijken. We merken, net als in voorgaande jaren, op dat de nominaties voornamelijk afkomstig zijn uit de Mathematische Statistiek en Operations Research. We suggereren promotoren uit de toegepaste statistiek om minder bescheiden te zijn en ook hun excellente dissertaties voor te dragen!'

De keuze voor de winnaar was unaniem: Rianne de Heide. Haar proefschrift behandelt het onderwerp Safe Testing, een methode waarmee het verzamelen van gegevens vroegtijdig kan worden gestopt en optioneel kan worden voortgezet. De jury was onder de indruk van haar boeiende schrijfstijl, die makkelijk van conceptueel naar technisch en van praktisch naar filosofisch gaat. Het was echt een genot om te lezen. Het technische niveau is hoog. De jury is van mening dat het in dit proefschrift beschreven werk van grote invloed zal zijn en feliciteert Rianne de Heide met een welverdiende onderscheiding.