

STAtOR

CAUSALITY & COMPLEXITY

Program Annual Meeting 2020 and the
VVSOR Conference 2020

Gebruik van sensoren om meetfouten in
enquêtes te corrigeren

Consistente schattingen voor categoriale data
gebaseerd op een mix van administratieve
databronnen en enquêtes

De regel van Cardano

From killer nurses to quantum entanglement,
and back; Part 1

Impact van synchromodale logistiek op
achterland transport



STATOR

Jaargang 21, nummer 1, maart 2020

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, Kristiaan Glorie, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stermerdink (eindredacteur), Vanessa Torres van Grinsven en Sanne Willems. Vaste medewerkers: John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, mobiel 06 55886747, j.a.dossantos.gromicho@vu.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Fred van Eeuwijk, db@vvsor.nl; Secretaris: Laura Zwep MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Bestuurscoördinator: Pieter Jongsma MSc, db@vvsor.nl; Webmaster: Mitra Ebrahimpoor MSc, Nikky van Buuren 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); prof. dr. Casper Albers (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Daniel Oberski (Section Data Science); Jonas Haslbeck MSc (Young Statisticians) Sanne Willems MSc (Section Statistics Communication).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Postbus 1058, 3860 BB Nijkerk, telefoon 033 2473408, 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

METEN

Zoals het een tijdschrift van een vereniging voor statistiek betaamt, lijkt toeval een grote rol te spelen in de samenstelling van STATOR. Of misschien lijkt dat alleen maar bij toeval zo.

Hoe dan ook, veel van de artikelen in dit nummer houden verband met meten. Zelfs de columns schenken hier aandacht aan. We kunnen u verzekeren dat we hier niet gericht op hebben gezocht bij het samenstellen van dit nummer, het is écht toeval.

Soms lijkt of het leven draait om meten. We willen steeds vaker van alles weten en daarvoor hebben we gegevens nodig. Die gegevens kunnen uit verschillende bronnen komen en moeten dan wel onderling kloppen. Zowel het artikel van Laura Boeschoten als dat van Jonas Klingwort et al. gaat in op de problemen bij het combineren van gegevens met zulk een verschillende oorsprong. Richard Gill heeft een uitgebreid artikel geschreven dat we in twee achtereenvolgende nummers zullen publiceren. Het is grotendeels gebaseerd op zijn afscheidscollege, de *ex-augural speech* zoals hij het zelf noemt, en legt de nadruk op het éerst kijken naar de juistheid van data alvorens er analyses op los te laten. Ten slotte hebben ook de drie columns in dit nummer min of meer het correct meten als onderwerp.

Naast al dat meten zijn er nog enkele interessante artikelen. Steven Tijms (ja, de zoon van) schrijft over de regel van Cardano, de auteur van een van de oudste geschriften over kansrekening, en Marijke Lijzenga vertelt over optimalisatie van transport.

Dit nummer opent, zoals nummer 1 van iedere jaargang, met het programma van en verdere gegevens over de Annual Meeting van onze vereniging. Het onderwerp, Causality and Complexity, is uiterst actueel en het programma is veelbelovend.

Wij wensen u veel meetbaar leesplezier en hopen veel van onze lezers op 12 maart te ontmoeten.



10



16



20



26



38

INHOUD

4 Redactioneel – Meten



VVSOR Annual Meeting 2020
VVSOR Conference 2020
Causality & Complexity

5 Letter from the President | FRED VAN EEUWIJK

6 Program Annual Meeting & Conference on March 12, 2020

10 Gebruik van sensoren om meetfouten in enquêtes te corrigeren | JONAS KLINGWORT, BART BUELENS, JOEP BURGER & RAINER SCHNELL

15 Weet wat je meet – column | GERRIT STEMERDINK

16 Consistente schattingen voor categoriale data gebaseerd op een mix van administratieve data- bronnen en enquêtes | LAURA BOESCHOTEN

19 International Prize in Statistics

20 De regel van Cardano | STEVEN TIJMS

26 From killer nurses to quantum entanglement, and back; Part 1 | RICHARD D. GILL

35 Buitenspel, meetfouten en millimetergemier – column | GERARD SIERKSMA

38 Impact van synchromodale logistiek op achter- land transport | MARIJKE LIJZENGA

42 De machine stopt – column | JOHN POPPELAARS

43 Good news from Young Statisticians World Statistics Congress 2021

VVSOR Conference 2020

Causality & Complexity

&

Annual Meeting of the Netherlands Society for Statistics and Operations Research (VVSOR)

Thursday March 12, 2020

LOCATION

Gertrudiskapel – In de Driehoek congres- en vergadercentrum
Willemsplantsoen 1 C, 3511 LA Utrecht



Letter from the President

As has become usual over the last years, I direct myself to the VVSOR members in the first *STATOR* issue of the new year. The main reason to do so is to make publicity for the annual meeting of the VVSOR and to inform our members about what the board of the VVSOR has in mind on how to maintain and develop our society.

The central theme for an VVSOR annual meeting should comply with two requirements. Firstly, the theme should be of interest to the statistical and OR community for being a hot research topic. Secondly, the theme should illustrate the impact of statistics and OR on science, technology and society. In that spirit, we had 'Climate Change' as central theme in 2018 and 'Big Data & Privacy' in 2019. For 2020, a special committee within the VVSOR has been created that fully dedicates itself to the organisation of the annual meeting. This committee chose the topic 'Causality and Complexity'.

VVSOR Conference 2020 – Causality & Complexity

The following eminent researchers have agreed to speak on this topic: Marloes Maathuis, Stijn Vansteelandt, Joris Mooij, Rick Quax and Jonas Peters. Their talks look at various aspects of causal modelling and modelling of complex systems. Adjustment for confounders is a major issue in causal modelling. Both Marloes Maathuis and Stijn Vansteelandt will elaborate their vision. Joris Mooij will address a new constraint-based method of causal discovery from multiple data sets. Rick Quax will look at causal interpretations of information flows among stochastic dynamical variables. Jonas Peters will introduce a new technique called Anchor Regression that presents a trade-off between predictive and causal modelling. For more detail on the scientific program, I refer to the abstracts that are published elsewhere in this issue of *STATOR* and to the VVSOR webpages

To me, the scientific program for the VVSOR annual meeting looks very exciting. Beyond the scientific program, the annual meeting also comprises the announcement of the winners of the best MSc thesis (Jan Hemelrijk Award), PhD thesis (Willem R. van Zwet Award) in statistics and operations research. The Hemelrijk and Van Zwet awards are annual awards. This year, we will also present the winner of the five-yearly Van Dantzig Award, an award for

outstanding work during the career of a young statistician or operations research scientist.

Agenda points for Algemene Ledenvergadering

Around noon, we will celebrate our annual members' meeting (Algemene Ledenvergadering). There are some serious issues to be discussed. Last year, the board has created a discussion group to think about the future structure of the VVSOR. Some of the ideas that have turned up for restructuring our society will be discussed with the members. And just as we did last year, we again ask for volunteers to assist the VVSOR board in elaborating new policies and ideas to keep our society viable. The areas in which we need new people remain organisation of the annual meeting, communication, and information on statistics and operations research to journalists and media. The committees on Education and on Certification (re-)started their activities after last year's appeal for members. The role of the Sections in central activities, like organizing the Annual Meeting, is another topic that deserves attention.

Last year I also asked for VVSOR members to reflect on a publication strategy for *Statistica Neerlandica*. This remains an issue that requires further input. Good news is that Miroslav Ristic has joined the chief editorial team, replacing Eric Cator (thanks Eric for all the good work). He runs the journal now together with Marijtje van Duijn and Nan van Geloven. Furthermore, Wiley has presented new data on *Statistica Neerlandica* that show that its impact has increased considerably.

Recently an important discussion took place around the plans on the intended restructuring of the curriculum of basic and secondary schools and the place of Statistics in that curriculum. Several members of our society participated very intensively in that discussion and their resulting recommendations seem to have had a considerable influence on the final plans.

We have a very attractive scientific program for the annual meeting on the 12th of March in Utrecht, therefore we count on your presence to discuss issues related to our society. And of course, the 12th of March will also provide an excellent opportunity to meet your fellow society members. Therefore, I am confident to see many of you in In de Driehoek in Utrecht.

FRED VAN EEUWIJK, President VVSOR

Causality & Complexity

DATE

Thursday, March 12, 2020

VENUE

Gertrudiskapel – In de Driehoek congres- en vergadercentrum, Willemsplantsoen 1C, 3511 LA Utrecht (very close to Utrecht CS).

REGISTRATION

Registration for the conference is mandatory at www.vvsor.nl/articles/vvsor-annual-meeting. Detailed information can be found on our website.

LANGUAGE

The talks at the annual meeting will be in English, the Annual General Meeting (ALV) will be in Dutch.

ALGEMENE LEDENVERGADERING (ALV)

The Annual General Meeting (ALV) takes place on March 12, 12:15 – 13:00. The relevant documents will be provided on the website two weeks before the meeting. You can also get them by e-mail if you send a request to <admin@vvsor.nl>.

LUNCH, COFFEE, TEA AND DRINKS

Lunch, coffee and tea during the breaks and drinks afterwards are offered by the Society.

DINNER

Dinner will take place on March 12, 19:00 – 21:00 at xxxxxxxx

ORGANIZING COMMITTEE

The annual meeting is organized by a special committee in cooperation with the board of the VVSOR. For questions, contact the administration by e-mail at <admin@vvsor.nl>.

REGISTRATION BEFORE MARCH 10!

PROGRAMME

09:15 – 09:45	Registration and coffee/tea
09:45 – 09:50	Welcome by FRED VAN EEUWIJK, President of the VVSOR
09:50 – 10:35	Causality and Covariate Adjustment MARLOES MAATHUIS, Eidgenössische Technische Hochschule (ETH), Zürich
10:35 – 11:20	Machine Learning for the Evaluation of Treatment Effects: Challenges, Solutions and Improvements STIJN VANSTEELENDT, Ghent University & London School of Hygiene and Tropical Medicine
11:20 – 11:40	Break
11:20 – 12:15	Ceremony of the Willem R. van Zwet Award, Jan Hemelrijk Award & Van Dantzig Award Prize winners will be presented by the juries, followed by a short presentation by the laureates
12:15 – 13:00	Annual General Meeting (ALV)
13:00 – 13:45	Lunch
13:45 – 14:30	How to Learn Causal Relations from Data? JORIS MOOIJ, University of Amsterdam
14:30 – 15:15	Defining and Discovering (Causal) Information Flows: Nudge Causality RICK QUAX, University of Amsterdam
15:15 – 15:45	Break
15:45 – 16:30	Causality and Robust Prediction JONAS PETERS, University of Copenhagen
16:30 – 18:00	Snacks and Drinks
19:00 – 21:00	Dinner At The Colour Kitchen (Oudegracht 214, 3511 NS Utrecht)

March 12 | VVSOR Conference 2020

09:50 – 10:35

CAUSALITY AND COVARIATE ADJUSTMENT

Marloes Maathuis

Eidgenössische Technische Hochschule (ETH), Zürich

Covariate adjustment is commonly used in practice to adjust for confounders. In this talk, we argue that the covariate set should be chosen with care. We will discuss graphical criteria that identify covariate sets that give consistent and efficient estimates of total effects in causal linear models. We will illustrate the concepts in examples and discuss various generalizations.

MARLOES MAATHUIS is Professor of Statistics at the Seminar for Statistics at ETH Zürich. She studied Applied Mathematics at Delft University of Technology and obtained a PhD in Statistics from the University of Washington, Seattle. Her research interests include causality, graphical models, high-dimensional models, and applications of statistics. In 2004 she was the recipient of the VVSOR prize for the best MSc thesis. Since 2014 this prize features as the Jan Hemelrijk Award. For more information, please see <https://stat.ethz.ch/~maathuis/>.

10:35 – 11:20

MACHINE LEARNING FOR THE EVALUATION OF TREATMENT EFFECTS: CHALLENGES, SOLUTIONS AND IMPROVEMENTS

Stijn Vansteelandt

Ghent University & London School of Hygiene and Tropical Medicine

The evaluation of treatment effects from observational studies typically requires adjustment for high-dimensional confounding. This is the result of a lack of comparability between treated and untreated subjects in possibly many (pre-treatment) factors that are also related to outcome. While such adjustment is routinely achieved via parametric modelling, it is not entirely satisfactory as model misspecification is likely, and even relatively minor misspecifications over the observed data range may induce large bias in the treatment effect estimate. Over the past 2 decades, there has therefore been growing interest in the use of machine learning methods to assist this task. This is not surprising if one considers the enormous contributions that the machine learning literature has offered on how to predict outcomes based on possibly high-dimensional predictors or features. In this talk, I will therefore focus on the use of machine learning for the evaluation of (causal) treatment effects. This turns out to be a challenging task: while the prediction performance of a given machine learning algorithm can be measured by contrasting observed and predicted outcomes, such evaluation becomes impossible when machine learning is used for treatment effect estimation since the 'true' treatment effect is always unknown. In this talk, I will demonstrate that naive use of existing machine learning algorithms is problematic for treatment evaluation and explain why that is the case. I will next give a gentle introduction to pioneering work on Targeted Learning and on Double Machine Learning, and will discuss improvements that we have made to these techniques. Throughout the talk, 'machine learning' will be considered in the broad sense as any algorithm that uses data to 'learn' a proper model for the data, thus including (though not being limited to) routine variable

March 12 | VVSOR Conference 2020

selection procedures. The talk is based on joint work with Oliver Dukes (Ghent University) and will be accessible to attendees without a detailed understanding of machine learning algorithms.

STIJN VANSTEELANDT is Professor of Statistics at Ghent University (Belgium) and Professor of Statistical Methodology at the London School of Hygiene and Tropical Medicine (UK). As a causal inference expert, he primarily develops methods for causal machine learning, mediation analysis, time-varying confounding control, and for handling intercurrent events in randomised experiments. He has authored over 150 peer-reviewed publications in international journals on a variety of topics in biostatistics, epidemiology and medicine, such as the analysis of longitudinal and clustered data, missing data, mediation and moderation/interaction, instrumental variables, family-based genetic association studies, analysis of outcome-dependent samples, phylogenetic inference, meta-analysis, post-selection inference and interim analysis. He is currently Associate Editor of the Journal of the Royal Statistical Society (Series B) and has previously served as Co-Editor of Biometrics, the leading flagship journal of the International Biometrics Society, and as Associate Editor for the journals Biometrics, Biostatistics, Epidemiology, Epidemiologic Methods and the Journal of Causal Inference.

13:45 – 14:30

HOW TO LEARN CAUSAL RELATIONS FROM DATA?

Joris Mooij
University of Amsterdam

Many questions in science, policy making and everyday life are of a causal nature: how would a change of A affect B? Causal inference, a branch of statistics and machine learning, studies how cause-effect relationships can be discovered from data and how these can be used for making predictions in situations where a system has been perturbed by an external intervention. In this talk, I will introduce the basics of two, apparently quite different, approaches to causal discovery. I will discuss how both approaches can be elegantly combined in Joint Causal Inference (JCI), a novel constraint-based approach to causal discovery from multiple data sets. This approach leads to a significant increase in the accuracy and identifiability of the predicted causal relations. One of the remaining big challenges is how to scale up the current algorithms such that large-scale causal discovery becomes feasible.

JORIS MOOIJ is Associate Professor at the University of Amsterdam, the Netherlands. He studied mathematics and physics and received his PhD degree with honours from the Radboud University Nijmegen, the Netherlands, in 2007 on the topic of approximate inference in graphical models. Afterwards, he was a research scientist at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany. In the following years, he has obtained an NWO VENI grant, an NWO VIDI grant and an ERC Starting grant on several topics in the area of causal inference. The research topics addressed by his group span the entire spectrum from causal modeling, discovery, prediction, validation and application and combine mathematical, algorithmic, statistical and modeling aspects. He has won several awards for his work.

March 12 | VVSOR Conference 2020

14:30 – 15:15

DEFINING AND DISCOVERING (CAUSAL) INFORMATION FLOWS: NUDGE CAUSALITY

Rick Quax
University of Amsterdam

Seemingly very different systems may actually display the same kind of (complex) emergent behaviour. We argue that (Shannon's) information theory – or an appropriately generalized version – may bring a revolution in the form of a unified theoretical framework in complexity science. A crucial concept in this framework is that of 'information flows' among (stochastic) dynamical variables. In this talk I will explore causal interpretations of information flows, a possible way to detect causal information flows, and present an outlook of challenges in the multivariate setting. The work presented is part of an ongoing effort to develop a novel viewpoint on causality based on small interventions.

RICK QUAX is Assistant Professor in the Computational Science Lab (CSL) at the University of Amsterdam. His interests include studying emergent behaviours exhibited by complex adaptive systems such as phase transitions, tipping points, and controllability, using information theory and computational modelling. The application domains are diverse and include financial markets, social systems, ecosystems, and marine and cell biology.

15:45 – 16:30

CAUSALITY AND ROBUST PREDICTION

JONAS PETERS
University of Copenhagen

Purely predictive methods have do not perform well when the test distribution changes too much from the training distribution. Causal models are known to be stable with respect to distributional shifts such as arbitrarily strong interventions on the covariates, but do not perform well when the test distribution differs only mildly from the training distribution. We introduce Anchor Regression, a framework that provides a trade off between causal and predictive models. The method poses different (convex and non-convex) optimization problems and relates to methods that are tailored for instrumental variable settings. If time allows, we show how similar principles can be used for inferring metabolic networks. No prior knowledge about causality is required.

JONAS PETERS is a professor in statistics at the Department of Mathematical Sciences at the University of Copenhagen. Previously, he has been at the MPI for Intelligent Systems in Tübingen and was a Marie Curie fellow at the Seminar for Statistics, ETH Zürich. He studied Mathematics at the University of Heidelberg and the University of Cambridge. In his research, Jonas is interested in inferring causal relationships from different types of data and in building statistical methods that are robust with respect to distributional shifts. He seeks to combine theory, methodology, and applications. For more information, please see <http://web.math.ku.dk/~peters/index.html>



Vrachtwagens op de A28, bij knooppunt Hattemerbroek. Foto: BakkeBaarend CC

GEBRUIK VAN SENSOREN OM MEETFOUTEN IN ENQUÊTES TE CORRIGEREN

Het CBS publiceert cijfers over het goederenvervoer over de weg. Deze cijfers geven inzicht in het gebruik van het Nederlandse wegennet en de doorvoer- en distributiefunctie van het Nederlandse wegtransport. Ook worden de resultaten gebruikt voor het bepalen van beleid op het gebied van milieu en verkeersveiligheid. Om het goederenvervoer over de weg in kaart te brengen voert het CBS een enquête uit onder de eigenaren van de Nederlandse bedrijfsvoertuigen. Door de uitkomsten van de enquêtes en administratieve gegevens te koppelen aan de data die een sensorsysteem van Rijkswaterstaat verzamelt, kan met vangst-hervangst-technieken in potentie onderrapportage worden gemeten.

JONAS KLINGWORT, BART BUELENS, JOEP BURGER, RAINER SCHNELL

Er is steeds meer belangstelling voor het gebruik van sensordata in het maken van officiële statistieken. Bij officiële statistieken wordt veel gebruik gemaakt van kanssteekproeven omdat daarmee efficiënt zuivere schattingen over de hele populatie gemaakt kunnen worden. Via een a-selecte steekproef worden mensen verzocht een vragenlijst in te vullen. Deze methode heeft wat nadelen: is potentieel belastend voor de respondent, niet iedereen doet mee en antwoorden zijn gevoelig voor fouten – al dan niet bewust gemaakt.

Sensoren kunnen de antwoorden op sommige vragen direct meten zonder tussenkomst van een respondent. Vaak hebben sensoren dan weer als nadeel dat niet elk element in de populatie een bekende en positieve kans heeft om gemeten te worden. In dit onderzoek zijn sen-

sordata geïntegreerd met enquête- en registerdata om met vangst-hervangst (VH) technieken de kwaliteit van de populatieschattingen te verbeteren. In de hier beschreven toepassing wordt gebruik gemaakt van de Enquête Wegvervoer, de bedrijfs- en kentekenregisters en Weigh-in-Motionwegsensoren. Dit onderzoek is de eerste keer dat het CBS sensorgegevens op microniveau koppelt aan enquêtes en administratieve gegevens om mogelijke vertekening door onderrapportage in dagboekenquêtes te beoordelen.

Data

De Enquête Wegvervoer is een dagboekonderzoek naar het goederenvervoer over de weg. De gepubliceerde cij-

fers geven inzicht in het gebruik van het wegennet en de functie van het wegtransport. Verder worden ze gebruikt voor het bepalen van beleid op het gebied van milieu en verkeersveiligheid. Het is een verplichte enquête met een hoog responspercentage van ongeveer 90% op een steekproef van ongeveer 34 duizend kentekens. De bezitter van de vrachtwagen wordt verzocht alle ritten en zendingen per dag gedurende een toegewezen week op te geven. Onderrapportage door onterechte opgave dat de vrachtwagen niet is gebruikt zal leiden tot onderschatting van het goederenvervoer over de weg.

De sensordata worden verzameld door het Weigh-in-Motionsensorsysteem van Rijkswaterstaat om overbelading van vrachtwagens op te sporen. In 2015 werd op 18 locaties (figuur 1) het gewicht van elke pas-

serende vrachtwagen gemeten en werd het type van de vrachtwagen geïdentificeerd, resulterend in ongeveer 36 miljoen waarnemingen. Daarnaast is een foto van de kentekenplaat van de voor- en achterkant gemaakt, die een directe koppeling met de registers mogelijk maakt (figuur 2). Hierdoor kan het lege vrachtwagengewicht van de sensormeting worden afgetrokken. Dit verschil resulteert in het vervoersgewicht, dat gelijk is aan de definitie van het gevraagde gewicht in de enquête. De mogelijkheid om de sensorwaarnemingen op microniveau te koppelen aan de enquête maakt toepassing van VH-methoden mogelijk.

De bedrijfs- en kentekenregisters leveren aanvullende gegevens op met informatie over de technische kenmerken van de vrachtwagens en de specifieke ken-



Figuur 1. Locatie van de wegsensoren van het Weigh-in-Motionsensorsysteem; elk van de negen systemen heeft twee stations, één in iedere richting, aangeduid met een plus en een cirkel



Figuur 2. Weigh-in-Motionsoftware met onder andere het gewicht per as (het kenteken is om privacyredenen vervaagd)

		SENSOREN		TOTAAL
		gedetecteerd	niet gedetecteerd	
ENQUÊTE	gerapporteerd	Hervangst	Alleen gerapporteerd	Eerste vangst
	niet gerapporteerd	Alleen gedetecteerd	?	?
TOTAAL		Tweede vangst	?	?

Tabel 1. Combinatie van de rapportages van enquêtes en van sensoren

merken van de vrachtwagenbezitters. Deze achtergrondkenmerken worden gebruikt in de VH-modellen. Fouten in de sensormetingen van de asgewichten, bijvoorbeeld door oneffenheden in de weg, kunnen leiden tot vertekening van het geschatte vervoersgewicht. Om dit risico te beperken zijn extreme waarden geïmputeerd met een conditioneel gemiddelde gegeven enkele achtergrondkenmerken.

Methode

Vangst-hervangst-technieken zijn ontwikkeld in de ecologie om de (onbekende) omvang van dierenpopulaties te schatten. In de hier beschreven toepassing is de steekproefenquête de eerste vangst en de sensorwaarneming de tweede vangst. VH-technieken corrigeren voor zowel nonrespons als onderrapportage in de enquête. Indien de selectieve nonrespons in de enquête volledig wordt verklaard door het weegmodel, kan het relatieve verschil tussen de gewogen enquêteschatting en een VH-schatting duiden op onderrapportage in de enquête.

Voor de toepassing van dergelijke technieken zijn de aantallen in tabel 1 nodig, die zijn af te leiden door

de koppeling van beide bronnen op kenteken en dag. Er zijn elementen (kenteken-dagcombinaties) die alleen in de enquête gerapporteerd worden, elementen die alleen door de sensoren gedetecteerd worden, en elementen in beide bronnen: de hervangst. Elementen die noch in de enquête worden gerapporteerd, noch door de sensoren worden gedetecteerd, worden geschat met behulp van VH-technieken. Op basis van de combinatie van beide bronnen kan dan een betere populatieschatting gemaakt worden dan op basis van een enkele bron, mits aan een aantal voorwaarden wordt voldaan (zie volgende paragraaf).

De vangkans in de enquête en de sensoren worden gemodelleerd met logistische regressie en loglineaire modellen met behulp van covariaten uit de twee registers. Als de covariaten de heterogeniteit in vangkans verklaren, maakt het niet uit hoeveel en waar de sensoren geïnstalleerd zijn. Er worden vijf schatters toegepast voor twee doelvariabelen: vrachtwagendagen (D) en vervoerd gewicht (W). Een vrachtwagendag (D) wordt gedefinieerd als een dag dat een vrachtwagen in Nederland op de weg is geweest.

De eerste schatter (ENQ) is gebaseerd op de antwoorden in de enquête gecorrigeerd voor selectieve

nonrespons met behulp van poststratificatie-weging. Deze schatter wordt vergeleken met de VH-schatters. Aangezien VH-schattingen voor zowel de selectieve nonrespons als meetfouten corrigeren, kan het verschil tussen VH-schattingen en ENQ worden toegeschreven aan meetfouten, dat wil zeggen onderrapportage in de enquête. Voor een naïeve uitbreiding van de huidige schatter worden de waarnemingen van de enquête aangevuld met de sensorwaarnemingen (ENQX). De Huggins-schatter (HUG) modelleert de vangkans met een logistisch regressiemodel als functie van achtergrondkenmerken, gegeven dat een vrachtwagen door ten minste één bron is waargenomen. De Lincoln-Petersen-schatter (LP) gebruikt alleen de hoeveelheden uit bovenstaande tabel. De eerste vangst wordt opgehoogd met de verhouding tussen de tweede vangst en de hervangst. De loglineaire schatter (LL) modelleert de hoeveelheden in bovenstaande tabel, uitgesplitst naar achtergrondkenmerken, met een Poisson-regressiemodel. Een stapsgewijze selectieprocedure (gebaseerd op BIC) wordt toegepast om de balans te vinden tussen modelfit en modeleenvoud (parsimonie). De steekproefvariantie wordt voor alle schatters geschat door middel van bootstrapping.

Aannames

De VH-schatting is gebaseerd op de volgende vijf aannames: onafhankelijke datasets, gesloten populatie, homogene vangkans, alle elementen behoren tot de doelpopulatie en perfecte koppeling van datasets. Het voldoen aan de vijf VH-aannames is een vereiste om onvertekende VH-schattingen te garanderen. Aan de eerste aanname (onafhankelijkheid) wordt voldaan. De tweede aanname (gesloten populatie) wordt niet ernstig geschonden, aangezien het eerste en laatste kwartaal van het voertuigregister meer dan 99% overlappen. Aan de derde aanname (homogene vangkans) wordt voldaan zolang de heterogeniteit in vangkans verklaard wordt door de achtergrondkenmerken in de modellen. Dezelfde aanname wordt gemaakt bij het corrigeren van de enquêterespons voor selectieve nonrespons. Aan de vierde aanname (elementen behoren tot de doelpopulatie) wordt voldaan omdat de enquêtesteekproef uit het populatieregister wordt getrokken en in de sensordata alleen wordt gezocht naar voertuigen in de steekproef.

De vijfde aanname (perfecte koppeling) kan soms worden geschonden in beide gegevensbronnen om verschillende redenen. Ten eerste kan het zijn dat voer-

tuigeigenaren te weinig of te veel dagen van de week rapporteren, of dat zij de verkeerde dagen van de week rapporteren. Ten tweede moeten de voertuigeigenaren de dag van laden melden, welke kan verschillen van de dag van rijden zoals gedetecteerd door de sensoren. Ten derde registreren de sensoren ritten waar niet over gerapporteerd hoeft te worden, zoals lege ritten en ritten voor onderhoud of tanken. Ten vierde kan de sensor een kentekenplaat soms niet herkennen. Om de gevoeligheid van de resultaten voor deze respons- en sensorfouten te beoordelen zijn simulatiestudies uitgevoerd.

Resultaten en discussie

Alle VH-schattingen leveren grotere schattingen op voor het aantal vrachtwagendagen en het vervoersgewicht dan de schattingen op basis van de enquête. We raden het gebruik van de loglineaire schatter (LL) aan, omdat deze rekening houdt met de heterogeniteit in vangkansen (in tegenstelling tot LP) en is gebaseerd op het gekoppelde bestand, onvoorwaardelijk op waarneming in ten minste één bron (in tegenstelling HUG).

De simulatiestudies met betrekking tot de vijfde aanname (perfecte koppeling) toonden het volgende aan. In de eerste plaats is LL robuust voor overrapportage in de enquête, terwijl ENQ zou afnemen, waardoor de geschatte onderrapportage zou toenemen. Ten tweede zou LL toenemen met onderrapportage in de enquête, maar langzamer dan ENQ, waardoor de geschatte onderrapportage zou afnemen. Maar zelfs als elke gerapporteerde lading over de volgende dagen zou worden verdeeld, blijft ENQ lager dan LL. De geschatte onderrapportage kan dus niet volledig worden verklaard door een verschil tussen de dag van laden zoals gerapporteerd in de enquête en de dag van rijden zoals gedetecteerd door de sensor. Ten derde verdwijnt de geschatte onderschatting geleidelijk naarmate de fractie foutpositieven in de sensor toeneemt. Het is een open vraag hoe waarschijnlijk het is dat een vrachtwagen in de steekproef die door een sensor wordt gedetecteerd, niet gerapporteerd hoeft te worden. Ten vierde is de zuiverheid van de LL-schattingen robuust tegen het niet herkennen van het kenteken door de sensorsoftware. Alleen de precisie zal afnemen naarmate minder kentekens gekoppeld kunnen worden.

Toekomstig onderzoek zou de hypothese kunnen testen dat het relatieve verschil tussen de enquêteschatting en de VH-schatting prominenter is in handmatige

responsmodi (internet- of papieren vragenlijsten) dan in geautomatiseerde responsmodi (XML). Het is ook aan te raden om te onderzoeken of een deel van het verschil alsnog door nonrespons verklaard kan worden met een beter weegmodel. Verder is het van belang na te denken over het behandelen van de antwoordcategorie ‘niet in bezit’ als kaderfout of (technische) nonrespons. Tot slot is er behoefte aan een methode om de kans op foutposities te kunnen schatten.

De hier gepresenteerde methode is van toepassing op elke validatiestudie waarbij enquête-, register en sensor-data (of een andere externe bigdatabron) op microniveau kunnen worden gekoppeld. Het voldoen aan de vijf VH-aannames garandeert onvertekende VH-schattingen. Simulatiestudies kunnen inzicht geven in de gevoeligheid van de resultaten voor het schenden van een of meerdere aannames.

LITERATUUR

Klingwort, J., Buelens, B., & Schnell, R. (2019). Capture-recapture techniques for Transport Survey estimate adjustment using permanently installed highway sensors. *Social Science Computer Review*, OnlineFirst, doi: 10.1177/0894439319874684.

JONAS KLINGWORT is wetenschappelijk medewerker en PhD-kandidaat in de Onderzoeksgroep Methodologie van prof. dr. Rainer Schnell aan de Universiteit van Duisburg–Essen. Daarnaast is hij statistisch onderzoeker bij het *Center for Big Data Statistics* van het Centraal Bureau voor de Statistiek (CBS).

E-mail: jonas.klingwort@uni-due.de

BART BUELENS is senior data scientist bij de Vlaamse Instelling voor Technologisch Onderzoek (VITO). Eerder was hij senior methodoloog bij het CBS.

E-mail: bart.buelens@vito.be

JOEP BURGER is methodoloog bij het CBS. Eerder was hij evolutionair ecooloog, evolutionair geneticus en entomoloog.

E-mail: j.burger@cbs.nl

RAINER SCHNELL heeft de leerstoel Empirisch Sociaalwetenschappelijk Onderzoek aan de Universiteit van Duisburg–Essen. Eerder was hij directeur van het Centre for Comparative Surveys van de City, Universiteit van Londen en de oprichter-redacteur van het ESRA wetenschappelijk tijdschrift *Survey Research Methods* tijdens de eerste zeven jaar. Sinds 2017 is hij adjunct-professor aan de Faculteit Gezondheidswetenschappen van de Universiteit van Curtin, Perth.

E-mail: sekretariat.schnell@uni-due.de

Foto: Naj



Er wordt wat afgemeten dezer dagen. Zo ongeveer ieder bedrijf dat zichzelf serieus neemt overspoelt zijn klanten met vragen over hun ervaringen. Zelfs mijn apotheek heeft mij de afgelopen maand drie maal een vragenlijstje gestuurd over mijn tevredenheid, zo'n mailtje wordt kennelijk automatisch verstuurd na ieder bezoek. Maar wat is de waarde van al deze gegevens, afgezien van mijn twijfel of iemand er überhaupt wel iets mee doet?

Mochten mijn kinderen deze column onder ogen krijgen dan zouden ze zuchten ‘daar heb je papa weer met zijn eeuwige obsessie’. Maar ik blijf volhouden, er worden veel te vaak gegevens verzameld die nutteloos zijn of op zijn minst minder geschikt voor het doel dat men er mee beoogde.

Een veel gemaakte fout is dat men niet de juiste doelgroep benadert. Neem bijvoorbeeld een bakker die een maand lang zijn klanten een vragenlijstje over hun tevredenheid meegEEft. Heeft zo'n bakker iets aan die antwoorden? Verspilde moeite en geld, ik kan hem zonder vragenlijst al direct zeggen dat de meerderheid van zijn klanten tevreden tot zeer tevreden zal zijn. De niet-tevreden klanten zullen immers allang geen klant meer zijn! Wat die bakker *eigenlijk* wil is zijn omzet vergroten. Dat is prima, maar dan moet hij andere vragen stellen aan andere mensen. Om meer klanten te trekken moet hij juist diegenen die géén klant zijn benaderen met vragen over de redenen daarvoor. Pas dan kan hij de juiste actie ondernemen om hen zijn winkel in te krijgen. En als hij tóch zijn bestaande klanten wil ondervragen: vraag dan niet naar hun tevredenheid, maar naar wat ze missen in zijn assortiment waardoor hij meer producten aan hen kan slijten en op die manier zijn omzet kan vergroten.

Mijn favoriete slachtoffer voor dit soort boutades is de NS. Regelmatig publiceren zij vol trots de laatste

punctualiteitscijfers zoals: afgelopen maand reed 90,8 % van de treinen op tijd. Nu weet ik heus wel dat achter dit simpele getal veel meer schuilt, maar alleen al het feit dat een trein die is uitgevallen niet meetelt kan mij doen stijgen. Maar dan nog zegt dit cijfer mij weinig, wordt een boemeltje met enkele tientallen passagiers net zo zwaar meegeteld als een intercity met vele honderden reizigers? En wat is op tijd: op tijd vertrokken of op tijd aangekomen, en worden vertragingen op tussenliggende overstapstations ook geteld? Het lijkt soms wel of NS denkt dat hun taak bestaat uit het laten rijden van treinen, maar dat is niet zo. Als ze dat zo graag willen gaan ze maar op hun zolder zitten met een Märklinbaan. Hun kernopdracht is namelijk het punctueel vervoeren van *reizigers*. Die moet je meten om te kijken of ze op tijd aankomen, niet de treinen. En kijk daarbij naar het totaaltraject, een vertraging van 3 of 4 minuten halverwege kan leiden tot een vertraging van minimaal 30 minuten omdat men een overstap mist. Kijk ook niet alleen naar de vertraging die de reiziger heeft, maar vooral naar de beleving daarvan. Iemand die slechts enkele keren per jaar een dagje naar familie reist zal een vertraging ongetwijfeld anders ervaren dan een dagelijkse forens. Al was het alleen maar omdat de meeste vertragingen in de spits voorkomen en forensen helaas meestal tot de spits veroordeeld zijn en vaker reizen. Ik heb ooit de voorzichtige schatting gemaakt dat als 90% van de treinen op tijd rijdt een dagelijkse spitsforens de kans loopt op twee tot drie vertragingen per week!

Metten is fantastisch, ik heb het veel gedaan en rekenen met de resultaten kan heel bevredigend zijn. Maar doe het wel met de juiste gegevens, anders is er weinig verschil met het analyseren van random gegenereerde data.

GERRIT STEMERDINK is eindredacteur van *STAtOR*.
E-mail: gjsterdink@hotmail.com



Foto: Andrey Popov

Consistente schattingen voor categoriale data gebaseerd op een mix van administratieve databronnen en enquêtes

Officiële statistieken moeten aan allerlei eisen voldoen en van zo hoog mogelijke kwaliteit zijn. Zo moeten meetfouten zo veel mogelijk worden opgespoord en verbeterd. Ook moeten de statistieken consistent zijn. In Boeschoten (2019) wordt een methode geïntroduceerd die de kwaliteitseisen simultaan aanpakt voor categoriale data. Boeschoten introduceerde in haar proefschrift een methode die de drie hierboven besproken problemen simultaan aanpakt specifiek voor categoriale data. Ze doet dit door verschillende bestaande methoden op een nieuwe manier met elkaar te combineren: enerzijds de *latenteklassanalyse* en anderzijds *multipele imputatie*.

LAURA BOESCHOTEN

Bij het produceren van officiële statistieken maakt het Centraal Bureau voor de Statistiek (CBS) zoveel mogelijk gebruik van bestaande administratieve bronnen. Soms omvatten die de gehele onderzoekspopulatie. Zo bevat de Basisregistratie Personen (BRP) alle inwoners van Nederland. Soms slechts een deel van de populatie. Zo heeft de Dienst Uit-

voering Onderwijs (DUO) geen gegevens over diploma's behaald in het buitenland of aan particuliere instellingen, en is het ook voor personen die langer geleden een opleiding hebben gevolgd minder compleet. Als er interesse is in een statistiek over een onderwerp dat helemaal niet in administratieve bronnen wordt gemeten, dan wordt de informatie

verkregen door uit de populatie een steekproef te nemen en een enquête af te nemen. Over het algemeen bevatten al deze datasets unieke *identifiers* waardoor deze bronnen op persoonsniveau aan elkaar kunnen worden gekoppeld.

Meetfouten

Data verkregen door middel van enquêtes kunnen allerlei soorten fouten kunnen bevatten. Van een fout spreken we wanneer de echte score van een eenheid (meestal een persoon of een bedrijf) op een bepaalde variabele niet overeenkomt met de score die is ingevuld in de enquête. In mijn proefschrift bespreek ik alleen meetfouten in het algemeen en ga ik niet verder in op het onderscheid tussen de verschillende soorten fouten zoals gedefinieerd door Groves et al. (2009). Minder bekend is dat administratieve data evengoed meetfouten kunnen bevatten. Bijvoorbeeld doordat iemand fouten maakt tijdens het invullen van het register, wanneer iemand vergeet de informatie in het register te updaten of wanneer iemand fouten maakt bij het ophalen van de informatie uit het register.

Ongeacht met wat voor type fouten we te maken hebben en waar in de data ze precies zitten, kunnen fouten altijd problematisch zijn en zorgen voor vertekening bij het uitvoeren van statistische analyses wanneer er niets aan wordt gedaan. Soms zijn fouten willekeurig. Wanneer personen bijvoorbeeld hun gewicht in moeten vullen in een enquête en dat niet precies weten, zal de ene persoon een hoger gewicht invullen en de andere persoon een lager gewicht, en het verschil tussen het *echte* gewicht en het *ingevulde* gewicht is de ene keer groter en de andere keer lager. Al deze verschillen zorgen bij elkaar dan vooral voor extra variantie (oftewel spreiding) en wanneer we gewicht in relatie tot een andere variabele onderzoeken, zoals bijvoorbeeld geslacht, dan zal de geschatte relatie als gevolg van deze extra spreiding door meetfout vooral minder sterk zijn. Wanneer fouten structureel dezelfde kant op wijzen, bijvoorbeeld wanneer verschillende personen altijd een lager gewicht invullen dan ze daadwerkelijk hebben, dan kunnen de effecten in sommige gevallen zelfs sterker worden. Behalve deze problematische *vertekening*, zorgen meetfouten bij het CBS nog voor twee andere problemen. Het eerste probleem heeft te maken met *onmogelijke combinaties van scores* op verschillende

variabelen, en is eerder besproken door Sander Scholtus in zijn artikel in *STATOR* (Scholtus, 2018). Het tweede probleem gaat over dat de verschillende statistieken die het CBS produceert (in de praktijk gaat dit vaak over grote hoeveelheden kruistabellen) allemaal *consistent* moeten zijn. Dit wil zeggen dat wanneer een kruistabel tussen de variabelen 'opleidingsniveau × geslacht × regio' wordt geproduceerd, en daarnaast ook een kruistabel 'opleidingsniveau × geslacht × burgerlijke staat', dat, bijvoorbeeld, het totaal aantal hoogopgeleide mannen in beide kruistabellen exact gelijk aan elkaar moet zijn. Dit probleem is eerder en uitgebreider besproken in *STATOR* door Jacco Daalmans (Daalmans, 2019).

In mijn proefschrift introduceer ik een methode die de drie hierboven besproken problemen simultaan aanpakt specifiek voor categoriale data. Ik doe dit door verschillende bestaande methoden op een nieuwe manier met elkaar te combineren: enerzijds de *latenteklassanalyse* en anderzijds *multipele imputatie*.

Latenteklassemodellen

Ten eerste maak ik gebruik van zogenaamde latenteklassemodellen. Latenteklassemodellen worden normaal gebruikt om (categoriale) variabelen te meten die je niet goed direct kunt meten (McCutcheon, 1987). Je kunt hier bijvoorbeeld denken aan persoonlijkheidstypen of type eetstoornissen. In dat laatste voorbeeld wordt er niet direct aan personen gevraagd 'welk type eetstoornis heb je?'. In plaats daarvan wordt er gevraagd naar een aantal zaken die indicatoren kunnen zijn voor bepaalde typen eetstoornissen, zoals overgewicht, overgeven of overmatig sporten. Een bepaalde combinatie van antwoorden op deze vragen kan een indicatie zijn voor een bepaald type eetstoornis. Bij het schatten van het model worden de relaties tussen de indicatoren en de latente variabelen bepaald.

Nadat zo'n model is geschat, geeft het ons ten eerste informatie over wat de kans is dat iemand een bepaalde score heeft op een indicator-variabele gegeven de latente klasse (bijvoorbeeld de kans dat iemand overgeeft gegeven dat de eetstoornis anorexia is), de *conditionele kansen*. Ten tweede geeft het ons informatie over wat de kans is dat iemand tot een bepaalde latente klasse behoort gegeven de scores op de indicatoren (bijvoorbeeld de kans dat iemand anorexia heeft gegeven dat hij geen overgewicht

heeft, overgeeft maar niet sport), de *posterior kansen*.

In mijn onderzoek, gebruik ik de latenteklassemodelen op een andere manier. Ik gebruik namelijk variabelen die hetzelfde meten, maar afkomstig zijn van de verschillende bronnen die we op eenheidsniveau aan elkaar hebben kunnen koppelen (Biemer, 2011). De variabelen kunnen dus zowel afkomstig zijn van administratieve bronnen als van enquêtes en kunnen zowel de gehele populatie als een deel van de populatie bevatten. Het zou bijvoorbeeld kunnen dat we drie variabelen hebben die *burgerlijke staat* meten, waarvan er twee afkomstig zijn uit administratieve bronnen en een afkomstig is uit een enquête. Deze variabelen worden gebruikt als indicatoren van een latente variabele die de *echte* burgerlijke staat meet. Om deze methode te kunnen toepassen moet er wel aan een aantal voorwaarden worden voldaan. In deze specifieke situatie is vooral een belangrijke voorwaarde dat de indicatoren onafhankelijk van elkaar gemeten moeten zijn. Wanneer een administratieve bron ook *burgerlijke staat* meet, maar deze informatie heeft verkregen uit het BRP, dan zijn deze metingen niet onafhankelijk en dus ook niet bruikbaar voor deze methode.

In het latenteklassemodel kunnen ook andere variabelen (covariaten) worden meegenomen. Dit is belangrijk, want wanneer het CBS bijvoorbeeld de kruistabel '*burgerlijke staat* × *geslacht*' moet produceren, dan moet de variabele *geslacht* worden meegenomen in het latenteklassemodel dat de *echte* scores van *burgerlijke staat* meet, omdat anders de uitkomsten van deze kruistabel vertekening kunnen bevatten. Daarnaast kunnen we in het latenteklassemodel ook restricties specificeren. Zo kunnen we ervoor zorgen dat bepaalde combinaties van scores die in de praktijk onmogelijk zijn, ook niet voor kunnen komen nadat we een specifieke score voor *burgerlijke staat* hebben toegewezen na de toepassing van het latenteklassemodel. Bijvoorbeeld, de combinatie '*burgerlijke staat* = gehuwd' en '*leeftijd*=jonger dan 5 jaar' is bij wet verboden en daarom niet mogelijk (als gevolg van meetfouten is het in de praktijk natuurlijk wel mogelijk dat deze combinatie van scores wordt geobserveerd in de data). Het zou dan ook onwenselijk zijn dat deze combinatie na de toepassing van het latenteklassemodel wel voor zou komen. Door middel van restricties op het model kunnen we daar voor zorgen.

Meerdere imputatie

Een manier om onderling consistente tabellen te krijgen is door de kansen uit het latenteklassemodel te gebruik-

ken om per eenheid een score toe te wijzen, oftewel een imputatie van de latente variabele *echte burgerlijke staat* te genereren. Daarnaast willen we de geproduceerde statistieken van variantieschattingen voorzien. De statistieken worden geproduceerd door gebruik te maken van de geïmputeerde data, maar het is wel wenselijk dat de variantieschattingen de onzekerheid bevat die wordt veroorzaakt door de ontbrekende en conflicterende waarden in de oorspronkelijke databronnen. Om dit mee te kunnen nemen maken we gebruik van meerdere imputatie (Rubin, 1987), een methode die doorgaans wordt gebruikt voor ontbrekende waarden in variabelen.

Toepassingen en uitbreidingen

In mijn proefschrift wordt de hierboven beschreven methode geïntroduceerd en op verschillende manieren uitgebreid. Zo wordt het aantal ernstige verkeersgewonden per voertuigtype geschat op basis van data afkomstig van politie en ziekenhuizen en wordt het aantal werkenden op meerdere tijdstippen geschat op basis van een gecombineerde dataset uit Italië. Daarnaast wordt onderzocht of de methode gebruikt kan worden om consistente volkstellingstabellen te genereren en wordt de methode uitgebreid zodat covariaten ook op een later tijdstip aan het model toegevoegd kunnen worden.

LITERATUUR

- Boeschoten, L. (2019). *Consistent estimates for categorical data based on a mix of administrative data sources and surveys*. Doctoral thesis. Tilburg University, Tilburg. Retrieved from: https://pure.uvt.nl/ws/portalfiles/portal/31415117/Boeschoten_Consistent_25_10_2019.pdf.
- Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*. Second Edition. Hoboken: Wiley.
- McCutcheon, A. L. (1987). *Latent Class Analysis*. Sage, Newbury Park.
- Biemer, P. P. (2011). *Latent Class Analysis of Survey Error*. Hoboken: Wiley.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Scholtus, S. (2018). Automatisch opsporen van fouten in data voor officiële statistiek. *STATOR*, 19(3), 14–18
- Daalmans, J. (2019). Automatisch corrigeren van inconsistenties in de officiële statistiek. *STATOR*, 20(3), 9–12

LAURA BOESCHOTEN werkt als postdoctoraal onderzoeker aan de Universiteit Utrecht. In oktober 2019 is zij gepromoveerd aan de Universiteit van Tilburg op het proefschrift dat in dit artikel wordt beschreven.
E-mail: l.boeschoten@uu.nl



International Prize in Statistics

De wetenschap kent veel prijzen voor excellent werk. De bekendste en meest prestigieuze zijn ongetwijfeld de Nobel-prijzen. Wiskundigen kennen dan verder de Fields Medal en de Abel Prize en de informatici hebben de Turing Award. In Nederland hebben we als een van de belangrijkste wetenschappelijke prijzen de Spinozapremie, enkele jaren geleden is die aan Aad van der Vaart toegekend. Op de website van de Koninklijke Nederlandse Akademie van Wetenschappen is verder een hele lijst van prijzen te vinden, zoals bijvoorbeeld de Heineken Prijzen en de Lorentzmedaille.

Ook onze eigen VVSOR kent prijzen toe, de Jan Hemelrijk en Willem van Zwet Awards en als belangrijkste de vijfjaarlijkse Van Dantzig Prijs die dit jaar weer aan de beurt is toegekend te worden.

De internationale statistische gemeenschap stelde enkele jaren geleden vast dat ook op hun vakgebied behoefte was aan een prijs voor zeer excellente wetenschappers. Er kwam een samenwerkingsverband tot stand tussen het International Statistical Institute, de American Statistical Association, het Institute of Mathematical Statistics, de International Biometric Society en de Royal Statistical Society. Samen hebben zij een stichting opgericht die tweejaarlijks de International Prize in Statistics toekent. De uitreiking vindt plaats tijdens het tweejaarlijks World Statistics Congress van het ISI. Aan de prijs is een geldbedrag van US\$ 75.000 verbonden.

Op de website van de organisatie staat de volgende informatie over deze prijs:

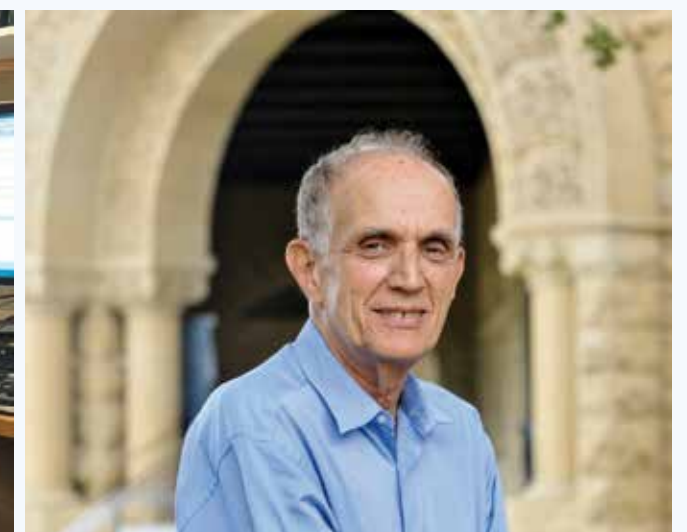
- The prize will be awarded for a single work or body of work, rather than for more diffuse reasons such as “lifetime achievement.” Not only should powerful and original ideas be recognized by the prize, but also contributions that lead to breakthroughs in other disciplines or works with important practical effects on the world.
- Generally, the prize will be awarded to individuals, but in some cases, groups of individuals working on similar ideas — or even teams of individuals or organizations — could be recognized.

Na de eerste toekenningen in 2017 aan Sir David R. Cox in 2019 aan Bradley Efron is het nu de beurt aan de prijs voor 2021, die tijdens het WSC-2021 in Den Haag zal worden uitgereikt. Het nominatieproces is onlangs gestart. Iedere statisticus kan een voordracht indienen, alle informatie hierover is te vinden op de genoemde website.

Voordrachten voor de International Prize in Statistics 2021 kunnen tot 15 augustus 2020 via de website worden aangemeld. Zie voor meer informatie <https://statprize.org/>



Sir David R. Cox



Bradley Efron

De regel van Cardano

In het Italië van de zestiende eeuw leefde één van de kleurrijkste geleerden uit de geschiedenis, Girolamo Cardano. Nu is hij bijna vergeten, alleen de zogeheten cardanische ophanging die naar hem is vernoemd maar niet door hem is uitgevonden, herinnert nog aan hem.

STEVEN TIJMS

Girolamo Cardano (1501–1576) was tijdens zijn leven een internationale beroemdheid. De erudiete geleerde las als een bezetene en schreef als een bezetene. Naast geleerde verhandelingen over onderwerpen uit de wiskunde, filosofie en natuurwetenschap schreef hij boeken over de meest uiteenlopende onderwerpen, van een boek over de interpretatie van dromen tot een lofzang op keizer Nero. Van al zijn werken heeft alleen zijn kleurrijke en zeer persoonlijke autobiografie *Over mijn leven* de eeuwen getrotseerd. Het is een klassieker geworden in het literaire genre van de autobiografie. In zijn autobiografie valt te lezen dat de geleerde een verwoed gokker was. Hij schreef er zelfs een boekje over, getiteld *Over kansspelen*, het eerste in zijn soort. In dit boekje introduceert hij een principe om verwachte winst en verwacht verlies gelijkelijk over de spelers te verdelen, de regel van Cardano. Deze regel is feitelijk een voorloper van de beroemde kansformule van Laplace.

Cardano's jonge jaren

Cardano was zijn leven lang een vechter. Als baby overleeft hij niet alleen een abortus, maar ook de pestepidemie van zijn geboortjaar. Minder fortuinlijk zijn de drie broers die hij heeft. Ze bezwijken allen aan de Zwarte Dood. Zijn vader is bevriend met Leonardo da Vinci, die in de jonge jaren van Cardano regelmatig over de vloer komt. Samen met vader Cardano bespreekt Leonardo dan de wiskun-

dige problemen die hem bezighouden. Vader Cardano is dan wel jurist, maar geeft eigenlijk meer om wiskunde en astrologie. Girolamo is amper twaalf wanneer zijn vader hem de eerste zes boeken van Euclides' *Elementen* laat bestuderen. Op zijn negentiende verlaat Girolamo het ouderlijk huis om medicijnen en filosofie te gaan studeren aan de universiteit in Pavia. Wiskunde is dan nog geen aparte, volwaardige studie zoals nu. Met het geven van bijlessen wiskunde aan medestudenten verdient hij er wat geld bij. Dan al maakt Cardano naam. Hij blinkt uit in universitaire debatten over de meest uiteenlopende vraagstukken. En in feestvieren!

Cardano en de Stotteraar

Eenmaal afgestudeerd leidt hij als arts aanvankelijk een armoedig bestaan. Samen met zijn briljante jonggestorven leerling Ludovico Ferrari (1522–1565) schrijft hij de *Ars Magna* (letterlijk vertaald 'De grote kunst'), een boek dat nog enige bekendheid geniet onder historici omdat het een overzicht biedt van de complete en meest recente kennis op het terrein van de algebra in zijn tijd. Het zou een kleine eeuw lang hét standaardwerk zijn. Het levert hem onder wiskundigen de eerste internationale erkenning op.

Het brengt hem ook de woede en haat van de Italiaanse wiskundige Niccolò Tartaglia (1499–1557), bijgenaamd De Stotteraar. Tartaglia had Cardano de door hem ge-



Girolamo Cardano. Gravure van Clemens Ammon, 1652



vonden methode om derdegraads vergelijkingen op te lossen onder geheimhouding in de vorm van een cryptisch gedicht toevertrouwd. Cardano had de formules van Tartaglia toch in zijn boek gepubliceerd.

Tartaglia was woest en beschuldigde Cardano ervan de methode onder zijn eigen naam te hebben gepubliceerd. Dat was niet zo. Bovendien wist Cardano dat niet Tartaglia de eerste was die dit soort vergelijkingen wist op te lossen maar Scipione del Ferro (1465–1526), die er echter niets over had gepubliceerd. Desondanks ontspint er zich een lange en venijnige correspondentie tussen Tartaglia enerzijds en Cardano en Ferrari anderzijds, waarbij de wiskundigen elkaar aftroeven in sarcasme. Uiteindelijk culmineert de briefwisseling in een openbaar debat tussen Tartaglia en Cardano's leerling Ferrari. Ferrari is overduidelijk de meerdere van Tartaglia, die voortijdig de zaal verlaat. Tartaglia zou zijn leven lang volhouden dat niet 'het schepsel van Cardano' zoals hij Ferrari consequent noemt, maar hij de strijd heeft gewonnen. De rancune van De Stotteraar zou Cardano tientallen jaren achtervolgen. Het zou niet de enige keer zijn dat Cardano slachtoffer werd van de afgunst van collega's.

Op de toppen van zijn roem

Na de publicatie van nog enkele andere boeken groeit Cardano's bekendheid buiten de grenzen van Italië. Maar zijn carrière bereikt pas echt een duizelingwekkend hoogtepunt wanneer hij onverwachts het verzoek uit Schotland ontvangt om de aartsbisschop van St. Andrews, een belangrijk en machtig man, van zijn ernstige vorm van astma te genezen. Cardano reist naar Schotland en schrijft de geestelijke een regiem van rust, reinheid en regelmaat voor. En inderdaad, na jaren van astmatische aanvallen geneest de aartsbisschop! Het nieuws gaat als een lopend vuurtje langs de hoven van Europa. Verscheidene vorsten waaronder de koning van Frankrijk en keizer Karel V willen hem inlijven als hofarts, maar Cardano weigert. Hij wil geen partij kiezen tussen de machtige rivalen en heeft bovendien zijn persoonlijke vrijheid lief.

Donkere tijden

Cardano is nu overal een beroemdheid geworden en eenmaal thuis wordt de deur van zijn praktijk platgelopen door gefortuneerde en hooggeplaatste patiënten. Maar

ook de tegenslagen van de fortuin worden hem niet bespaard. Zijn ene zoon zwerft stelend en dobbelend door Italië – hij besteelt zelfs zijn eigen vader – en zijn andere zoon wordt berecht en onthoofd omdat hij in een vlaag van woede zijn vrouw heeft vergiftigd nadat zij op het kraambed had bekend dat hij van geen van zijn kinderen de vader was. Het zijn donkere tijden voor Cardano, maar het zou nog erger worden. Cardano wordt door de inquisitie gearresteerd en gevangen gezet. De redenen zijn nooit helemaal opgehelderd. Deels zal het zijn religieuze vrijzinnigheid zijn geweest, deels zal het de paus misnoegd hebben dat Cardano in het verleden de hoscop van Jezus Christus had gepubliceerd. Door een goed woordje van twee bevriende kardinalen ontloopt hij de brandstapel en wordt hij na 77 loodzware dagen van gevangenschap op borgtocht vrijgelaten. In ruil voor de genade van de paus verplicht hij zich zijn leven lang niets meer te publiceren. Hij vestigt zich in Rome, in de nabijheid van zijn beschermheren. De inwoners van de stad zien de excentriek geklede en oud geworden geleerde dagelijks moeizaam door de straten en over de bruggen van Rome voortschuifelen. Thuis voltooit hij aan zijn schrijftafel nog juist voordat hij sterft, zijn autobiografie.

Een handboek voor gokkers

Cardano was een verwoed gokker. In perioden dat hij krap bij kas zat, of wanneer hij zijn verdriet wilde verdrijven, stortte hij zich met volle overgave op het gokken. Dag in, dag uit. Jarenlang. En Cardano zou Cardano niet zijn als hij niet ook over het gokken had geschreven. In het Latijn omdat het de internationale voertaal was waarin bijna alle geleerde en wetenschappelijke literatuur werd geschreven. Zo kon er ook buiten de eigen landsgrenzen kennis van worden genomen. Het boekje is geen wiskundige verhandeling. Het is eerder een geleerd handboek over gokken, het eerste in zijn soort. Lange tijd heeft het niet de waardering gekregen die het verdient. Dat heeft zowel te maken met Cardano's snelle manier van schrijven – fouten liet hij vaak staan te midden van latere correcties – als met het schijnbare gebrek aan enige interne logica in het boek. Men vond het maar een warboel van notities waarin geen enkele lijn viel te ontdekken. Maar die lijn is er wel degelijk.

Kritiek op het gokken

Er was door de eeuwen heen nogal wat kritiek op het gok-

ken. In de eerste plaats van de zijde van de kerk, maar ook de antieke Griekse filosoof Aristoteles (384–323 v.Chr.) veroordeelde al het gokken in zijn boek over het ware geluk, de *Ethica Nicomachea*. De filosoof noemt gokkers in één adem met dieven en rovers. De gokker is hebzuchtig en slechts uit op schaamteloos winstbejag. Cardano pareert in een apart hoofdstuk de kritiek van Aristoteles. Dat is misschien verbazingwekkend. De Griekse filosoof was immers al zo'n tweeduizend jaar eerder gestorven. Maar Aristoteles was niet zomaar een criticus. Aristoteles was in de traditionele filosofie die aan de universiteiten werd onderwezen verreweg de best bekende filosoof uit de Oudheid en gold als gezaghebbend. Zijn *Ethica Nicomachea* was in de tijd van Cardano een veel gelezen boek en Cardano was één van de meest belezen geleerden van zijn tijd.

Fair play

In de ogen van Cardano, zelf immers een fervent gokker, bestond er wel degelijk een eerlijke vorm van gokken. Namelijk wanneer er volkomen gelijkwaardigheid bestaat tussen beide spelers. 'Gelijkwaardigheid,' schrijft hij, 'is het allerbelangrijkste beginsel van het gokspel.' Het is de rode draad in zijn boek. Cardano schetst een aanschouwelijk beeld van hoe het er tijdens het gokken in de praktijk aan toeging. Partijdige toeschouwers die zich luidruchtig met het gokken bemoeien, meningsverschillen die ont-aarden in slaande ruzies, list en bedrog, onevenwichtige dobbelborden en valse dobbelstenen. In al deze gevallen – de keuze van je tegenspeler, de toeschouwers, de omstandigheden van de locatie, het geld waarom gespeeld wordt, de dobbelstele en ook de dobbelsteen zelf – zou gelijkwaardigheid moeten heersen.

Maar Cardano gaat nog een stap verder. Niet alleen de omstandigheden moeten voor elk van beide spelers gelijk zijn, gelijkwaardigheid moet ook een eigenschap van het spel zelf zijn! Dat wil zeggen dat geen van beide spelers in het voordeel mag zijn. Alleen, hoe bereik je deze gelijkwaardigheid in het dobbelspel? Als Cardano daar eenmaal een antwoord op had gevonden, kon hij de kritiek van Aristoteles pareren dat gokken niets anders is dan oneerlijk winstbejag. Het antwoord vindt hij verrassenderwijs bij Aristoteles zelf, in dezelfde *Ethica Nicomachea*.

Gelijkwaardigheid volgens Aristoteles

Het vijfde deel van de *Ethica Nicomachea*, een klassieke

tekst in de filosofie van het recht, gaat over rechtvaardigheid. Rechtvaardigheid tussen mensen, schrijft de filosoof, is gebaseerd op gelijkwaardigheid. Niet toevalligerwijs precies het criterium van Cardano voor een eerlijk spel. Volgens Aristoteles kan je op twee manieren gelijkwaardigheid bereiken. Beide manieren geven een eerlijke verdeling. Allereerst is er een gelijkwaardigheid die in verhouding staat tot de verdiensten van mensen. Als iemand bijvoorbeeld twee keer zoveel werk heeft geleverd als een ander, dan heeft hij ook recht op een twee keer zo grote beloning. Dus als er 30 muntstukken te verdelen zijn, krijgt de één er 20 en de ander 10. Deze eerlijke verdeling zou je 'proportionele gelijkwaardigheid' kunnen noemen.

De tweede vorm van gelijkwaardigheid geldt als mensen gelijke rechten hebben. Stel dat twee personen recht hebben op hetzelfde bedrag aan geld, maar dat de één onterecht winst heeft gemaakt ten koste van de ander. Dan tel je, schrijft de filosoof, eerst de bedragen van beiden bij elkaar op en daarvan neem je het midden, d.w.z. de helft van het totaal. Stel ze hebben samen 30 muntstukken, de één 20, de ander 10. Om gelijkwaardigheid te verkrijgen krijgt de benadeelde nu dus 5 muntstukken en verliest de ander hetzelfde bedrag zodat beiden op het 'midden', 15 muntstukken uitkomen. Deze vorm van een eerlijke verdeling zou je 'evenwichtige gelijkwaardigheid' kunnen noemen. Cardano zelf noemt het de methode van het midden.

Cardano's methode van het midden

Cardano zegt het niet met zoveel woorden, maar een eerlijk spel is duidelijk een spel waarbij geen van beide spelers in het voordeel is. Dat is het geval wanneer de verwachte winst en het verwachte verlies evenwichtig tussen beide spelers zijn verdeeld. Het lijkt op het eerste gezicht dus logisch om van een *evenwichtige verdeling* uit te gaan. En dat is inderdaad ook de strategie die Cardano aanvankelijk kiest. Een simpel voorbeeld kan Cardano's methode van het midden illustreren.

Stel je speelt met zijn tweeën een spel met één dobbelsteen. Het is een zuivere dobbelsteen zodat de zes mogelijke uitkomsten allemaal even goed mogelijk zijn. Stel je gokt op de uitkomst zes en je zet elk van beiden eenzelfde bedrag in. Dan heb je bij één worp maar één mogelijke uitkomst die winst geeft. Namelijk een zes. Gooi je nu drie keer dan heb je drie keer de mogelijkheid





om een zes te gooien. En dat is precies het berekende midden van het aantal mogelijke uitkomsten. Dus als je een weddenschap afsluit om in drie worpen ten minste één keer zes te gooien, zo redeneert Cardano in het voetspoor van Aristoteles, is de verwachte winst en het verwachte verlies voor beide spelers gelijk verdeeld. Alleen, klopt de methode van het midden wel? Nee!

De fout van Cardano

Om makkelijker in te zien wat er scheelt aan Cardano's methode van het midden, vertalen we zijn redenering in kansen. Cardano kende het begrip kans nog niet. Daarom moeten we hem de fout die hij maakt ook niet te sterk aanrekenen. Als je nog niets van kansrekening weet, dan is je eerste ingeving misschien ook wel om de kansen evenredig met het aantal worpen op te tellen. In één worp heb je een kans van 1 op 6 ofwel $1/6$ om zes te gooien. Als je nog een keer gooit heb je weer een kans van $1/6$ om zes te gooien. Wat is er dan logischer dan te denken dat voor de kans om in twee worpen een zes te gooien je alleen maar de twee gevonden kansen hoeft op te tellen? En voor de kans om in drie worpen een zes te gooien daar nogmaals $1/6$ bij op te tellen? Je komt dan op een kans van $3/6$ ofwel $1/2$. Precies het midden dat Cardano zoekt. Alleen, het is fout. Je kunt kansen of, zoals Cardano deed, mogelijke uitkomsten niet zomaar evenredig aan het aantal worpen optellen. Dat leidt tot vreemde resultaten. Stel dat je zes keer zou werpen, dan zou de kans om een zes te gooien zes keer $1/6$ zijn ofwel 1 (100%). Je zou er dan zeker van zijn dat je in zes worpen een zes gooit. Maar iedereen weet wel dat je best zes keer kan gooien zonder dat er ook maar één keer een zes valt. Bovendien, wat voor zes geldt, geldt ook voor de andere mogelijke uitkomsten van de dobbelsteen. Het zou dus als je zo redeneert zeker zijn dat in zes worpen elke uitkomst één keer voorkomt. De conclusie kan alleen maar luiden dat kansen niet evenredig aan het aantal worpen toenemen, ook al leek dat intuïtief zo logisch.

Halverwege hoofdstuk veertien van zijn boek heeft Cardano echter een moment van inzicht. De kansen om een zes te gooien in drie worpen zijn helemaal niet fifty-fifty verdeeld! Want als je driemaal met een dobbelsteen gooit, dan zijn er in totaal 216 verschillende mogelijke uitkomsten. Alle zes de mogelijke uitkomsten van de eerste worp kun je combineren met elk van de

zes mogelijke uitkomsten van de tweede worp. En de 36 mogelijke combinaties die je zo krijgt vallen op hun beurt weer te combineren met elk van de zes mogelijke uitkomsten van de derde worp. Ofwel bij drie worpen zijn er 216 verschillende mogelijke uitkomsten. Maar als je goed telt en rekent, zijn er onder deze 216 mogelijke uitkomsten niet 108 (de helft) maar slechts 91 mogelijke uitkomsten waarbij je ten minste één zes gooit. Er zijn namelijk 75 mogelijke uitkomsten met precies één zes. De zes kan immers elk van de 3 worpen verschijnen, terwijl er voor elk van de overgebleven twee worpen dan 5 mogelijke uitkomsten overblijven. Verder zijn er 15 mogelijke uitkomsten met twee zessen. Want er zijn drie manieren waarop je in drie worpen twee zessen kan gooien, terwijl er voor de resterende worp dan nog 5 mogelijke uitkomsten overblijven. En tenslotte kan je op één manier drie zessen gooien. Als je deze aantallen bij elkaar optelt, kom je dus aan 91 mogelijke uitkomsten waarbij je tenminste één keer zes gooit. En dus zijn er $216 - 91 = 125$ mogelijke uitkomsten waarbij je géén zes gooit. Ofwel de gunstige en ongunstige uitkomsten verhouden zich helemaal niet fifty-fifty maar als 91 staat tot 125 (Cardano gebruikt het Latijnse woord *proportio* dat 'verhouding' betekent; in het Engels zou je zeggen *the odds*). De methode van het midden kan dus niet goed zijn.

Cardano's regel

Hoe nu verder? Cardano zoekt zijn toevlucht tot Aristoteles' andere vorm van een eerlijke verdeling: proportionele gelijkwaardigheid. De geleerde gokker legt het in zijn boek uit. Eerst stel je vast hoeveel uitkomsten er in totaal mogelijk zijn. Vervolgens hoeveel daarvan gunstig voor jou zijn. Dit laatste aantal vergelijk je met het aantal mogelijke uitkomsten dat overblijft, ofwel de ongunstige uitkomsten. Die krijg je vanzelfsprekend door het aantal gunstige uitkomsten af te trekken van het totaal aantal mogelijke uitkomsten. Cardano's regel is nu dat elk van beide spelers overeenkomstig de zo gevonden verhouding een bedrag moet inzetten. Anders gezegd, de hoogte van de geldelijke bijdragen van beide spelers moet in dezelfde verhouding staan als de *odds*. Dan zijn het verwachte verlies en de verwachte winst voor elk van de twee spelers gelijk en kan het spel eerlijk worden genoemd. Een simpel voorbeeld kan dit illustreren.

Stel je gokt dat je in twee worpen tenminste één keer een zes gooit. Wat is dan het aantal mogelijke uitkomsten waarbij je wint? Je kunt op verschillende manieren

tenminste één zes in twee worpen gooien: 1 6, 6 1, 6 6, 6 3, enzovoort. Als je deze bij elkaar optelt, zal je zien dat er 11 verschillende mogelijkheden zijn waarbij je wint. Het totaal aantal mogelijke uitkomsten bij twee worpen is 36 (6×6). Je vindt het aantal ongunstige uitkomsten door het gevonden aantal gunstige uitkomsten af te trekken van het totale aantal mogelijke uitkomsten. In ons geval is het aantal ongunstige uitkomsten dus gelijk aan $36 - 11$, dat is 25 (of zoals we nog makkelijker hadden kunnen berekenen: 5×5). Het aantal gunstige mogelijke uitkomsten verhoudt zich dus tot het aantal ongunstige mogelijke uitkomsten als 11 staat tot 25. Dat wil zeggen dat je verwacht gemiddeld in 11 van de 36 gevallen te winnen en in 25 van de 36 gevallen te verliezen.

Cardano's regel zegt nu dat je de gezochte gelijkwaardigheid bereikt als de ingezette bedragen zich net zo verhouden als deze *odds*. Dat is het geval wanneer jij om 11 muntstukken wedt dat je ten minste één zes gooit en je tegenstander om 25 muntstukken van niet. Aangezien je gemiddeld 11 van de 36 keer wint, is jouw verwachte winst in 36 worpen gelijk aan 11×25 muntstukken = 275 muntstukken. En aangezien je gemiddeld 25 keer van de 36 verliest, is de verwachte winst van je opponent gelijk aan 25×11 muntstukken, ofwel ook 275 muntstukken. Verwachte winst en verlies zijn zo precies in evenwicht en dus is er sprake van een eerlijk spel.

Een eerlijke manier van gokken

Je zou Cardano's boekje kunnen lezen als een fictief debat waarin hij de handschoen opneemt tegen Aristoteles. Cardano is dan de terechte winnaar. Hij heeft de grote filosoof met zijn eigen wapens verslagen. Hij heeft laten zien dat er een eerlijke vorm van gokken bestaat. Als je je maar aan zijn regel houdt, dan zijn de verwachte winst en het verwachte verlies voor beide spelers aan elkaar gelijk. Niet alle gokkers vallen dus over één kam te scheren met dieven en rovers.

Cardano presenteert zijn vondst als een nieuw inzicht. In een tijd waarin niemand nog over kansen sprak, was deze verdeling van de inzetten overeenkomstig de berekende *odds* heel wat minder vanzelfsprekend dan nu. Cardano wil, zo schrijft hij, zijn ontdekking niet geheim houden. Wat natuurlijk een steek onder water is aan Niccolò Tartaglia. Veel mensen die Aristoteles niet begrijpen, gaan de fout in – namelijk doordat ze zich met oneerlijke weddenschappen inlaten – en verliezen op die manier veel geld. En daar wil hij naïeve gokkers voor behoeden.

Een klassieke kansformule



Cardano's regel is een algemene regel. Gokkers die de odds kennen, kunnen nu eerlijke weddenschappen aangaan. Of, als ze minder goede bedoelingen hebben, van deze kennis profiteren. Cardano is de eerste die verder kijkt dan individuele gokspelen. Belangrijker nog, hij had alle ingrediënten in handen om een algemene formule voor kansen te formuleren! Als hij het aantal gunstige mogelijke uitkomsten had gedeeld door het totaal aantal mogelijke uitkomsten, had hij een formule voor de kans op winst gehad. In ons voorafgaande voorbeeld om ten minste één keer zes te gooien met twee dobbelstenen, is die kans dus $11/36$ ofwel een kans van ongeveer 30,6%. Deze bekende kansformule vind je terug bij diverse latere wiskundigen. De klassieke verwoording is afkomstig van het wiskundige genie Pierre-Simon Laplace (1749–1827). De formule heet daarom ook wel de kansregel van Laplace. Laplace noemt de formule het eerste beginsel van de kansrekening en schrijft dat de kans op een gebeurtenis (bijvoorbeeld een zes gooien in drie worpen) gelijk is aan het aantal gunstige mogelijke uitkomsten gedeeld door het totaal van alle mogelijke uitkomsten. Op voorwaarde natuurlijk dat alle mogelijke uitkomsten dezelfde waarschijnlijkheid bezitten om plaats te vinden. De uitdrukkingen 'gunstige uitkomst' en 'ongunstige uitkomst' zijn in de kansrekening standaardtermen geworden, al gaat het tegenwoordig al lang niet meer alleen om dobbelspele en weddenschappen aan de goktafel.

De vraag is waarom Cardano, als hij daar zo dicht bij was, niet rekende met kansen in plaats van met wat de Engelsen de *odds* noemen. Met kansen kun je namelijk een stuk eenvoudiger rekenen. Bovendien zijn ze veel makkelijker onderling te vergelijken. Maar Cardano's boek was niet voor wiskundigen bedoeld, maar voor gokkers. En gokkers gaan weddenschappen aan. En dan zijn *odds* heel inzichtelijk en een makkelijke manier om je bijdrage aan de pot te berekenen. Waarom zou je dan nog eens de gevonden *odds* delen door het totaal aantal mogelijke uitkomsten? Het is nog steeds gebruikelijk om bij weddenschappen van de *odds* uit te gaan, denk maar aan paardenraces. Anders gezegd, *odds* zijn zo'n slechte keuze nog niet.

STEVEN TIJMS heeft Wiskunde en Klassieke Talen gestudeerd. Dit artikel is een bewerkte versie van een hoofdstuk uit zijn boek *Toeval is altijd logisch* dat in 2018 bij VU University Press verschenen is. E-mail: stijms@gmail.com

From killer nurses to quantum entanglement, and back Part 1

This article is loosely based on the farewell lecture I gave on the occasion of my retirement as professor of Mathematical Statistics in Leiden, one and a half years ago. A talk on statistics is not complete without some data and a talk on mathematics is not complete without the proof of a theorem, so you will see both. But, while reviewing the content of the formal lecture, I took a new look at some of the data from one of my killer nurse cases, and made some exciting new discoveries. So the article grew, and now has to be published in two parts. Welcome to Part 1. In the lecture, I presented two of my recent research passions: one of which enjoyed some measure of success, the other seemingly ending in failure. There are, I think, amusing links between the two, at various levels. And working on this article, I made new statistical discoveries concerning the case of the English nurse Ben Geen – that’s the research passion I was so unhappy about. I now feel some optimism.

RICHARD D. GILL

For me, the Ben Geen case starts with my earlier experience in the case of the Dutch nurse Lucia de B., and that starts with a small number of 2×2 tables (Figure 1). These tables played an enormous role in the conviction for serial murder of the Dutch nurse Lucia de Berk.

The data is also presented in a recent paper *Elementary*

Statistics on Trial (the case of Lucia de Berk) from R. D. Gill, P. Groeneboom and P. de Jong (2018). This data was analysed for the court in 2002 by my old colleague and old friend Henk Elffers, who had been contacted in 2001 by police investigators at the beginning of what became a ten year saga. I only got involved half way through.

JKZ *					RKZ-41 **					RKZ-42 ***				
		Incident in shift					Incident in shift					Incident in shift		
		Yes	No				Yes	No				Yes	No	
Lucia on duty	Yes	8	134	142	Lucia on duty	Yes	1	0	1	Lucia on duty	Yes	5	53	58
	No	0	887	887		No	4	361	365		No	9	272	281
		8	1021	1029			5	361	366			14	325	339

* Juliana Kinderziekenhuis, Medium Care Unit 1, 1 Oct 2000 – 9 Sep 2001; ** Rode Kruis Ziekenhuis, afdeling 41, 6 Aug – 26 Nov 1997; *** Rode Kruis Ziekenhuis, afdeling 42, 1 Aug – 30 Nov 1997

Figure 1. Roster data from the case of Lucia de B.



Foto: sudoki

During nearly a year ending in September 2001 when police investigations started, on a medium care ward at the Juliana Children’s Hospital (JKZ) in the Hague, there were in total 1029 (3×343) 8-hour shifts (3 shifts a day, 7 days a week). In 8 of them, an “incident” occurred. All 8 in the shifts of a certain nurse called Lucia. Several years earlier, at another hospital, the Red Cross Hospital (RKZ), during the same four months in two intensive care wards (Wards 41 and 42; the data from ward 42 misses 9 days at the beginning and end of the 4 month period) there were 5 and 14 incidents respectively. Lucia was only on duty once in RKZ-41 but on just that one occasion she netted one of the 5 incidents! She mainly worked on Ward 42, where she netted disproportionately many of the 14 incidents (she had one third of the incidents in only one sixth of the shifts). Does one need a statistician to interpret it to the board of judges of a criminal court where Lucia is being tried for serial murder? The data speak for themselves.

But do the data speak the truth, the whole truth, and nothing but the truth? As you perhaps already know, the Lucia case can be considered a success story. Her

life sentence for ten murders of children and old people got reversed. At the retrial, the judges, in their summing up, congratulated the nurses on their devotion and their professional efforts to save the lives of their patients, lives which (they said) were unnecessarily shortened through medical errors. The errors were caused by misdiagnosis, chaotic management, ignorance of the content of the patients’ medical dossiers ... they were committed by hospital specialists and hospital managers. A witch-hunt ensued and an intelligent nurse with a charismatic personality and a colourful past was a natural scapegoat.

Inspired by our success in getting Lucia a retrial (I was one of a group of people fighting for a retrial, from 2006) I got involved in several similar cases, in particular, in the case of one Ben Geen (*not* “Green”), who also got a life sentence for an enormous number of attacks on patients in a short time period in a small hospital in a provincial town in England ... at the same time as the Lucia case was playing out. I will show you some data from the Ben Geen case in a moment. But first, here are some more 2×2 tables.

Quantum entanglement

The data in the four 2 × 2 tables in Figure 2 were collected in an experiment performed in Delft in 2015 in two labs at opposite ends of the university campus. The results were rapidly published in the paper *Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres* from B. Hensen, H. Bernien, A. Dréau et al. (2015).

In the two labs, 1300 meters apart, we imagine some lab assistants at work, conventionally named Alice and Bob. The labs are connected by dedicated glass fibre cables and ordinary internet connections. It takes light four micro-seconds to traverse that distance (there are a million microseconds in one second). A lot will be done in both labs in many short time intervals each of length four microseconds: so short that there cannot be any influence on events in Alice’s lab, from events taking place in that time interval in Bob’s lab; or vice versa.

In each lab there is a diamond, and in each diamond there is a “nitrogen-vacancy” defect: two adjacent carbon atoms are missing and in their place are one nitrogen atom and one little hole. This “vacancy” is home of an electron whose “spin” in various directions can be addressed and manipulated and read-out with the help

of state-of-the-art lasers and electronics. If asked, it only ever answers “up” or “down”.

To start with, the two distant *spins* are prepared in a way which certainly can induce correlations between their properties. It is not necessary to explain how that is done, and indeed, perhaps it cannot be *explained*. We could run through the mathematics (a few lines of simple linear algebra of tensor-products), but that would not make things less mysterious. Then, at each of the two labs, a fair coin is tossed, heads marked “1” and tails marked “2”. In each lab, the resulting “1” or “2” is used as an *input* or a *setting* to set a switch to position “1” or position “2” on a measurement device which actually houses the diamond. The spin is immediately measured and a result comes out. This *measurement outcome* or *output* is binary, we call it “+” or “-” (these are just labels). Let me emphasize: **this happens within such a small time interval that there can be no knowledge at Alice’s side of Bob’s setting, till after Alice’s outcome is determined, and vice versa.**

All apparatus is then reset. The two spins are put into the “quantum entangled singlet state” anew. Repeat.

Call each repetition one *trial*. Stop when you have performed 245 trials. The measurement settings

correspond to directions in which “spin” is measured, and the outcome is “up” or “down” in the direction measured. It’s not possible for humans to actually toss fair coins and switch switches in four microseconds, so the random settings are generated by arguably equivalent but much faster physical devices.

Anyway, each trial has two *inputs* taking values in the set of two symbols {“1”, “2”}, let’s call them *a* and *b*, and two *outputs* taking values in the set of two symbols {“+”, “-”}, let’s call them *x* and *y*. So the data generated in one trial are represented by one string of four symbols *abxy*. There are 16 possible strings of four binary symbols. They were counted and are tabulated in the spreadsheet above.

Each *trial* is classified as being either a “success” or a “failure”. The definition of “success” is: “at least one input is “1” and the outputs are the same, or both inputs are “2” and the outputs are different”. There were 196 successes out of 245 trials, which is a success rate of 80%, and 80% is somewhat larger than 75% – in fact, statistically significant, (one-sided test) at the 5% level.

This experimental outcome resulted in a scientific paper very rapidly published in the prestigious journal *Nature*, and the news was reported in all the serious science supplements of decent newspapers everywhere in the world. In fact, a big race had been on. Three other experimental groups, in Munich, Vienna and at NIST in Boulder, Colorado, had performed variants of the same experiment at about the same time, but only published their results (in full agreement with the Delft findings) a little bit later. Delft was actually dark horse in the race, an unexpected contender.

Imagine tossing two fair coins repeatedly a total number of 245 times (trials). Each trial, the chance of seeing two tails is one quarter. We would expect around 75% of the trials to show one or two heads. The exciting thing about the Delft experiment is that the chance of seeing one or two heads in 196 or more of such trials (i.e., at least 80% of the trials) is only 3.9%, which is less than the magic “significance level” 5%. Why that is exciting, in fact, perhaps earth-shattering, I will try to explain in a moment. Just significant with a one-sided test at the 5% level, though perhaps guaranteeing you a publication in regional economics or social psychology, would not usually be enough to guarantee a publication in *Nature* in the field of physics. Despite this, it is fair to say that this

experiment represents a milestone reached – a milestone we had been trying 50 years to reach.

There is plenty of evidence that this was not just a lucky chance. Aside from the copious support from quantum physics itself, reported in the published paper, we have the results of the other three experiments, each on their own passing the same milestone while using quite different quantum technologies. Taken together, the statistical evidence seems to me to be overwhelming.

By the way, quantum theory says that it is impossible to get beyond 85%, or, more precisely $100 \times (2 + \sqrt{2}) / 4 \%$. That’s a bound which needs sophisticated mathematics to derive and from which, alas, one does not gain much intuition. I could drop the name of Grootendieck and I could say something about the geometry of Hilbert space ... This means that the question arises that maybe, there is a better theory than quantum mechanics which would allow a higher limit? There is surely no reason to suppose that quantum mechanics is a “last stop” for physics. There are arguments that the real bound could be 100%. Some progress has been made finding an intuitive reason behind the mysterious $(2 + \sqrt{2}) / 4$, but the problem is still essentially open.

I will try to explain how I had something to do with the whole thing (also with the results obtained in Vienna and Munich and at NIST). The connection has to do with *martingale theory*, which has a lot to do with statistics. In particular, it has a lot to do with randomised clinical trials and the whole idea of randomisation in statistics. (I’ll also explain how Lucia’s exoneration and release from jail also had something to do with my interests in quantum mechanics. Everything is connected!)

First a little bit of probability theory. Consider four random variables X_1, X_2, Y_1, Y_2 which take the values ± 1 . Consider the sum of four indicator variables

$$I(X_1 = Y_1) + I(X_1 = Y_2) + I(X_2 = Y_1) + I(X_2 \neq Y_2).$$

Obviously, this sum can only take the values 0, 1, 2, 3 and 4. It can only take the value 4 if all four events occur simultaneously. But

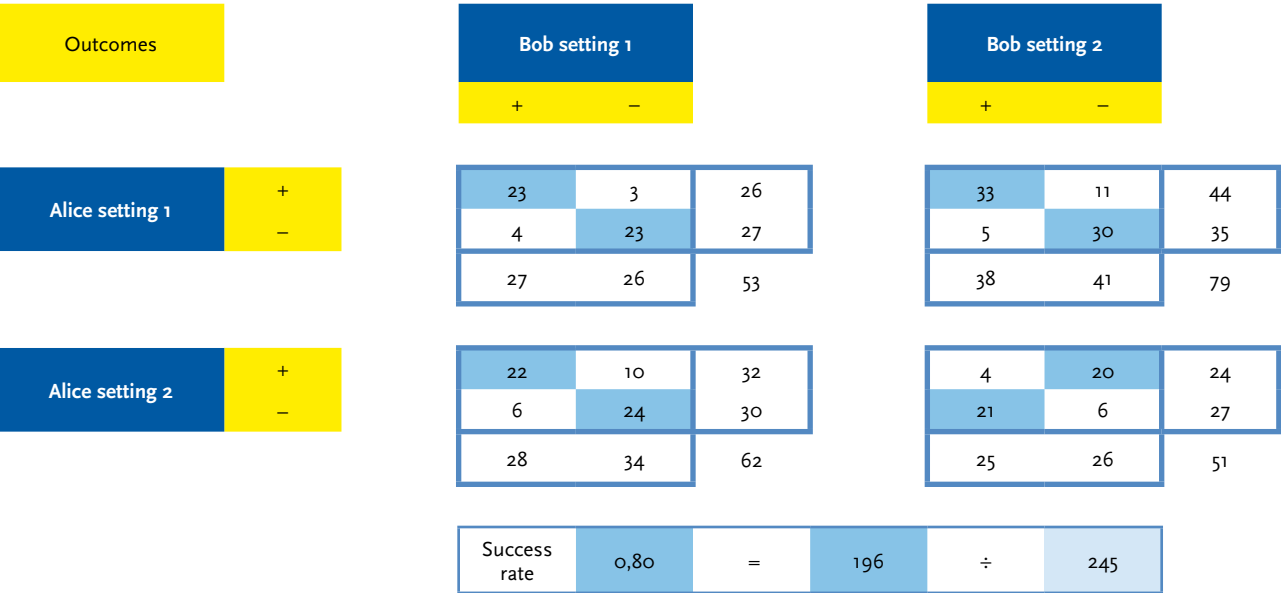
$$X_1 = Y_1 \ \& \ X_1 = Y_2 \ \& \ X_2 = Y_1 \ \Rightarrow \ X_2 = Y_2.$$


Figure 2. The Delft Bell experiment

So, the value 4 cannot be attained. The expression can't exceed 3. Divide the resulting inequality throughout by 4. We obtain

$$\frac{1}{4} I(X_1 = Y_1) + \frac{1}{4} I(X_1 = Y_2) + \frac{1}{4} I(X_2 = Y_1) + \frac{1}{4} I(X_2 \neq Y_2) \leq \frac{3}{4}.$$

Now I connect this to the Delft experiment. Suppose that for each trial, counterfactual outcomes could be defined, X_1, X_2, Y_1, Y_2 , which stand for: "the outcome which Alice would see were her setting equal to 1", "the outcome which Alice would see were her setting equal to 2", "the outcome which Bob would see were his setting equal to 1", "the outcome which Bob would see were his setting equal to 2". These could be constructed (perhaps in many ways) in any crypto-deterministic (i.e., having a behind-the-scenes mechanistic explanation) mathematical physical theory which does not have action at a distance: Alice's possible outcomes can't depend on Bob's actual setting; and vice versa. Four microseconds is too short for the information to have travelled from Bob's side to Alice's. Both of Alice's counterfactual outcomes (the outcomes she would have seen, if either of her two possible inputs had been supplied) are functions of the stuff on Alice's side of the experiment, and the stuff connecting it to the source and at the central source itself, but can't possibly depend on the actual input at Bob's side.

Everything might depend on everything which happened in the past in both wings of the experiment. Moreover, there might be correlated time trends and time jumps in the physics on the two sides of the experiment. But if the two settings are chosen at each trial, anew, completely at random, independently of the past, then for each trial, conditional on all preceding ones, the just proven inequality says that the conditional probability of success (as it was defined earlier), conditional on the past trials in both wings of the experiment, is less than or equal to $\frac{3}{4}$. We have found that if we define a stochastic process by taking its increments to equal $I(\text{success}) - \frac{3}{4}$, this process is a supermartingale.

It's not difficult to prove from this (by a recursive coupling argument, and taking account of the 0/1 nature of the indicator random variable) that the random variable #successes is stochastically smaller than the Binomial(#trials, $\frac{3}{4}$) distribution; in other words, right-tail probabilities are less than or equal to Binomial right-

tail probabilities.

I developed these martingale ideas back in 2001, introducing the idea of exploiting the randomisation of the settings instead of making i.i.d. assumptions about the physics of the outcomes in order to get safe bounds on the probability of large deviations from the Bell-CHSH local realism bound. I wanted such bounds because I wanted to hold a bet, and indeed, to win a bet, against an opponent who claimed they could simulate the quantum mechanical correlations on a network of computers, where the computer network reflected the spatio-temporal constraints of a "loophole-free" Bell-type experiment. I wanted to be pretty sure that I would win! My opponent complained about the rules which I wanted to impose. The independent jury we had recruited chickened out of their scientific duty to agree that my experimental protocol was fair. These are the rules which are imposed in the experiments of 2015 and later, and which had been written down 35 years before then.

For a fairly recent survey, see my paper "Statistics, Causality and Bell's Theorem" (Gill, 2014). A new survey is badly needed, thanks to the 2015 experiments.

Killer nurses

I now go back to (probably innocent) convicted serial killer nurses. The dataset presented in Figure 3 helped get the young English nurse Ben Geen a life sentence for two counts of murder and 15 of grievous bodily harm (a 16th count of grievous bodily harm was not considered proven), in the three consecutive months of December 2003, January 2004 and February 2004. I am certain he is innocent, just as I am certain that Lucia is innocent. And for much the same reasons. The reasons have little to do with statistics. The reasons have to do with the social structures in a modern hospital and the facts that (a) sick people do die in hospitals, (b) doctors do make mistakes, (c) top hospital managers and top medical specialists need to protect the reputation of their hospital. A fourth reason is (d) the coincidence that this case occurred shortly after the Shipman Enquiry, which blamed health-care administrators for not earlier noticing serial killer doctor Harold Shipman, who maybe murdered 250 patients.

Obviously, one of the most important factors in a doctor-patient relationship is that the patient has trust in his or her doctor. The medical establishment consequently has a strong interest in patients collectively having trust in doctors, and more generally in their health-care system. Similarly, we need to have confidence in our judges and in our legal systems. The guardians of our legal system strongly believe that our judges and our legal systems must not be seen to make mistakes. Unfortunately, a system which cannot admit to making mistakes can never learn from mistakes and is doomed to repeat them.

Learning from mistakes is good, but a new danger then arises that by learning the wrong lessons from one kind of mistake, one might increase the chance of making the opposite mistake. If the rate of false convictions goes down but nothing else really changes, the rate of false acquittals will go up. The more easily a health-care system goes into alarm-mode because of suspicion that it harbours a health-care serial killer, the more often innocent health-care professionals will trigger an alarm.

This key data-set in the Ben Geen case was later presented to the court by Michelle Brock, head-nurse of the Accidents and Emergency department where Ben Geen worked, at Horton General Hospital, a rather

small hospital in the provincial market town Banbury in North Oxfordshire. Together with a dossier of perhaps 30 incidents all from December 2003 onwards, it had initially been compiled in great haste before the case was reported to the police. Michelle and some colleagues based their work on patient records and nurse attendance records at the hospital, looking only at what happened during Ben's shifts, their investigation triggered and guided by recent memory and gossip. Ben, who was a trainee nurse, had won a higher qualification at the beginning of December, allowing him to work under less supervision than before. The trigger for their investigation had been two sudden and surprising collapses of patients who had just entered A and E (also known as ED: Emergency Department) on Thursday 5 February. Ben had reported sick on Friday, and had had a free weekend after that. He was arrested on Monday morning, 9 February 2004, as he arrived for work (one third of the way into the last bar of the bar-chart). The bar-chart was later presented to the court and was also known to the medical experts who were consulted on the 18 individual cases. There is no doubt it had a big impact on everyone involved in the trial, including journalists covering the trial.

We catch a glimpse from the chart of the fact that a lot of old people and people with existing serious health

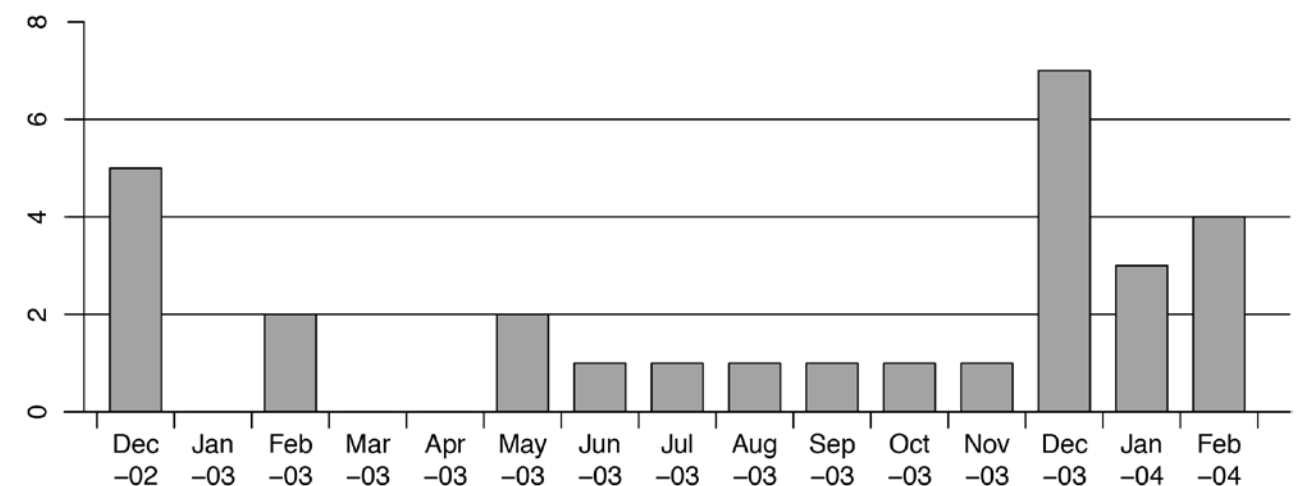


Figure 3. Admissions to critical care from the emergency department with a diagnosis of cardio-respiratory or respiratory arrest or hypoglycaemia, data: Head Nurse Brock

problems get brought to emergency care during the winter months of December, January and February with acute problems involving heart and lungs (a hot summer is also a danger period). A common diagnosis is *cardio-respiratory arrest* (the heart has stopped working and consequently the lungs too), much less common is “pure” *respiratory arrest* (the lungs have stopped working); fairly common is *hypoglycaemia*: a fall in blood glucose level. It *causes* fainting; breathing stops or is much suppressed. It can be *caused* by too much insulin or other glucose lowering diabetes tablets, delaying or missing a meal, not eating enough carbohydrate, unplanned physical activity, more strenuous exercise than usual, drinking alcohol – the risk of hypoglycaemia increases, the more alcohol you drink. In the bar-chart, nurse Brock has combined the three “standard” categories cardio-respiratory, respiratory, and hypoglycaemic arrest; but what is the correct category is hard to guess when a patient presents (arrives at the hospital). Past medical history, *and future medical events* will give clues as to what was actually going on. In an emergency situation, past medical history may be unknown.

Hospital nurses and authorities had been worried by the behaviour of the young male trainee nurse Ben Geen already before December 2003. His father was in the army, his mother was a nurse. He had been in the territorial army medical corps, and his ambition was to be qualified and then transferred as a combat medic to a military field hospital in Iraq. He was energetic and very “present”, keen to get action and to get experience. He made some other nurses nervous. They were calling him “Ben Allitt” behind his back, not such a nice joke, since Bev Allitt is the very well-known name of a pre-Shipman famous English convicted serial killer nurse (an interesting case which, in my opinion, deserves fresh study).

In December 2003 the numbers of patients reaching an overstressed emergency ward in an underfunded minor hospital in a provincial English town, threatened by closure because of its small size, and with perhaps not the very best of medical specialists and not the very best of management, was bigger than ever. There were a couple of “surprising” events when patients who were initially thought to be in fairly good shape suddenly, and at the time unexplainably, worsened. Ben was usually around when anything happened simply because he was

usually around: he was working double shifts in order to gain more and more experience as fast as possible, and also often fell in for absent colleagues.

On Thursday 5 February 2004, at the end of a hectic day, a chronic alcoholic diabetic was brought into hospital by his drinking mates, throwing up repeatedly and feeling very poorly, suffering fainting fits. Ben took a blood sample. The patient suddenly worsened and later had little idea what else Ben did to him. He certainly inserted a *canula* (a tube that can be inserted into the body, often for the delivery or removal of fluid or for the gathering of samples) and the patient was rapidly transferred to “critical care”. The real problem for Ben came later: Ben went home with, unknown to himself (he said), a used plastic needle-less syringe containing some muscle relaxant in his nurse’s smock. Such a syringe is used to administer necessary medications, including a muscle relaxant, through the *canula* prior to inserting breathing and feeding tubes into patients in the critical care ward. Ben stayed home sick on Friday, and then had the weekend free. His girlfriend, another nurse, doing the washing, told him off for this (she said) and told him to take it back as soon as possible. So, on Monday morning – with the syringe in his coat pocket – he was met by policemen as he entered the hospital. In some panic (he said) he stupidly further emptied the remaining contents of the syringe into his pocket. Obviously, he tried to harm patients by injecting them with this stuff so that he could then play the hero, helping to resuscitate them! The so-called “Munchhausen by proxy” syndrome.

At his trial, the Crown secured the services of a famous and experienced expert (a highly distinguished professor of Anaesthesiology), who found a number of the events highly suspicious; another confidently swore that never ever in his long experience had he met with an *unexplained* respiratory arrest. They all agreed on that ...

Of course they did. All respiratory arrests are “explainable”, though different experts often give different explanations. Actually, whether a collapse is diagnosed as cardio-respiratory, respiratory or hypoglycaemic can be pretty arbitrary. When either heart or lungs get into difficulty, the other organ rapidly gets into difficulties too. Hypoglycaemic arrest (critically low blood glucose levels) always involves breathing problems (you faint when not enough oxygen is reaching your brain) and can trigger

further deterioration of heart and lung function. Reduced oxygen levels affect brain, heart, lung. Muscles burn oxygen, the brain burns oxygen. All arrests are explainable, but the categories which are ticked on forms in the patient’s dossier and in the hospital’s administrative records may differ and may be revised in the light of later events. The categories which tend to be chosen by nurses, doctors and administrators may depend on who is doing it, and may show trends and jumps as time goes by. Just one occurrence of an unusual diagnosis alerts people to its existence, and they start seeing it every day: the well known *Baader-Meinhof phenomenon*.

At the time each had actually occurred, each of the 18 cases in the criminal charges against Ben had been “normal”. The last two had surprised some people (certainly not all), but because of earlier suspicion and gossip, they had triggered an emergency weekend-long internal hospital investigation, in which more than 30 dossiers of patients who had in recent months gone through Emergency *while Ben was on duty* were combed through, resulting in a dossier of 18 cases to hand over to the police on Monday. In fact the teams had access to 4000 patient medical records but were not interested in what happened when Ben was not there. Expert witnesses for the defence later explained how explainable each of the 18 was, though they were honest enough to admit that some cases were too complex to come to any clear conclusion. The prosecution had more expensive and more court-experienced experts than the defence. The prosecution experts were of course specifically hired to point out anomalies in each of the selected 18 cases, and tended to be rather confident of their diagnoses. Prosecution experts are “instructed” by the prosecution, defence experts are “instructed” by the defence.

Ben must have used a myriad of different techniques to cause all these unexplained medical emergencies and in many cases the expert witnesses called by the Crown in fact had conflicting ideas of what he might have done; though they did of course agree that he must have done something. All of the 18 patients were very sick, and what happened to each was what you may well have expected to happen in view of their existing severe and often complex conditions. But sometimes developments are fast, you do not “see them coming”, and so a sudden worsening takes some nurses or some doctors by surprise. People,

including Ben himself, did notice Ben often being there when such events took place. He had said, and said it in court again, that he thought he had been jinxed.

Ben’s unemotional and careful account of what he could recall that he had seen and done in each case, the impression he gave that he knew the law better than the lawyers, the eminent professor’s categorical statement that he had never seen an unexplained respiratory collapse in all his career, and the smoking gun which was the syringe, together clinched the matter for the jury. It mainly consisted of decent retired folk who had spent most of the trial napping during the presentation of interminable medical evidence (in 18 cases). The judge in his summing up made it very clear what verdict he expected from the jury.

Blood and urine samples from the trigger case showed traces of a muscle relaxant as well as of plenty of sedatives, but unfortunately the samples were not dated – one has no idea when they were taken nor by whom! Sedatives and muscle relaxant *should* have been present. The traces of muscle relaxant were of the same kind as was in the syringe. The consultant anaesthesiologist who had attended to the trigger patient as he went into intensive care said that she had asked (another nurse, later) for a different one. Ben said that he was not told to administer muscle relaxant, so, of course, had not done so. Hospital records were woefully incomplete. Since the earlier cases were not at the time thought to be suspicious, all samples of blood and urine had long ago been thrown away.

The annual pattern we see in that data can be seen in data which I analysed from many similar hospitals all over Britain. Of course, there is no data whatsoever about *unexplained* respiratory arrests. The data stored in a hospital database are administrative data. Every event has been put into a pre-existing category with an explanation, because it is not possible to enter it into the data base otherwise. The data in the database determines the fees of the medical consultants (the medical specialists) and the funding of the hospital. The data is not collected for scientific research or forensic investigation.

The three standard categories relevant to this case are *cardio-respiratory arrest*, *respiratory arrest*, and *hypoglycaemic arrest*. I already showed you the data supplied to the court by Ben’s head nurse, combining

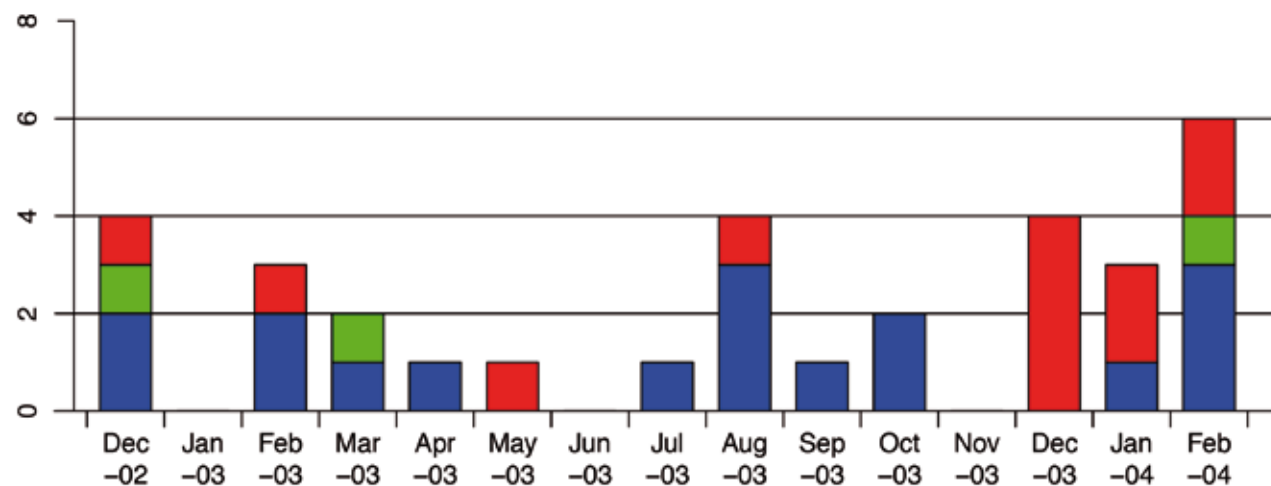


Figure 4. Admissions to CC from ED with CR, Hypo or Resp arrest, FOI data, Cardio-respiratory (blue), hypoglycaemic (green), respiratory (red)

those three categories. Much later, we got from Ben's hospital the data as presently archived in official hospital records. It was different! Also, the categories are still separate, and we have data from many more years.

The total numbers of relevant cases in December 2002 and in December 2003 are now equal to one another – both an unremarkable 4. Not 5 in 2002 versus an incredible 7 in 2003. The split between categories in the two periods of winter months is markedly different. In winter 2002 – 2003 it is normal, spread out over all three, but mostly *cardio-respiratory*. In winter 2003 – 2004 *almost everything* is being categorised as *respiratory*. The total number of cases in January, in both winters, is much less than in adjacent months, this is normal.

Normal case-mix (for the three categories of interest), both in this hospital and in all others (we have similar data from about 40 other hospitals all over England, for the thirteen year period 2000 – 2012), is a mix mainly of *cardio-respiratory*, with *respiratory* and *hypoglycaemic* normally each at roughly a fifth of the level of *cardio-respiratory* (Figure 4). They are both much less usual, but neither can be called *rare*.

There is also data in the official public enquiry held after Ben's conviction, held to find out why Ben wasn't caught earlier and to prevent such a tragedy from ever occurring again. "The number in December 2003 was six and this was only one more than in December 2002". Two

different numbers, yet again. The inquiry did suggest that the very large numbers of incidents while Ben was carrying out his attacks might have been expected anyway, due to the winter season, perhaps masking incidents caused by Ben. It did heavily criticise the Emergency department for poor record keeping when updating patient medical notes and poor registration of withdrawals of dangerous medications.

The allocated space for this article has now run out. In Part 2 I will show results from a new analysis of some of the data which we have on the Ben Geen case, which I think could be part of the key to getting him a fair re-trial.

REFERENCES

- Gill, R. D. (2014). Statistics, Causality and Bell's Theorem. *Statistical Science*, 29(4), 512–528. doi:10.1214/14-STS490.
- Gill, R. D., Groeneboom, P., & De Jong, P. (2018). Elementary Statistics on Trial (the case of Lucia de Berk). *Chance*, 31(4), 9–15.
- Hensen, B., Bernien, H., Dréau, A. et al. (2015). Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature*, 526, 682–686, doi:10.1038/nature15759.

RICHARD GILL is professor emeritus of Statistics at Leiden University. This article is mainly based on his 'exauguration' speech. Apart from other distinctions he is an Honorary Member of the VVSOR.
E-mail: gill@math.leidenuniv.nl

Buitenspel, meetfouten en millimetergemier

Met de invoering van de Video Assistant Referee ofwel de VAR leek het er voor alles en iedereen een stuk eerlijker op te worden in de voetballerij. Met name de doellijnteknologie werd met gejuich begroet. Nooit meer eindeloze discussies over doelpunten die al of niet terecht waren goedgekeurd. En ook over buitenspel geen twijfel meer, zo dacht men optimistisch. Maar hoe goed de VAR ook werkt voor de eerstgenoemde situatie, buitenspel en de VAR is een minder gelukkige combinatie gebleken.

De gruwel van lange spelonderbrekingen

Hoewel ook het gebruik van de VAR bij buitenspel met hoge verwachtingen werd geïntroduceerd roept zij voornamelijk onvrede op. Denk alleen maar aan het afgekeurde doelpunt van Quincy Promes in de recente thuiswedstrijd van Ajax tegen Chelsea. De onvrede wordt veroorzaakt door het minutenlang VAR-overleg bij vermeend buitenspel en de vaak uiterst omstreden beslissing die daarop volgt. Het oog van de scheidsrechter blijkt niet altijd minder onrechtvaardig dan het oog van de computer, de hoge verwachtingen omtrent de VAR als snelle en eerlijke hulpscheidsrechter ten spijt.



Quincy Promes buitenspel?

Met de VAR kan worden gemeten, heel nauwkeurig tot op de millimeter. Maar helaas, meten is lang niet altijd zeker weten! Meten is weten met een foutmarge! En als de VAR dus aan het tijdrovende millimetermieren is over wel of niet buitenspel, dan is de terechte vraag of het gebruikte computerbeeld van de VAR wel de realiteit weergeeft en de beslissing dus wel correct en eerlijk is. Natuurlijk klopt het dat je met elkaar kunt afspreken dat de VAR-beslissing DE beslissing is en we ons niet bekommeren over onnauwkeurigheden van de metingen. Maar de terechte vraag is dan of we dat de sporters kunnen aandoen. Een foute beslissing van de scheidsrechter (denk aan de iconische handsbal van Thierry Henry), en dus ook van de VAR, heeft soms miljoenen euro-consequenties! Ook andere sporten dan voetbal kennen dit nijpende probleem dat alleen maar groter wordt omdat de prestatiedichtheid aan de top toeneemt en de onderlinge verschillen tussen de toppers steeds kleiner worden.



Zichtbaar teleurgestelde Koen Verweij mist goud op 3-duizendste seconde

Was het eerlijk dat Koen Verweij in Sotchi zilver kreeg terwijl het verschil met de nummer één slechts drieduizendste van een seconde was, een verschil zo klein dat die binnen de erkende foutmarges van de (tijd)metingsystemen lag? Zelfs de officiële tijdwaarnemers kennen die onnauwkeurigheidsmarges. Saillant punt hierbij was

dat beide atleten niet in dezelfde rit reden, zodat er geen finishfoto beschikbaar was en niet met een loop op de millimeter nauwkeurig gekeken kon worden. In Sochi zat het hem in de onnauwkeurigheidsmarge van het tijdregistratiesysteem. Millisecondenwerk, binnen de foutmarge. Onterecht zilver dus voor Koen. *Ex aequo* goud was de juiste uitslag geweest.



Marcel Kittel winnaar?

En hoe zat het ook al weer met de zevende etappe in de Tour de France van 2017? Een heldere finishfoto met een kaarsrechte finishlijn was hier beschikbaar. Maar pas na minutenlang overleg werd Marcel Kittel uitgeroepen tot winnaar met een nauwelijks waarneembaar verschil met de nummer twee, Edvald Boasson Hagen. Erger nog, de lijn op de foto is niet de echte finishlijn die op het asfalt is gekalkt. De lijn is ingetekend in de finishfoto en kan dus nooit voor 100% zekerheid exact op de plek staan van de finishlijn op de weg. Er is dus sprake van een foutmarge en die kan net groot genoeg zijn geweest om niet Kittel maar Boasson Hagen tot winnaar uit te roepen. Gemier op de millimeter dus en een typisch voorbeeld van een *ex-aequo*-situatie met twee winnaars.

Hoe zit het met de onnauwkeurigheid van de VAR-lijn?

Volgens de FIFA-spelregels is de buitenspellijn een rech-

te lijn evenwijdig aan de doellijn van de ‘verdedigende’ keeper (eigenlijk het vlak verticaal op deze lijn); de grootte van de zone tussen de buitenspellijn en die doellijn wordt bepaald door de voorlaatste speler van het verdedigende team, die dus zelf net niet in de buitenspelzone staat (alleen een deel van zijn hand of arm mag volgens de regels over de grens zijn). Op het moment dat de VAR zich gaat bemoeien met buitenspel, dan tekent de VAR de bekende lijn in het computerbeeld, net zo’n lijn als bij Kittel en Boasson Hagen. Een lijn dus die niemand ziet ‘in het echt’, het is geen vaste kalklijn op het veld. Volgens de theorie en de praktijk van meten kan de positie van die lijn nooit 100% nauwkeurig samenvallen met de theoretische buitenspellijn volgens de FIFA-regels. Weliswaar zal de fout klein zijn, maar wel groot genoeg om flink de plank mis te slaan, zoals we hebben gezien. Het probleem ontstaat als de bal ‘naar voren’ wordt gespeeld naar een teamgenoot, die daardoor wordt betrokken in het spel en mogelijk strafbaar buitenspel staat. Van belang is dan het tijdstip dat de bal wordt aangespeeld. Precies op dat tijdstip zou de VAR-lijn in het computerbeeld moeten worden getekend. Als het gaat om millimetertjes (en de mogelijk miljoenen euro’s) beginnen de bekende lange wachttijden en slaat bij de VAR-scheidsrechters de onzekerheid toe. En begrijpelijk. Waarom? Allereerst is het tijdstip van aanspelen van de bal sowieso niet met 100% nauwkeurigheid te bepalen: in de praktijk hebben we te maken met een (weliswaar heel klein) tijdsverschil. Verder moet volgens de FIFA de VAR-lijn ‘strak’ aanliggen tegen de ‘voorlaatste’ speler. Maar helaas, die speler beweegt meestal in het tijdsvenstertje en waar precies zijn arm overgaat in de schouder is al evenmin exact te bepalen. Verder is het onmogelijk om de VAR-lijn precies evenwijdig te trekken aan de doellijn van de tegenstander. Conclusie: de VAR-lijn valt niet met 100% zekerheid samen met de theoretische FIFA-buitenspellijn.

Schijneerlijkheid met de VAR

Zodra de VAR-lijn is ingetekend wordt bepaald of de bewuste (armloze) speler inderdaad strafbaar buitenspel staat. En dan kan het gaan om millimeters, zoals onlangs bij de Ajacied Promes. De daaropvolgende minutenlange discussie en onderbreking van het spel is echter verspilde tijd en geeft slechts schijneerlijkheid. Het is zoeken naar een speld in een hooiberg, omdat dat millimetertje buitenspel veel kleiner is dan de foutmarge op de positie van

de ingetekende VAR-lijn. En ook als het over millimeters gaat kun je niet volstaan met de simpele vaststelling dat de VAR-lijn is-zoals-hij-is en de inherente onnauwkeurigheden worden genegeerd. Een foute referee-beoordeling van strafbaar buitenspel (volgens de FIFA-regels dus) kan grote gevolgen hebben voor het verloop van de wedstrijd, zeker ook als er weinig gescoord wordt, zoals bij EK’s en WK’s waar meer dan een derde van de wedstrijden eindigt met maximaal slecht één doelpunt.

De oplossing: tolerantiezones

De onzekerheidsmarge van de VAR-lijn kan eenvoudig in beeld worden gebracht met het aanbrengen van een zogenaamde tolerantiezone tegen de VAR-lijn aan in de richting van de goal van de tegenstander. Binnen die tolerantiezone is er dan geen sprake van strafbaar buitenspel. Het ligt voor de hand om de breedte van de tolerantiezone te laten bepalen door de grootte van de onnauwkeurigheid op de positie van de VAR-lijn. Die is evenwel wat lastig uit te rekenen en is ook afhankelijk van de meetapparatuur. Beter is het om een vaste breedte te kiezen die duidelijk en zichtbaar is: 30 cm moet volstaan. De scheidsrechters zien dan op het scherm de eigen VAR-lijn plus daar tegenaan de, zeg grijze, tolerantiezone van 30 cm breed.

Maar is daarmee de discussie niet 30 cm opgeschoven?

Nee, en dat zit zo. In de buurt van de VAR-lijn is het, zoals we zagen, volstrekt onzeker of er sprake is van strafbaar buitenspel. Maar omdat de tolerantielijn (de grens van de tolerantiezone richting de doellijn) zo is gekozen dat die ver genoeg verwijderd is van de VAR-lijn, is zelfs in de buurt ervan zeker sprake van strafbaar buitenspel. Dus als de betreffende speler geheel of gedeeltelijk de tolerantielijn is gepasseerd, dan geen VAR-discussies: onmiddellijk strafbaar buitenspel geven. Maar wat als het voor de VAR-scheidsrechter(s) toch niet zo duidelijk is dat de tolerantielijn (gedeeltelijk) is gepasseerd? Ook dan dus geen discussie, de tolerantielijn ligt daarvoor te ver af van de VAR-lijn: zonder discussie onmiddellijk strafbaar buitenspel geven. Op deze wijze is de buitenspelonzekerheid ‘op’ de VAR-lijn vervangen door zekerheid ‘op’ de tolerantielijn.

Een metafoor

Nogmaals, met de tolerantiezone naast de VAR-lijn wordt recht gedaan aan het feit dat in de buurt van de huidige VAR-lijn het strafbaar buitenspel staan niet zinvol en eerlijk is te bepalen, omdat de VAR-lijn niet precies samenvalt met de FIFA-buitenspellijn. Je zou dit kunnen vergelijken met de tolerantie bij het overtreden van, zeg, de 50 km/u snelheidsbeperking in de bebouwde kom. De ‘harde’ grens ligt bij 50 km/u: sneller dan 50 is fout. Maar vanwege de meetonzekerheid worden gemeten snelheden binnen de 57 km/u getolereerd en niet bestraft. Het grote verschil met voetbal is natuurlijk wel dat spelers op het veld niet kunnen anticiperen op die tolerantie wat automobilisten wel kunnen met hun snelheidsmeter.

Is de VAR de voorloper van de robotscheids?

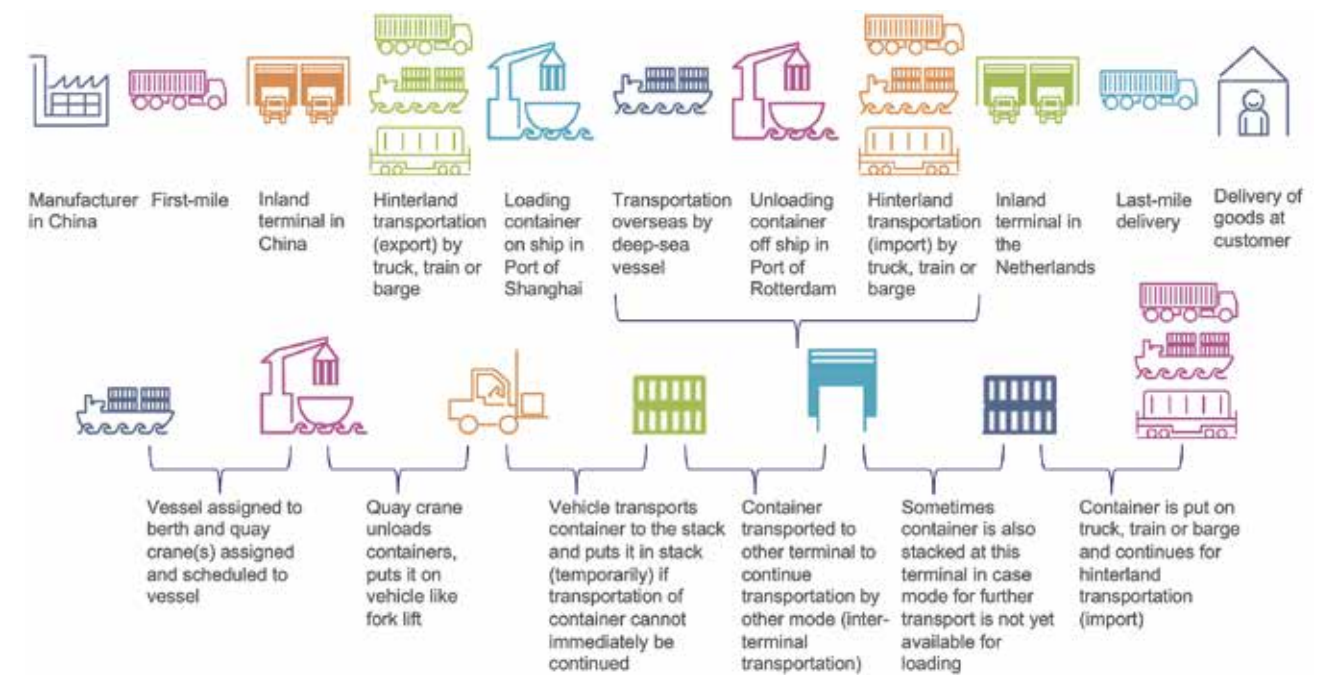
De meetapparatuur wordt steeds nauwkeuriger en we kunnen ons afvragen of met de komst van VAR de robotscheidsrechter zijn intrede heeft gedaan in de sport. Zeker is het dat strafbaar buitenspel ooit nano-nauwkeurig is vast te stellen. Maar of dat de VAR-tolerantiezone bij buitenspel overbodig maakt valt zeer te betwijfelen. Geen enkele sport zit te wachten op beslissingen die alleen met de computer, een robot, genomen worden en ‘onzichtbaar’ zijn voor het publiek. Het is van cruciaal belang dat wij, de toeschouwers, beelden zien van de redenen van computerbeslissingen. Op de stadionschermen, de smartphones en de televisieschermen moeten we deze beslissingen *real time* kunnen meeleven. Het in beeld brengen van de tolerantiezone is daarvan een mooi voorbeeld. De tonnen voor millimetergemier met nano-nauwkeurige robotapparatuur kunnen beter besteedt worden aan de optimalisatie van de communicatie met de fans. Als toeschouwer wil je zien, niet geloven.

P.S. Een opmerkelijke lezer van een vorige versie van dit verhaal, dat *mutatis mutandis* in januari dit jaar verschenen is in *SportknowhowXL*, wees mij erop dat de zo te noemen buitenspel-VARlijn door mij in het televisie-interview op 19 januari jl. in Studio Voetbal de docterbibberlijn werd genoemd (over de precieze schrijfwijze van dit woord zijn de meningen trouwens verdeeld) en dat dit kennelijke neologisme een dag later in Lexiton Taalbank is uitgeroepen tot ‘Woord van de Dag’. Waarvan acte.

GERARD SIERKSMA is emeritus hoogleraar Kwantitatieve Logistiek en Sportstatistiek aan de Rijksuniversiteit Groningen. E-mail: g.sierksma@rug.nl



Foto: Alexander Kilem via Pixabay



Figuur 1. Overzicht van de verschillende processen in de container supply chain en hoe deze processen met elkaar samenhangen

Impact van synchromodale logistiek op achterland transport

MARIJKE LIJZENGA

Synchromodaliteit is een concept dat de laatste jaren steeds meer aandacht heeft gekregen door ontwikkelingen binnen de containerindustrie. Om een beter inzicht te krijgen in de rol die het kan spelen in de transportwereld van de toekomst, hebben we geanalyseerd wat de mogelijke impact is van synchromodale logistiek op transportkosten, CO₂-uitstoot en vertragingen.

Synchromodale logistiek focust op het optimaal toekennen van containers aan verschillende transportmodaliteiten (truck, trein en vrachtschip) waarbij het mogelijk is om real-time aanpassingen te doen als verstoringen plaatsvinden. Doordat het aantal containers dat wereldwijd vervoerd wordt de laatste jaren significant is toegenomen, wordt het steeds belangrijker dat de connectie tussen de haven en het achterland goed georganiseerd is. Daarnaast neemt de capaciteit van

containerschepen steeds meer toe. Vandaag de dag heeft het grootste containerschip een capaciteit van maar liefst 23.756 TEU, een standaardmaat die gebruikt wordt in de containerindustrie om de grootte van een container aan te duiden. Als gevolg van deze steeds groter wordende containerschepen, neemt de druk op havens toe. Door de toenemende concurrentie tussen havens die hetzelfde achterland kunnen bedienen, zijn havens gedwongen om hun infrastructuur hierop aan te passen en daarbij dus ook rekening te houden met de toegang tot het achterland. Dit is waar synchromodaliteit een belangrijke rol kan spelen. Als alle beschikbare modaliteiten optimaal gebruikt worden, kan dat resulteren in een betere verbinding tussen de haven en het achterland, wat de haven aantrekkelijker maakt voor scheepvaart.

Complexiteit synchromodale logistiek

Het is echter behoorlijk complex om de modaliteiten optimaal te gebruiken, gezien veel processen hierbij op elkaar afgestemd moeten worden. Zodra een container aankomt in de haven, moet het gelost worden van het schip, waarna die vaak tijdelijk wordt opgeslagen in de haven. Daarna moet de container op het haventerrein verplaatst worden naar de terminal vanuit waar het mogelijk is om de container naar het achterland te vervoeren. Idealiter zijn al deze processen perfect op elkaar afgestemd, maar dat blijkt in de praktijk vaak een grote uitdaging.

Zoals in figuur 1 te zien is, hangen al deze processen tot op zekere hoogte van elkaar af. Een containerschip dat zijn tijdslot mist, een kraan die vertraagd is, een trein die vertraagd is, het zijn allemaal processen die een significante invloed hebben op de daaropvolgende processen. Goede communicatie en het delen van informatie is cruciaal om dit goed te laten verlopen, maar dat is niet vanzelfsprekend. Vanwege de complexiteit om alle processen mee te nemen en de verschillende partijen die daarbij betrokken zijn, hebben we voor dit onderzoek de keten gelimiteerd tot het achterland transport, waarbij aangenomen wordt dat alle processen in de haven vast staan. Ook wordt de zogenoemde *last-mile delivery*, het vervoer van en naar de eindgebruiker, niet meegenomen.

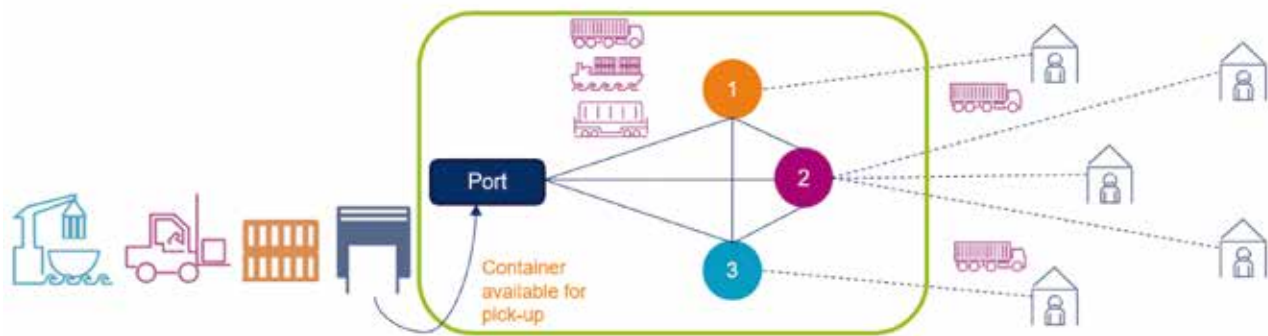
Rekening houdend met deze afbakening, hebben we twee verschillende synchromodale planningsmethodes gedefinieerd om de mogelijke impact van synchromoda-

liteit te bepalen. In deze methodes bekijken we een logistieke serviceprovider (LSP) die orders ontvangt van containers die vervoerd moeten worden. Omdat we de *last-mile delivery* niet meenemen, gaat al het containervervoer dat we bekijken tussen de haven en een binnenlandse terminal of vice versa (zie ook figuur 2). De taak van de LSP is om de orders zo goed mogelijk toe te kennen aan de transportmodaliteiten, daarbij rekening houdend met onder andere capaciteit en leveringstermijnen. Daarnaast wordt de planning geüpdatet in geval van verstoringen. Mocht een trein bijvoorbeeld vertraagd zijn, dan kan het beter zijn om een container met een truck te sturen in plaats van te wachten op de vertraagde trein. Deze mogelijkheid om te 'reschedulen' is er vanaf elke terminal.

De optimale allocatie van een order wordt voor alle containers in de order los bepaald. Dat betekent dat we voor alle containers bekijken wat de combinatie van transportmodaliteiten is die de gewogen som van transportkosten, vertragingen en CO₂-uitstoot minimaliseert.

Synchromodale planningsmethodes

Voordat de containers worden gealloceerd aan transportmodaliteiten, reduceren we het netwerk voor alle combinaties van start- en eindpunten. Dit houdt in dat de routes met een te grote omweg worden verwijderd. Het is bijvoorbeeld niet rendabel om van Rotterdam via Leeuwarden naar Maastricht te gaan. Voor deze reductie gebruiken we een factor α . Het netwerk wordt geli-



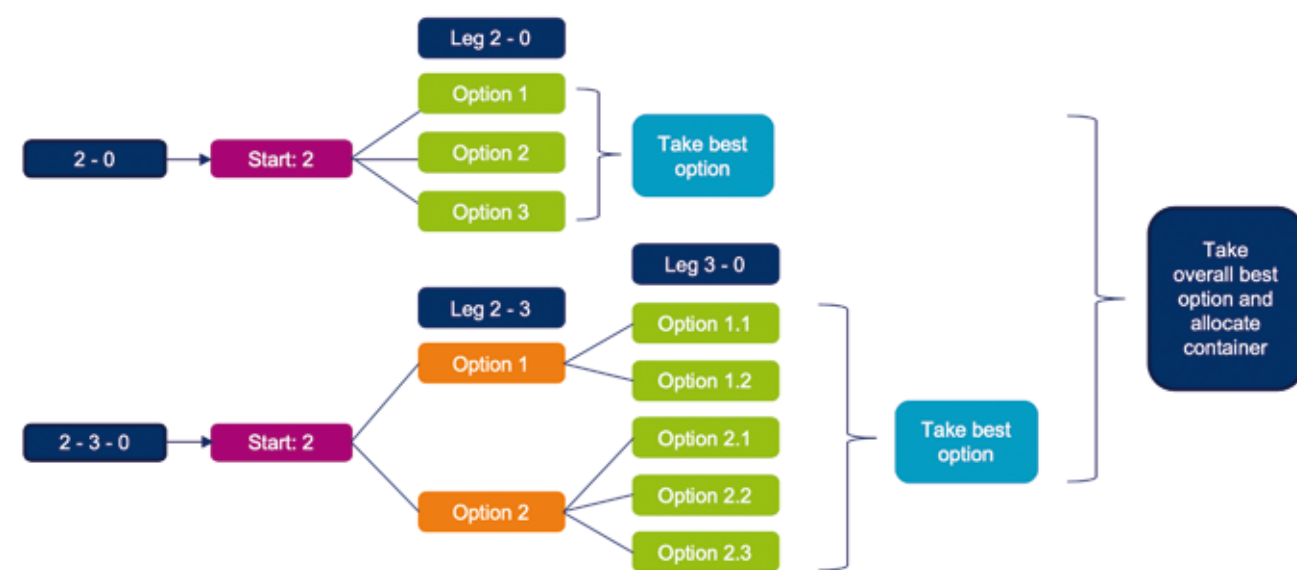
Figuur 2. Het onderzoek focust zich op het groen omlijnde stuk, dat wil zeggen alleen het containervervoer tussen de haven en achterlandterminals wordt meegenomen

miteerd tot de routes die binnen een afstand van factor α vermenigvuldigd met de directe afstand van startpunt naar eindpunt vallen. Dit is afgeleid van de planningsmethode geïntroduceerd in het onderzoek van Mes en Iacob (2016). Het gereduceerde netwerk wordt vervolgens als input genomen voor de planningsmethodes.

Voor de eerste planningsmethode hebben we volledige enumeratie gebruikt. In deze methode wordt een inkomende order toegewezen aan één of meerdere transportmodaliteiten zodra de LSP de order ontvangt. Eerst worden alle mogelijke transportopties gevonden op alle mogelijke routes tussen start- en eindpunt van de order. Deze mogelijke routes volgen uit de netwerk reductie. Om te bepalen of een combinatie van modaliteiten een optie is om een container te vervoeren, wordt er onder andere gekeken naar de beschikbare capaciteit en de

eventuele vertraging die het oplevert. Zodra we alle mogelijke transportopties gevonden hebben, bekijken we welke optie de beste waarde geeft. Deze waarde wordt bepaald aan de hand van de gewogen som van transportkosten, CO₂-uitstoot en vertraging die de optie oplevert. De optie met de beste waarde is de optie waar de container aan toegekend wordt. Deze optie heeft dus de laagste gewogen som, waarbij de gewichten van tevoren bepaald worden. In figuur 3 is een schematisch overzicht van deze methode weergegeven.

Naast bovenstaande methode hebben we een methode bestudeerd waarbij historische data worden meegenomen. Als een order al een jaar lang iedere maandag binnenkomt, dan is de kans groot dat dit komende maandag ook het geval zal zijn. Aan de hand van simulaties van de toekomst gebaseerd op historische data, bepalen



Figuur 3. Overzicht van de werking van het eerste planningsalgoritme. Voor alle mogelijke paden tussen terminal 2 en terminal 0 (de haven) worden alle transportopties bekeken, om vervolgens de container aan de beste optie te alloceren

we waar een container het best gealloceerd kan worden. Om de invloed van bepaalde orders te beperken, simuleren we verschillende scenario's van de toekomst. Evenals in de eerste planningsmethode, vinden we ook hier alle mogelijke transportopties. De transportoptie die gekeken naar alle scenario's samen gemiddeld het beste resultaat geeft, is de optie waaraan gealloceerd wordt. Dit wordt wederom bepaald aan de hand van de gewogen som van transportkosten, CO₂-uitstoot en vertragingen.

Resultaten

Om de mogelijke impact van synchromodaliteit te bepalen, hebben we bovenstaande methodes getest op een gesimuleerde dataset. Voor de data waarop we getest hebben, bleek synchromodaliteit veel potentie te hebben. Er was een duidelijke verschuiving in modaliteiten te zien, waarbij veel minder gebruik werd gemaakt van trucks en meer van treinen en vrachtschepen. Truck-transport is een flexibele transportoptie waar vandaag de dag veel gebruik van wordt gemaakt. Echter, door de toename in containervervoer wordt de belasting op wegen groter, wat een negatieve impact heeft op milieu en maatschappij. Door bij synchromodale logistiek de mogelijkheid te hebben om te optimaliseren over verschillende modaliteiten en real-time de planning aan te passen, wordt het aantrekkelijker om de trein of vrachtschip te gebruiken voor containervervoer. Dit zagen we in de resultaten duidelijk naar voren komen.

Sensitiviteit

Bij bovengenoemd resultaat moet echter wel een kanttekening geplaatst worden. Om data te simuleren, moesten veel aannames gemaakt worden en tijdens de sensitiviteitsanalyse bleek een aantal aannames significante invloed te hebben op het resultaat, wat van tevoren ook te verwachten was. Doordat we geen data beschikbaar hadden, moesten we onder andere tijdschema's, beschikbare modaliteiten, trein- en waterwegverbindingen definiëren. De beschikbaarheid van modaliteiten en verbindingen heeft veel invloed op de mogelijkheden om een container met een bepaalde modaliteit te vervoeren. Echter, om een zo realistisch mogelijk beeld te krijgen, hebben we onze aannames in overleg met verscheidene experts op dit gebied gemaakt en ook vergeleken met aannames in eerdere onderzoeken, onder andere de onderzoeken van Van Riessen (2013) en Van Riessen (2018). Op deze ma-

nier kunnen we toch een beeld krijgen van de impact van synchromodaliteit.

In beide planningsmethodes wordt de gewogen som van transportkosten, CO₂-uitstoot en vertragingen als maatstaf genomen om te bepalen aan welke modaliteiten een container toegekend wordt. Hierbij geven verschillende gewichten een verschillende uitkomst. Hoe minder vertragingen zijn toegestaan en hoe hoger daarvoor de penalty is, hoe meer truck transport gebruikt zal worden, gezien trucks flexibeler inzetbaar zijn dan treinen en vrachtschepen. Daarentegen, wanneer de focus op CO₂-uitstoot ligt, zien we dat treinen zo vaak mogelijk ingezet worden. Dit is iets wat in de praktijk belangrijk is om mee te nemen. Op welk element of op welke elementen de focus ligt, heeft een grote invloed op de mate waarin synchromodaliteit succesvol kan zijn.

Conclusie

Al met al kunnen we concluderen dat er in de toekomst zeker mogelijkheden liggen voor synchromodaliteit. Echter, hierbij is het wel van cruciaal belang dat de verschillende partijen in de keten goed met elkaar communiceren en dat data gedeeld wordt. Alleen op die manier is het mogelijk om real-time updates te doen en de keten zo optimaal mogelijk te maken. Daarbij komt ook dat transportplanners en mensen die een container willen vervoeren ervan overtuigd moeten zijn dat synchromodaliteit inderdaad voordelen biedt. Dit blijkt vandaag de dag nog een grote uitdaging, onder andere doordat synchromodale logistiek vaak een duurdere transport optie is dan truck transport. Vanwege de complexiteit van de hele containerketen, zal het nog wel enige tijd duren voordat synchromodaliteit optimaal toegepast kan worden, maar er liggen zeker mogelijkheden op dit gebied.

LITERATUUR

- Mes, M. R., & Iacob, M. E. (2016). Synchromodal transport planning at a logistics service provider. *Logistics and Supply Chain Innovation*, 23–36.
- Van Riessen, B. (2013). *Planning of hinterland transportation in the EGS network*. Erasmus University.
- Van Riessen, B. (2018). *Optimal Transportation Plans and Portfolios for Synchromodal Container Networks*. Erasmus Research Institute of Management.

MARIJKE LIJZENGA is werkzaam als Analytics Consultant bij ORTEC. Dit onderzoek heeft zij gedaan voor haar afstudeerscriptie voor de master Econometrics and Operations Research aan Maastricht University, waarvoor zij stage liep bij ORTEC. E-mail: marijkelijzenga@gmail.com

Charly Chaplin in *Modern Times* (1936)

DE MACHINE STOPT

Amsterdam, 25 maart 2027. De rechtbank van Amsterdam heeft een 41-jarige arts van het Academisch Medisch Centrum in Amsterdam schuldig bevonden aan dood door schuld door het negeren van de aanbevelingen van Dr Daneeka¹, een op kunstmatige intelligentie gebaseerd diagnose en behandelplan-adviesstelsel dat sinds een jaar bij het AMC in gebruik is. De nabestaanden van de patiënt hebben de zaak aanhangig gemaakt nadat bekend werd dat de arts een andere behandeling voorschreef dan door het algoritme werd aanbevolen. De tegenwerping van de arts dat de adviezen van het algoritme vaker fout dan goed waren werden door de rechtbank als irrelevant afgedaan. De arts is in voorlopige hechtenis genomen in afwachting van de uitspraak van de Strafmaatbepaler, een algoritme dat al sinds 2022 met succes wordt gebruikt door het Openbaar Ministerie om de werkdruk te verlagen. Het Medisch tuchtcollege heeft de arts direct uit het BIG-register geschrapt en de eerder afgegeven richtlijnen voor het gebruik van algoritmisch bepaalde behandelplannen bekrachtigd.

Is bovenstaande schets een absurd scenario? Wellicht, feit is wel dat we steeds meer van onze beslissingen overlaten aan algoritmes en daarmee kennelijk minder waarde toekennen aan het zelf maken van een afweging. De inzet van algoritmes verandert daarmee van beslissingsondersteuning naar autonome besluitvorming. Ongecontroleerde groei van deze afhankelijkheid zou als extreme uitkomst een soort beslissings-singulariteit kunnen opleveren waarin het slaafs volgen van de uitkomst van een algoritme het vermogen van mensen om zelfstandig beslissingen te kunnen nemen volledig vernietigt.

Hoe groot is de invloed van algoritmes op onze besluiten eigenlijk? Het is algemeen bekend dat algoritmes gebruikt worden in internet search, dating, kredietverstrekking, recruitment en voorwaardelijke invrijheidsstellingen. In China gaat men een stap verder, daar wordt

met big data en algoritmes een sociaal kredietsysteem geïmplementeerd en daar stopt het niet. Volgens Human Rights Watch² zouden in 2018 in de regio Xinjiang de eerste burgers opgepakt zijn nadat algoritmes hen preventief hadden aangeduid als staatsgevaarlijk. In China is het kennelijk al 2054³. In zijn boek *A Human's Guide to Machine Intelligence: How Algorithms Are Shaping Our Lives and How We Can Stay in Control*, beschrijft professor Kartik Hosanagar⁴ van Wharton University op welke manier algoritmes invloed hebben op beslissingen die door ons of over ons worden genomen. Vaak gebruiken we bewust algoritmes omdat het ons tijd bespaart, zoals bijvoorbeeld in internet search. Door PageRank biedt Google ons de meest relevante antwoorden op onze vragen. Als we echter te passief worden en de antwoorden van algoritmes kritiekloos overnemen kan dat vervelende gevolgen hebben. Als oplossing voor dit probleem adviseert Hosanagar dat we ons veel bewuster moeten zijn van de mogelijke gevolgen van het gebruik van algoritmes.

Een algoritme is objectief. Het streeft naar een concreet meetbaar doel waarbij voldaan moet worden aan vooraf bepaalde randvoorwaarden. Vaak leidt een algoritme tot een enkelvoudig resultaat (u weet wel 42), waarbij de mogelijkheid om te kunnen kiezen op voorhand lijkt te worden uitgesloten. Ondanks dat de algoritmes eigenlijk niet meer opleveren dan een voorstel voor een besluit (zoals bijvoorbeeld de aanbeveling op Netflix of Amazon) worden ze door velen van ons toch als een beslissing ervaren. Dat komt omdat we algoritmes als superieur beschouwen (wie wil er nu tegen DeepMind een partijtje Go of Schaak spelen?) of omdat het alternatief om zelf het vraagstuk op te lossen ons te veel moeite kost. Natuurlijk kan een algoritme toegevoegde waarde creëren en kan het ons veel (en onaantrekkelijk) werk uit handen nemen. Keerzijde is dat we de controle op de ontwikkeling en toepassing van algoritmes verliezen waarbij uiteindelijk een situatie kan ontstaan waarin het algoritme gewoon

blindelings wordt gevolgd zonder te begrijpen wat het algoritme precies doet of wat je moet doen om het algoritme bij te sturen en te overrulen. Dat gaat soms mis, Death by GPS⁵ of autonoom rijdende Tesla's die op stilstaande hulpvoertuigen knallen⁶ illustreren dat. En ander voorbeeld zijn de gecrashte Boeing 737 Max vliegtuigen, de piloten begrepen het systeem onvoldoende en konden als gevolg daarvan niet tijdig bijsturen.

Ik las recent *The Machine Stops*⁷, een essay van Edward Morgan Forster⁸, bekend van onder andere *A Room with a View* en *Howards End*. Hij schreef het essay in 1909 met een vooruitziende blik, zo beschrijft hij onder andere het internet en videobellen. In het essay stelt hij zich een post-apocalyptische wereld voor waarin verzwakte individuen ondergronds leven in een totalitair, technologie-afhankelijk systeem. Ze vertrouwen op de 'Machine' voor communicatie, amusement en spiritueel en fysiek onderhoud. De mensen zijn zo afhankelijk geworden van de Machine dat ze het zijn gaan aanbidden als een levend wezen. Het beslissingsvermogen van de meeste mensen is erg verzwakt. Degenen die rebelleren worden verbannen naar het aardoppervlak. Echter, wanneer de Machine catastrofaal faalt, stort de ondergrondse samenleving in. De strekking van het verhaal is dat een te grote afhankelijkheid van technologie uiteindelijk zal leiden tot een samenleving die geen kans heeft om te overleven.

Dat algoritmes een nog grotere rol gaan spelen in besluitvorming is onvermijdbaar, het zal ons vele voordelen bieden. De gevaren zoals hierboven beschreven zijn vermijdbaar zolang we menselijke intelligentie op een intelligente manier in de loop houden bij de ontwikkeling en toepassingen van algoritmes. Dat vraagt om *awareness* en educatie, niet alleen voor de algoritmebouwers maar ook voor degene waarover of waarvoor het algoritme 'besluit'. Als we te veel beslissingen aan algoritmes delegeren, verliezen we het vermogen om te innoveren en belanden we wellicht in een wereld zoals Forster die beschrijft.

Ik wil graag benadrukken dat deze column zonder hulp van een algoritme tot stand is gekomen.

1. naar Doc Daneeka, uit Joseph Hellers *Catch-22*
2. <http://bit.ly/STAtOR37yV8si>
3. https://nl.wikipedia.org/wiki/Minority_Report
4. <https://oid.wharton.upenn.edu/profile/kartikh/>
5. <http://bit.ly/STAtOR2OY1yLg>
6. <http://bit.ly/STAtOR38IFD2j>
7. Gratis ebook via <http://bit.ly/STAtOR37CmE8v>
8. https://en.wikipedia.org/wiki/E._M._Forster

JOHN POPPELAARS, Directeur LITIC B.V.
E-mail: john.poppelaars@litic.com



Good news from
Young Statisticians

Dear statistics enthusiasts,

In the new decade, the **Young Statisticians** are still organizing events for young (and of course older) statistics enthusiasts throughout the Netherlands.

... We are currently busy with organizing visits to several interesting companies to discover their ways to use statistics and data science in daily practice.

... Furthermore, we are organizing a Statistics Café (symposium) about a relevant and interesting statistical topic where everyone can learn from renowned speakers and meet other statistics enthusiasts in an informal manner between the talks.

... Finally, we are present at the VVSOR Annual Meeting where we host a pub quiz at the evening.

Looking forward to a great year with many interesting events. We hope to see you all at our events. For information about the **Young Statisticians** and our events, visit the website <https://www.vvsor.nl/young-statisticians/> where you can also subscribe for our newsletter.

World Statistics Congress 2021

De voorbereidingen voor het WSC 2021 zijn in volle gang. Op de website isi2021.org staat een oproep om voorstellen voor Invited Paper Sessions in te dienen. Inmiddels is ook de voorzitter van het Scientific Program Committee bekend, dat is Nalini Ravishanker. Begin februari was er in Den Haag een vergadering van het National Organizing Committee, de foto toont alle aanwezigen.



Van links naar rechts: Geert Bruinooge (Penningmeester ISI WSC 2021 Foundation), Lisette de Ruiter-Schröter (CBS), Sofia Kapnisi (ISI), Henriëtte de Jong-de Heer (CBS), Carina Franssen (CBS), Roeland Beerten (Statistics Flanders), Ada van Krimpen (ISI), Kees Zeelenberg (ISI), John Bailer (Co-Chair/ ISI President), Paul van der Laan (ISI), Shabani Mehta (ISI), Marjolein Verweij (Congress by Design), Eric Schulte Nordholt (Co-Chair/CBS), Nynke Krol (CBS en voormalig voorzitter van de VVS YS), Maarten Kampert (VVS).

