STAtOR

thema **CORONAVRAAGSTUKKEN**

periodiek van de VVSOR jaargang 22, nummer 4, december 2021

STAtOR

Jaargang 22, nummer 4, december 2021

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

Redactie
Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Joep Burger, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stemerdink (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), Universiteit van Amsterdam Faculteit Economie en Bedrijfskunde, Sectie Operations Management | Amsterdam Business School, Plantage Muidergracht 12, 1018 TV Amsterdam, j.a.s.gromicho@uva.nl

Bestuur van de VVSOR
Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: Pieter Jongsma MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Algemeen bestuurslid: Thomas Wise MSc, db@vvsor.nl; Webmaster: Eugenio Traini MSc: webmaster@vvsor.nl.
Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Daniel Oberski (Section Data Science); Marije Sluiskes MSc (Young Statisticians) dr. Sanne Willems (Section Statistics Communication).

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

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

Ontwerp en opmaak
Pharos, Nijmegen

Uitgever

ISSN 1567-3383

LNMB & NGB SEMINAR

OPERATIONS RESEARCH FOR SUSTAINABLE DEVELOPMENT

Conferentiecentrum De Werelt in Lunteren
19 januari 2022, 10.00 – 17.00
Meer informatie over het programma
op www.lnmb.nl/conferences

Ronald Does geridderd

Professor Ronald Does is op 21 oktober 2021 benoemd tot Ridder in de Orde van de Nederlandse Leeuw. Hij ontving deze onderscheiding tijdens zijn afscheid op de Business School van de UvA. Een prachtige onderscheiding voor een lange carrière die in het teken stond van baanbrekend onderzoek van bedrijfskundige statistiek en vooral van Lean Six Sigma.

First Young Statisticians event

The Young Statisticians-section organized its first (live!) event on Thursday September 23rd: a Statistics Café in Leiden, with interesting contributions from three great speakers. Marjan Sjerps told us all about forensic statistics, Erik-Jan van Kesteren advised us on do's and don'ts in multidisciplinary collaborations and Valentina Masarotto discussed how statistical techniques can be used to distinguish sounds in Mandarin. The maximum number of 40 sign-ups was quickly reached, which illustrates how everyone was just as excited about this event as we were. After the talks, there was time to socialize while enjoying a beer or soda. It was great to finally meet and get to know fellow young statisticians again.

GitHub-workshop & New Year's Statistics Café
The coming months, we plan to organize a GitHub-workshop and a New Year's Statistics Café. The dates are yet to be determined: if you want to stay up to date, make sure to subscribe to our newsletter (vvsor.nl/young-statisticians/ > Mailing list sign-up).

Pandemie

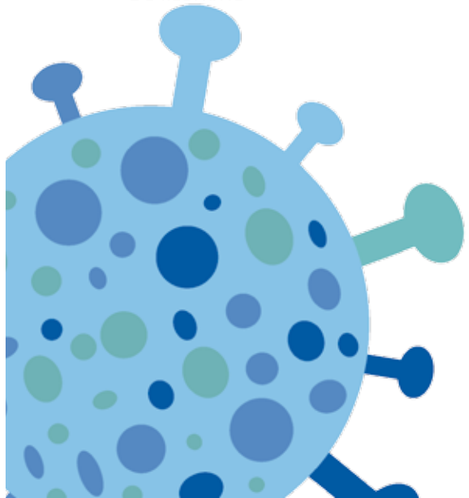
Vrijwel direct na de eerste coronagevallen, in februari en maart 2020, ontstond binnen de redactie het idee om vanuit de specifieke kennis in ons vakgebied ruim aandacht te besteden aan deze pandemie. De afgelopen anderhalf jaar verschenen er in *STAtOR* dan ook enkele gerelateerde artikelen.

We wilden meer, we dachten eerst aan een bescheiden themanummer, te verschijnen in de zomer van 2020, maar daar stapten we al snel van af. Gaandeweg kwamen er zóveel verschillende aspecten aan het licht dat we liever op een later tijdstip een uitgebreid nummer wilden brengen, met meer onderwerpen en met meer mogelijkheden tot retrospectie. Het leek ons goed om voor dit nummer gastredacteuren in te schakelen, deskundigen die vanuit hun netwerk een scala aan auteurs zouden kunnen benaderen.

Daartoe hebben we Gerrit Timmer en Casper Albers bereid gevonden. Zij hebben de Corona Gedragsunit van het RIVM geadviseerd en waren betrokken bij het Landelijk Coördinatiecentrum Patiëntenspreiding, corona-apps en coronagerelateerde projecten in ziekenhuizen en kennen daardoor het veld als geen ander. Hun inbreng heeft geresulteerd in een boeiend nummer, met maar liefst tien artikelen over corona. Wij willen hen van harte bedanken voor hun inspanningen. Zij geven hiernaast zelf een korte inleiding op dit nummer.

Ook bevat deze *STAtOR* de gebruikelijke elementen, zoals columns en mededelingen. Twee van deze mededelingen willen we hier noemen, een droevig en een blij bericht. We staan stil bij het overlijden van Constance van Eeden in een kort In Memoriam, in een volgend nummer zal zij uitgebreider worden herdacht. En we besteden aandacht aan de Huibregtsen-prijs die aan Sandjai Bhulai en Rob van der Mei is toegekend. Ons redactielid Caroline Jagtenberg heeft hen geïnterviewd.

De redactie



Coronavraagstukken

Eén onderwerp houdt de wereld nu al ruim anderhalf jaar in de houdgreep: corona. Met naar schatting wereldwijd inmiddels zo'n kwart miljard besmettingen en vijf miljoen sterfgevallen is de impact van deze pandemie ongekend.

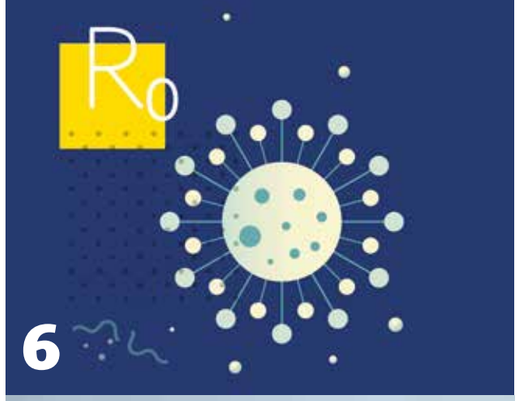
Naast al het leed bracht de pandemie ons ook een enorme berg aan data. Besmettingscijfers, het reproductiegetal, vaccinatiegraden, noem maar op. En doorgaans waren dit *messy data*: aan het begin van de pandemie was er enorme schaarste aan testkits. Wie niet getest is, wordt niet gemeten, dus de besmettingscijfers waren toen te laag. En zo zijn er tal van andere aspecten waardoor het niet gewoon mogelijk is een standaard tijdreeksmodel los te laten op de cijfers. Om echt te begrijpen wat de cijfers vertellen, zijn degelijke statistische analyses nodig. En als je dan relevante data hebt verkregen, hoe kunnen die je dan helpen om de beste beleids- en operationele keuzes te maken?

In dit nummer van *STAtOR* presenteren wij u maar liefst tien artikelen over coronadata. Vanuit verschillende disciplines, zoals medische statistiek, psychometrie en operations research, wordt een blik geworpen op de coronacijfers, en dan met name de situatie hier in Nederland. Hoe spreid je de patiënten het meest efficiënt over ziekenhuizen? Hoe modelleert het RIVM een nieuwe pandemie? Hoeveel en welke operaties zijn nog mogelijk gegeven het verwachte beslag van corona op de IC-capaciteit? Kunnen we aan de hand van de cijfers voorspellen wanneer er coronaprotesten plaatsvinden? Het antwoord op deze en zes andere coronavraagstukken vindt u in dit nummer. Wij danken alle auteurs hartelijk voor hun bijdragen.

Wij wensen u veel leesplezier.

Gerrit Timmer & Casper Albers
Gastredacteuren

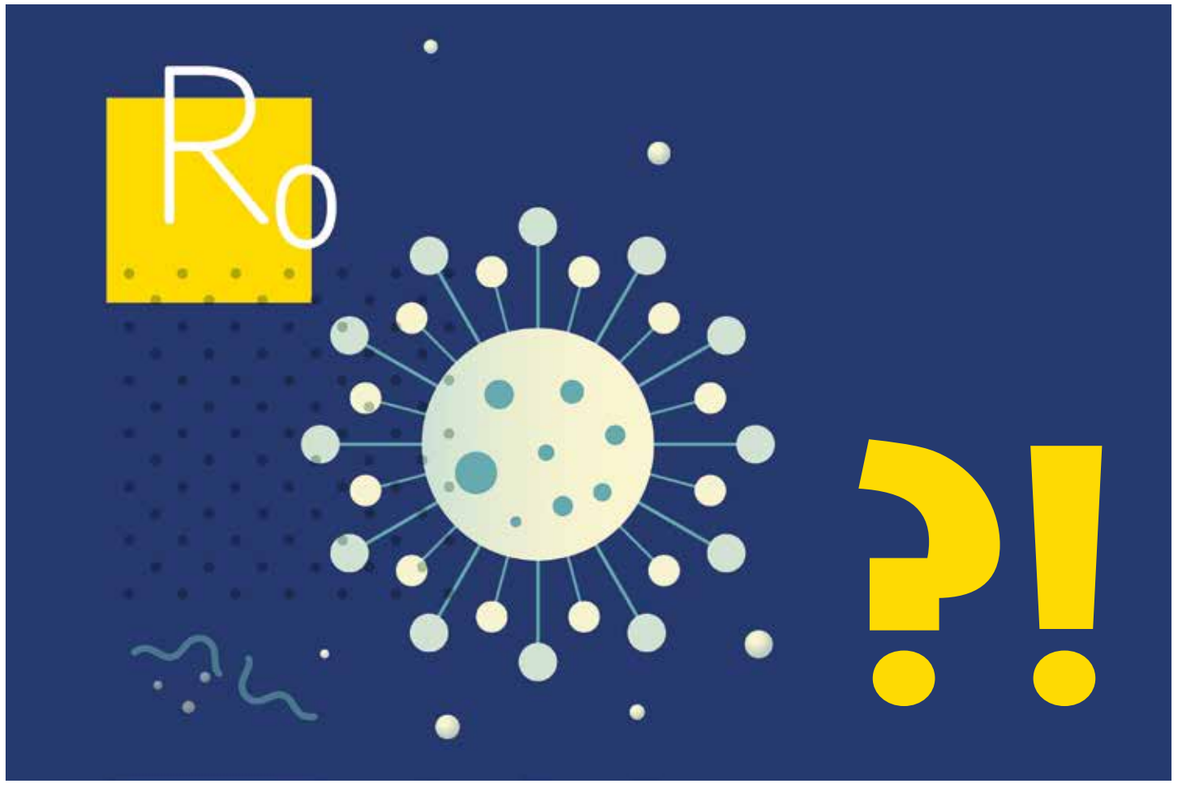
PS Dit nummer verschijnt eind december 2021. De artikelen in dit nummer zijn minimaal anderhalve maand geleden geschreven. De situatie rond de pandemie is in die tussentijd zorgwekkender geworden. Daardoor kan het voorkomen dat er zinsneden zijn die de auteur mogelijk iets anders zou hebben geformuleerd als het artikel pas op het ogenblik van verschijnen zou zijn geschreven. Denk niet dat dit nummer daarom achterhaald is, het tegendeel is waar. Vrijwel alle artikelen kijken namelijk ook naar toepassingen in de toekomst, rekening houdend met een scenario als dat waar we nu in zitten.



INHOUD

Het bestrijden van een nieuwe infectieziekte:
METEN, MODELLEREN, COMMUNICEREN

Een grote infectieziektepandemie met lege straten en overvolle ziekenhuizen komt zelden voor. Je kan erover lezen in boeken die een beschrijving geven van de pestepidemieën, de cholerapandemieën en de Spaanse griep.
En nu kunnen we het zelf meemaken. Vroeger werd de bestrijding zeer lokaal aangepakt, nu wordt de bestrijding in Nederland als vanzelfsprekend landelijk gecoördineerd. Vroeger werd het aantal sterfgevallen gedocumenteerd, nu worden allerlei statistieken bijgehouden, zoals aantallen infecties, aantallen positieve testen en ziekenhuisopnames om beter zicht op de epidemie te houden. Op basis van deze gegevens wordt nu berekend wat we van de bestrijding mogen verwachten.



Jacco Wallinga & Don Klinkenberg

In Nederland is het verzamelen van relevante gegevens voor infectieziekten een taak voor het Centrum voor Infectieziektebestrijding van het RIVM. Dat centrum is opgericht na de SARS-uitbraak in 2003. Het centrum omvat ook een afdeling modellering van infectieziekten die voor het rekenwerk zorgt. Bij deze afdeling worden methoden ontwikkeld voor het beschrijven, analyseren en voorspellen van de dynamiek van endemische infecties, en van kleine uitbraken en grote epidemieën. Deze methoden worden ingezet bij de advisering over de preventie en bestrijding van infecties. Tezamen met infectiologen, epidemiologen, immunologen, virologen, medisch microbiologen worden adviezen gegeven aan het ministerie voor Volksgezondheid, Welzijn en Sport (VWS), soms direct, soms via het Outbreak Management Team of de Gezondheidsraad. Via deze routes draagt het rekenen aan infectieziekten in Nederland bij aan de advisering voor infectieziektebestrijding.

Het werk aan COVID-19 bij het RIVM begon halverwege januari 2020. In Wuhan werd een uitbraak van een infectieziekte gerapporteerd, en de ontdekking van besmette reizigers uit Wuhan in Thailand en Japan suggereerde dat deze uitbraak veel groter moest zijn dan tot dan toe gerapporteerd. De aandacht richtte zich eerst op het vaststellen van de epidemiologische karakteristieken van deze nieuwe infectieziekte. Dit betrof onder andere de verdelingen voor incubatietijd en generatietijd (Backer et al. 2020; Ganyani et al. 2020). De incubatietijd is de tijdsduur tussen het besmet raken en het vertonen van de eerste symptomen van de ziekte. De generatietijd is de tijdsduur tussen de dag waarop een persoon is besmet en de dag waarop diens besmetter was besmet. Met incubatietijd- en generatietijdverdeling kan een schatting gemaakt worden van het aandeel pre-symptomatische infecties (het aandeel van de besmettingen dat plaatsvindt voordat de besmetter symptomen vertoont; Tindale et al. 2020). Deze informatie werd gebruikt bij het opzetten van bron- en contactopsporing; mensen met COVID-19 werden ook gevraagd naar contacten die gemaakt waren voordat de eerste symptomen verschenen.

Rekenwerk in context: hoe, wat en waarom

Al vrij snel werd ook berekend wat COVID-19 zou betekenen voor de belasting van de zorg in Nederland (Anderson et al., 2020, Wallinga et al., 2020). Nadat de pandemie ook Nederland bereikt had, werd het rekenen aan de verwachte zorgvraag een enorme hoeveelheid werk. De tijdsperiode tussen vraag en advies was soms maar enkele dagen of zelfs uren. Terwijl het werk drukker was dan ooit, bleek het ook belangrijk uit te leggen welke gegevens verzameld waren, en hoe de berekeningen bij een advies tot stand waren gekomen. Die uitleg moest worden gegeven aan beleidsmakers en politici, aan andere experts die adviseren, aan collegae, aan journalisten, en aan het grote publiek. De communicatie werd een baan op zich.

Normaliter worden de rekenstappen opgeschreven in een publicatie, aangeboden voor peer review door vakgenoten, en na verbetering van eventuele onvolkomenheden wordt het gepubliceerd. Deze route is moeilijk te nemen als een advies in een heel korte tijdsperiode van dagen of soms uren tot stand moet komen. Het is natuurlijk wel mogelijk om gebruik te maken van eerder ontwikkelde en gepubliceerde methoden, en dit aan te geven door enkele referenties, zodat anderen de rekenstappen in grote lijnen kunnen volgen. Een uitgebreidere uitleg van een berekening verschijnt na het advies in een rapport, gericht op een brede groep van geïnteresseerden (inclusief beleidsmakers) of als publicatie in een peer reviewed tijdschrift, gericht op vakgenoten.

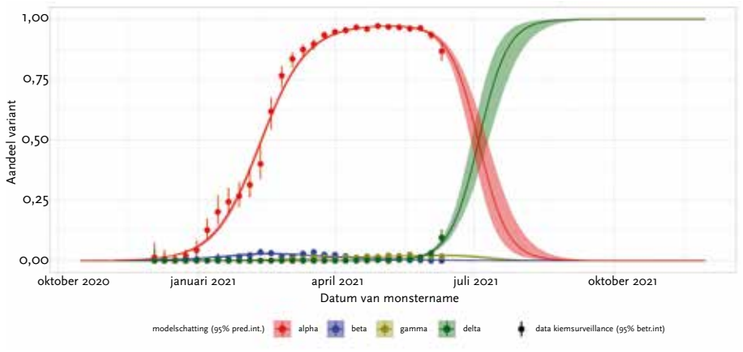
In de tussentijd wordt de uitleg gegeven aan andere experts, journalisten en een breder publiek van geïnteresseerden. Het is daarbij belangrijk om aan te geven wat de epidemiologische betekenis is van de berekende uitkomsten, en om daarbij aan te geven op welke gegevens de berekening is gebaseerd. Een aantal IC-opnames kan zijn berekend op basis van gerapporteerde IC-opnames per leeftijdsgroep per ziekenhuis gedurende de afgelopen dag, maar ook op basis van prognoses voor de situatie over 14 dagen. Om dat onderscheid te maken, moeten we het doel en de context van de berekening aangeven. Gaat het om het verwerken van binnengekomen gerapporteerde gegevens van vorige week, of gaat het om een prognose voor de toestand van volgende week? Of gaat het om scenarioanalyses waarbij een variabele andere waarden aan mag nemen dan we hebben gemeten? Het doel van een berekening is vaak evident voor degenen die de berekeningen uitvoeren maar niet altijd gemakkelijk begrepen door media en publiek, en dit kan tot verwarring leiden. Hieronder enkele ervaringen die dit illustreren.

Waarom we een berekening doen

IC-bezetting. Een prognose van de IC-bezetting door COVID-19-patiënten kan gemaakt worden op basis van het aantal nieuwe IC-opnames per dag met COVID-19-patiënten, en de verdeling van de ligduur op de IC van een COVID-19-patiënt. Sinds de tweede helft van maart 2020 was het mogelijk om die ligduur van een COVID-19-patiënt op een Nederlandse IC-afdeling te schatten, en die ligduur was in de ordegrootte van 20 dagen. Sinds die tijd worden er prognoses gegeven van de IC-bezetting. Voordat de ligduur op de IC in Nederland gemeten kon worden werden er ook al scenarioanalyses gedaan om effecten van interventies te verkennen, en hierbij werd een ligduur van 10 dagen aangehouden, conform aannames in andere studies. Iemand die de eerste prognose met die eerdere scenarioanalyses vergelijkt kan concluderen dat de scenario analyses gemaakt waren met een te korte ligduur en dat de analyses dus fout waren. Dat was genoeg om het nieuws te halen als een rekenfout.

Reproductiegetal. Als maat voor transmissie wordt het 'effectief reproductiegetal' genomen. Dit reproductiegetal wordt gedefinieerd als het aantal besmettingen dat wordt veroorzaakt door een typisch besmet individu. Het reproductiegetal is terug te rekenen uit gerapporteerde aantallen nieuwe besmettingen naar eerste ziektedag (Wallinga en Lipsitch 2007, Gostic et al. 2020). Aan de verandering in de waarde van het reproductiegetal kunnen we achteraf de effectiviteit van genomen maatregelenpakketten aflezen. Omdat wordt teruggerekend welk aantal mensen besmet werd door iemand met een eerste ziektedag op dag t, kan de uitkomst pas geven worden op dag t+14, als alle mogelijke besmettingen hebben plaatsgevonden en nadat de besmette mensen hun eerste symptomen hebben gehad, getest zijn, en de positieve test resultaten zijn doorgegeven. Andere wetenschappers geven in een blog hun schattingen van het reproductiegetal van vandaag op basis van prognoses waarin een aanname is verwerkt wat de effectiviteit is van de maatregelen gaat worden. Iemand die deze twee rapportages van een reproductiegetal vergelijkt kan concluderen dat het wenselijker is om het meest recente reproductiegetal te hebben, en eisen om de rekenmethoden te veranderen zodat er geen vertraging van 14 dagen meer is. Maar daarmee verdwijnt het oorspronkelijke doel om de effectiviteit van maatregelen te evalueren op basis van waarnemingen uit het zicht.

Percentage nieuwe virusvariant. Begin 2021 kwam een nieuwe variant van het SARS-CoV-2 virus op. Destijds werd dit de Britse variant genoemd, tegenwoordig wordt het aangeduid als de alfa-variant van het virus. Op basis van sequentieanalyse van binnengekomen monsters van besmette personen kan het aandeel van deze variant worden vastgesteld. Laboratoria vergelijken hun monsters naar de datum van monsterafname. Met een model kon al in januari 2021 een prognose worden gemaakt voor het aandeel van de nieuwe variant naar datum van monsterafname. De ontwikkeling van het aandeel van de Britse variant in Nederland gedurende maart 2021 blijkt verrassend goed overeen te komen met deze gemaakte prognose; de voorspelde datum van monstername waarop het aandeel van de besmettingen voor 50% uit de nieuwe variant bestaat zit er maar enkele dagen naast. In de nieuwsberichten rondom de oprukkende variant worden echter verschillende meetmomenten gebruikt van het aandeel nieuwe variant, zonder dit expliciet te noemen. Soms wordt een percentage genoemd van infecties naar eerste ziektedag, soms naar dag van melding, soms naar dag van laboratoriumanalyse, soms naar dag waarop het percentage wordt berekend. Er is een flink tijdsverschil tussen al deze meetmomenten. Dat kostte veel tijd om uit te leggen, en de indruk bleef hangen dat de prognose ernaast zat.



Voorbeeld van een modelleringsgrafiek: gemodelleerde en gemeten aandelen van de verschillende varianten (versie van 23 juni 2021); Inschatting toename B.1.1.7 (alpha), B.1.351 (beta), P.1 (gamma) en B.1.671.2 (delta) in Nederland.

Tot slot

Veel meer dan bij eerdere pandemieën worden berekeningen gebruikt bij het adviseren over de bestrijding. Deze berekeningen moeten in een korte tijdsperiode worden gemaakt. Het is daarom essentieel goed uit te leggen wat, hoe, en vooral waarom iets uitgerekend wordt en hoe de uitkomst geduid moet worden.

Literatuur

Backer, J. A., Klinkenberg, D., Wallinga J. (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. Euro Surveill., 25(5). pii=2000062. https://doi.org/10.2807/1560-7917.ES.2020.25.5.2000062

Ganyani T., Kremer C., Chen D., Torneri A., Faes C., Wallinga J., & Hens N. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Euro Surveill, 25*(17). pii=2000257. https://doi.org/10.2807/1560-7917.ES.2020.25.17.2000257

Tindale, L. C., Stockdale, J. E., Coombe, M.,Garlock, E. S., Lau, W. Y. V., Saraswat, M., Zhang, L., Chen, D., Wallinga, J., & Colijn, C. (2020). Evidence for transmission of COVID-19 prior to symptom onset. *eLife, 9*. e57149 doi: 10.7554/eLife.57149

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *Lancet, 395*(10228), 931–934. https://doi.org/10.1016/S0140-6736(20)30567-5

Wallinga, J., Backer, J. A., Klinkenberg, D., Hoek, A. J. van, Hahné, S. J. M., Van der Hoek, W., & Hof, S. van den. (2020). De COVID-19-epidemie: indammen en afvlakken: Bestrijdingsmaatregelen tegen piekbelasting in de zorg. *Ned Tijdschr Geneeskd, 2020*. 164:D4961

Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B., 274*, 599–604. http://doi.org/10.1098/rspb.2006.3754

Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J. A., De Salazar, P. M., Hellewell, J., Sophie Meakin, S., James D. Munday, J. D., Nikos I. Bosse, N. I., Sherrat, K., Thompson, R. N. White, L. F. Huisman, J. S., Scire, J., Bonhoeffer, S., Stadler, T., Wallinga, J., Funk, S., Lipsitch, M., & Cobey S. (2020). Practical considerations for measuring the effective reproductive number. *Rt. PLoS computational biology, 16*(12), e1008409.

Jacco Wallinga is buitengewoon hoogleraar modellering van infectieziekten aan het Leids Universitair Medisch Centrum en hoofd van de afdeling modellering van infectieziekten bij het RIVM.
E-mail: jacco.wallinga@rivm.nl

Don Klinkenberg is onderzoeker modellering van infectieziekten bij het RIVM
E-mail: Don.Klinkenberg@rivm.nl

DE ANDERHALVEMETERSAMENLEVING IN ZIEKENHUISWACHTKAMERS

Sander Dijkstra, Maarten Otten, Gréanne Leeftink, Bas Kamphorst & Richard Boucherie

Anderhalvemetersamenleving was in het coronajaar 2020 het Woord van het Jaar. Deze afstandsmaatregel had een enorme impact op onze maatschappij. In wachtkamers van ziekenhuizen werd het aantal gelijktijdig aanwezige personen de limiterende factor. De reductie van het beschikbare aantal wachtplaatsen had directe gevolgen voor het mogelijke aantal fysieke afspraken. Een deel van de afspraken werd daarom vervangen door digitale afspraken, of zelfs helemaal afgezegd. We ontwikkelden een roostermethode die het mogelijk maakt om alle afspraken doorgang te laten vinden, waarvan zoveel mogelijk fysiek, binnen de beperking van de capaciteit van de wachtkamers.

De capaciteit van wachtkamers in ziekenhuizen is door de anderhalvemetermaatregel vanwege de coronapandemie drastisch verlaagd, soms wel tot een derde van de voor de pandemie beschikbare capaciteit. Patiënten vragen precies op tijd te komen voor hun afspraak lijkt uitkomst te bieden om de grenzen van de wachtkamercapaciteit te respecteren. In de praktijk is dit niet realistisch, omdat enerzijds de afspraak van een voorgaande patiënt kan uitlopen, en anderzijds patiënten arriveren vanuit een andere afspraak in het ziekenhuis. Patiënten hebben veelal meerdere afspraken op een dag en nemen in de tijd tussen deze afspraken, de zogenoemde overbruggingstijd, plaats in de wachtkamer. De overbruggingstijd heeft vaak een minimale duur, bijvoorbeeld wanneer een arts bepaalde testresultaten nodig heeft vóór aanvang van een afspraak. De druk op de wachtkamercapaciteit als gevolg van overbruggingstijden, variabiliteit in aankomst van patiënten en afspraakduur resulteert in een beperking van het aantal planbare fysieke afspraken. Afzeggen van afspraken of omzetten van fysieke in digitale afspraken verlaagt deze druk, maar resulteert in zorg in een minder wenselijke vorm, of afschalen van zorg.

We beschrijven een roostermethode om alle afspraken doorgang te laten vinden, met een maximaal aantal fysieke afspraken, binnen de door afstandsmaatregelen beperkte capaciteit van de wachtkamers. Naast losse afspraken voor een deel van de patiënten beschouwen we het effect van de hele keten van afspraken die via de overbruggingstijden een grote druk legt op de wachtkamercapaciteit. Vanuit historische data uit het ZIS (ziekenhuisinformatiesysteem) en aanvullende data over de werking van de afdelingen bepalen we de parameters van een *Integer Linear Program* (ILP), waarmee we op basis van gemiddelden een zogenaamde blauwdruk opstellen voor de roosters op de polikliniek en dagbehandeling, met als doel dat zoveel mogelijk afspraken fysiek doorgang kunnen vinden. Het effect van variabiliteit van de aankomsttijdstippen, consult- en behandelduren voor de resulterende optimale blauwdruk bepalen we met een Monte Carlo Simulatie (MCS). Op basis van deze resultaten passen we indien nodig iteratief de parameters van het ILP aan om een nieuwe blauwdruk te bepalen, die robuust is tegen overbezetting van de wachtkamers als gevolg van de in de praktijk altijd aanwezige variabiliteit.

Roostermethode voor de polikliniek

Patiënten hebben losse afspraken of volgen patiëntpaden, die bestaan uit een serie afspraken op dezelfde dag. Op deze paden onderscheiden we verschillende fasen, zoals bloedonderzoek, een consult op de polikliniek, en een behandeling. Per fase zijn verschillende typen afspraken mogelijk, zoals een vervolgafspraak of een afspraak voor een nieuwe patiënt, waaraan verschillende resources kunnen worden gekoppeld, zoals een arts, verpleegkundige en bed. Afspraken worden gepland aan de hand van een blauwdruk, waarin afspraakslots voor typen afspraken zijn opgenomen, waarop door een planner specifieke patiënten van het corresponderende type worden geboekt. Voor adequate beschikbaarheid van zorg moet de blauwdruk de juiste mix bevatten van aantallen en typen afspraken op de juiste tijdstippen.

Onze roostermethode is voor iedere polikliniek en behandelcentrum inzetbaar en bestaat uit vier onderdelen: data verzamelen, optimaliseren, simuleren en verfijnen. De verzamelde data bevatten twee delen: data over de werking van de afdelingen, waaronder het aantal artsen, verpleegkundigen en bedden, de tijden waarop afspraken ingepland kunnen worden en de verschillende soorten afspraken met bijbehorende duren en mogelijkheid tot vervanging door een digitaal alternatief vanuit medisch perspectief, en ZIS-data om patiëntpaden te identificeren, en hun frequentie te bepalen. Deze data samen met de beperkte wachtkamercapaciteit geven de randvoorwaarden waaraan een blauwdruk moet voldoen.

Met behulp van het door ons ontwikkelde ILP bepalen we een optimale blauwdruk, waarin het aantal afspraken dat fysiek doorgang kan vinden maximaal is onder de aanname dat alle afspraken precies op tijd beginnen en een vaste duur hebben. Om inzicht te krijgen in deze blauwdruk in de dagelijkse praktijk, onderzoeken we middels MCS de effectuering van het rooster voor variabele aankomsttijden en afspraakduren en bepalen een 95% betrouwbaarheidsinterval voor de wachtkamerbezetting. Dit interval ligt veelal boven de wachtkamerbezetting die door het ILP wordt bepaald, omdat variabiliteit verstoringen in het rooster veroorzaakt, die de bezetting van de wachtkamer verhogen. Indien de bovengrens van het 95% betrouwbaarheidsinterval van de wachtkamerbezetting op een deel van de dag boven de beschikbare capaciteit komt, stellen we de parameters van het ILP zo bij dat de bezetting op deze tijden wordt gereduceerd en herhalen we zowel de ILP-optimalisatie als de MCS tot aan de beperking op de wachtkamercapaciteit is voldaan.

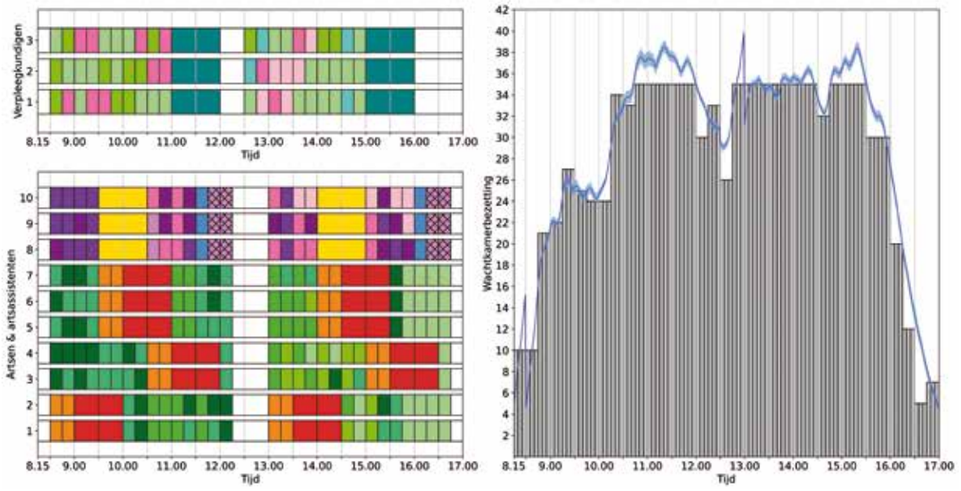
Reumatologie in de Sint Maartenskliniek

Onze methode hebben we onder andere toegepast op de reumatologieafdeling van de Sint Maartenskliniek (SMK), een ziekenhuis in Nijmegen gespecialiseerd in houding en beweging. Op deze afdeling werken iedere dag 3 verpleegkundigen, 3 arts-assistenten en 7 artsen. Patiënten die voor een afspraak komen wachten voor hun eerste afspraak of tussen twee afspraken in een gezamenlijke wachtkamer, waarvan de capaciteit door de anderhalvemetermaatregel is gereduceerd van 32 naar maximaal 18 personen. Een patiëntpad kan bestaan uit (een subset van) 4 fasen: bloedonderzoek in fase 1, een afspraak bij een verpleegkundige in fase 2, een consult met een arts of arts-assistent in fase 3 en in fase 4 een bezoek aan de apotheek voor medicatie. Uit de data-analyse blijkt dat er 16 unieke patiëntpaden zijn die gemiddeld minimaal 1 keer per dag plaatsvinden.

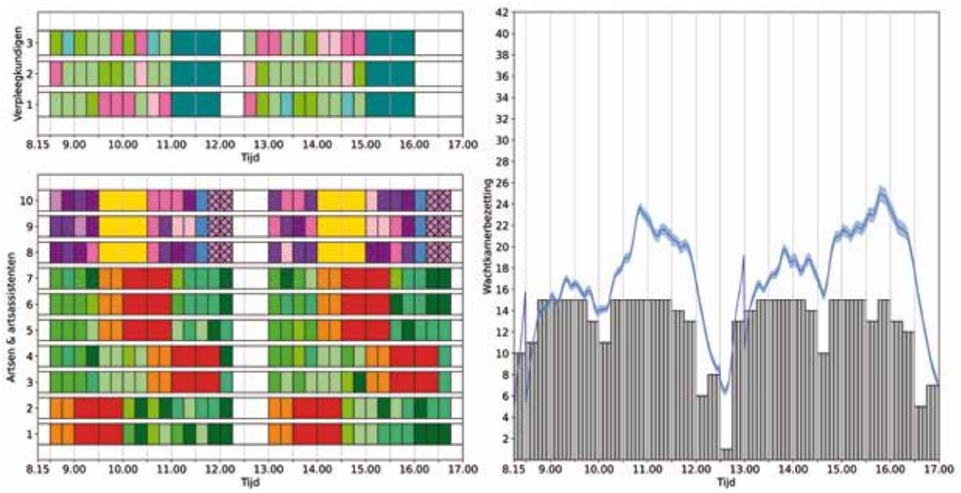
We hebben de wachtkamerbezetting van de pre-corona blauwdruk geanalyseerd om inzicht te krijgen in de waarde van onze interventie. Wachtkamerbezetting is bij het opstellen van de pre-corona blauwdruk niet als prestatiemaat meegenomen. Deze blauwdruk is gemaakt op basis van afspraaktypen en niet op basis van patiëntpaden. Hierdoor kan grote variatie optreden in wachtkamerbezetting, afhankelijk van de planning op een afspraakslot van patiënten van hetzelfde type, maar met verschillende patiëntpaden. Met behulp van een kleine aanpassing van ons ILP (om de pre-corona blauwdruk vast te leggen), hebben we de best en de slechtst mogelijke realisatie ten aanzien van wachtkamerbezetting van patiëntpaden in de pre-corona blauwdruk bepaald.

Figuren 1(links) en 2(links) tonen de met ons ILP bepaalde slechtste en beste realisatie van de pre-corona blauwdruk en figuren 1(rechts) en 2(rechts) de bijbehorende bezetting van de wachtkamer (zie kader voor toelichting). We zien dat in het slechtste geval de bezetting van de wachtkamer vrijwel de gehele dag boven het maximum van 18 aanwezige patiënten ligt en kan oplopen tot 40 patiënten. Ook in het beste geval loopt de bezetting aan het einde van zowel de ochtend als de middag geruime tijd op tot boven de 18 met als piek 24 patiënten. We zien verder dat vooral in het beste geval (figuur 2, rechts) variabiliteit een grote rol speelt in de gerealiseerde bezetting van de wachtkamer: de maximale bezetting voldoet met 15 ruimschoots aan de coronabeperking van 18 indien alle variabiliteit kan worden uitgesloten en dus alle afspraken precies op tijd beginnen en eindigen en alle patiënten precies op tijd arriveren. Aangezien variabiliteit niet valt uit te sluiten en de beste realisatie slechts sporadisch voor zal komen, concluderen we dat de pre-corona blauwdruk niet bruikbaar is in de anderhalvemetersamenleving.

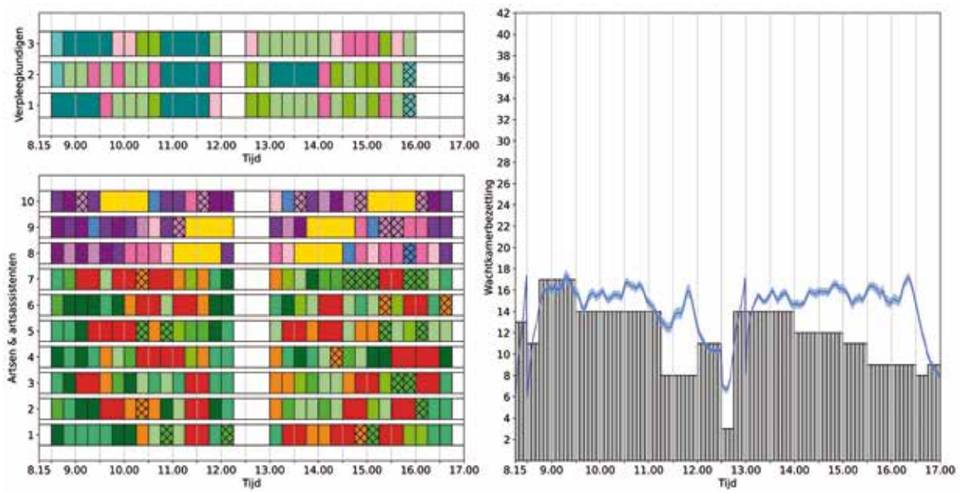
Figuren 1(links) en 2(links) tonen de slechtste en beste realisatie van de pre-corona blauwdruk voor de polikliniek, waarin de extra details over afspraaktypen zijn opgenomen die we hebben bepaald met onze methode. Vergelijkbare afspraaktypen hebben kleuren in dezelfde toon, bijvoorbeeld verschillende kleuren paars staan voor follow-up afspraken met een arts-assistent op de polikliniek, waarvan een deel digitaal is ingevuld (deze zijn gearceerd). Donkerpaars heeft een volgende afspraak bij de apotheek en roze een bloedtest na de poliklinie-kafspraak. De bovenste figuur geeft de blauwdruk voor de verpleegkundigen, de onderste voor de artsen. Figuren 1(rechts) en 2(rechts) tonen de wachtkamerbezetting per kwartier: de grijze staven als oplossing van het ILP, de blauwe lijn uit de MCS, waarbij het blauwe gebied het 95% betrouwbaarheidsinterval weergeeft. Figuren 3(links) en 3(rechts) tonen de optimale blauwdruk, waarin we ook digitale consulten (gearceerd) zien voor andere afspraaktypen.



Figuur 1. Slechtste realisatie van de pre-corona blauwdruk



Figuur 2. Beste realisatie van de pre-corona blauwdruk



Figuur 3. Optimale blauwdruk

Figuur 3(links) toont de blauwdruk die bepaald is met onze roostermethode en figuur 3(rechts) de bijbehorende bezetting van de wachtkamer. In deze blauwdruk worden alle patiëntpaden gepland, met inachtneming van de medische beperking op digitale afspraken, met maximaal aantal fysieke afspraken, binnen de grenzen van de beperking op de wachtkamercapaciteit. In de ILP-oplossing overbruggen patiënten zo kort mogelijk in de wachtkamer. Dit blijkt niet voldoende om alle afspraken fysiek doorgang te laten vinden. Indien 12% van de afspraken in de blauwdruk wordt vervangen door een digitale afspraak kunnen onder de coronabeperkingen alle afspraken doorgang vinden op de afdeling reumatologie van de SMK.

Conclusie

Met wiskundige optimalisatie en simulatie kunnen ziekenhuizen in een heel uitdagende tijd de zorg zo goed mogelijk voort blijven zetten. Naast een succesvolle implementatie in de SMK hebben we onze roostermethode ook ingezet voor de medische oncologie & hematologie van UMC Utrecht, waarbij slechts 13% van de afspraken moest worden omgezet naar digitaal om alle afspraken doorgang te laten vinden. Meer informatie over deze casus, de benodigde data en software, en mogelijkheden om deze modellen toe te passen, is beschikbaar in de referenties en via https://www.utwente.nl/en/choir/research/Covid19-outpatientclinic/.

Literatuur

Otten, M., Dijkstra, S., Leeftink, G., Kamphorst, B., Olde Meierink, A., Heinen, A., Bijlsma, R., & Boucherie, R. J. (2021). Outpatient clinic scheduling with limited waiting area capacity. *JORS*. DOI: 10.1080/01605682.2021.1978347

Dijkstra, S., Otten, M., Leeftink, G., Kamphorst, B., & Boucherie, R. J. (2021). Limited waiting areas in outpatient clinics: an intervention to incorporate the effect of bridging times in blueprint schedules. *Manuscript aangeboden voor publicatie.*

Sander Dijkstra (s.dijkstra-1@utwente.nl), Maarten Otten (j.w.m.otten@utwente.nl), Gréanne Leeftink (a.g.leeftink@utwente.nl) en Richard Boucherie (r.j.boucherie@utwente.nl) zijn verbonden aan het Center for Healthcare Operations Improvement and Research (CHOIR) van de Universiteit Twente. Op basis van toonaangevend wetenschappelijk onderzoek en in nauwe samenwerking met de praktijk ontwikkelt en implementeert CHOIR logistieke oplossingen in de zorg. Bas Kamphorst (bas.kamphorst@rhythm.nl) werkt bij Rhythm BV, de spin-off van CHOIR. Rhythm maakt opgedane kennis op gebied van zorglogistiek breder toegankelijk door begeleiding van zorgaanbieders in capaciteitsmanagement en opleiden van zorgprofessionals in zorglogistiek.

UITMUNTENDE MASTER'S OF PH.D. THESIS BEGELEID?

OPROEP OM KANDIDATEN TE NOMINEREN VOOR DE JAN HEMELRIJK EN WILLEM R. VAN ZWET AWARDS 2021

Ter bekroning van een uitzonderlijke afstudeerprestatie aan een Nederlandse instelling voor wetenschappelijk onderwijs/hoger beroepsonderwijs looft de VVSOR al vanaf 1989 een scriptieprijs uit. In 2014 kreeg deze de naam *Jan Hemelrijk Award*. Sinds 2012 is er ook een prijs voor dissertaties: de *Willem R. van Zwet Award*.

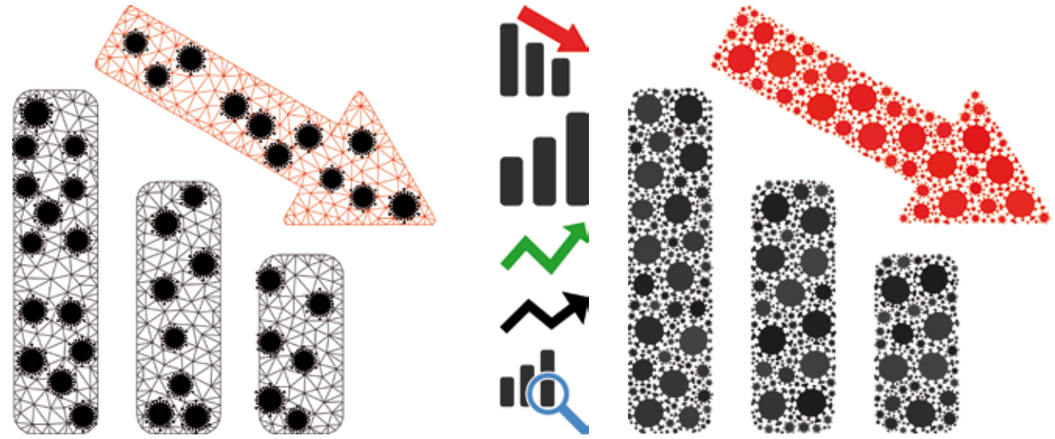
De VVSOR roept op tot nominaties voor deze prijzen. Beide prijzen bestaan uit een oorkonde en een geldbedrag van 1.000 euro. De prijzen zullen worden uitgereikt tijdens de Annual Meeting van de VVSOR, op donderdag 17 maart 2022. Genomineerd kunnen worden personen die van september 2019 tot en met augustus 2021 zijn afgestudeerd respectievelijk gepromoveerd – dat wil zeggen gedurende de twee meest recente afgesloten academische jaren – en die nog niet eerder zijn genomineerd.

Hierbij worden supervisors (begeleiders) opgeroepen om een uitmuntende afstudeerscriptie (Master) of dissertatie (Ph.D.) te nomineren voor de Jan Hemelrijk dan wel Willem R. van Zwet Award 2020. In aanmerking voor de prijzen komen zowel originele theoretische bijdragen aan als inventieve toepassingen van bestaande theoretische concepten uit de statistiek en/of operations research.

De indiening van een nominatie dient vergezeld te gaan van een aanbevelingsbrief van de supervisor van de genomineerde. De precieze procedure voor beide prijzen, alsmede de reglementen en het nominatieformulier zijn te downloaden via de website van de VVSOR, www.vvsor.nl. De nominatie dient uiterlijk 23 januari 2022 binnen te zijn.

Namens de VVSOR,
Dr. Ad Ridder, juryvoorzitter Jan Hemelrijk Award
Prof. dr. Jelle Goeman, juryvoorzitter Willem R. van Zwet Award
Dr. Sander Scholtus, Secretaris der beide jury's

Het voornemen is om vanaf volgend jaar voor beide prijzen alleen nog nominaties toe te laten uit het meest recente afgesloten academische jaar.



OVERZICHTELIJKE WEERGAVE VAN CORONADATA

In een tijd van grote onzekerheid zien we dat mensen vaak krampachtig zoeken naar betrouwbare informatie, die op een begrijpelijke manier wordt gecommuniceerd. Datavisualisatie en statistiek bieden ruimte om duidelijk te communiceren over de ontwikkeling van een epidemie. Tegelijkertijd is het belangrijk om zich te realiseren dat de data en statistische methoden uit allerlei expertisegebieden komen (zoals de medische, sociale, en economische hoek) en dat vereist brede samenwerking tussen experts uit die gebieden en statistici. Hoe rapporteer en duid je een groot scala aan data voor een breed publiek tijdens een crisis zoals deze?

Marino van Zelst en Yorick Bleijenberg

Afgelopen jaar hebben wij ons ingezet om de data rondom corona zo overzichtelijk mogelijk weer te geven via sociale media en websites. Het RIVM rapporteerde de dagelijkse coronadata over besmettingen, opnames, en sterfgevallen, via een epidemiologische rapportage en sociale media zoals Twitter. Vanaf 1 juli 2020 is het RIVM overgestapt op een wekelijkse rapportage omdat de epidemie toen op een laag punt was, waardoor het niet meer zinvol werd geacht om de meldingen dagelijks door te geven. Tegelijkertijd bemerkten wij dat een breder publiek nog steeds geïnteresseerd was om op de hoogte te worden gehouden over het verloop van de epidemie. Het RIVM stelde vanaf 1 juni 2020 ook dagelijks open data beschikbaar over het aantal nieuw geconstateerde besmettingen, opnames, en sterfgevallen.

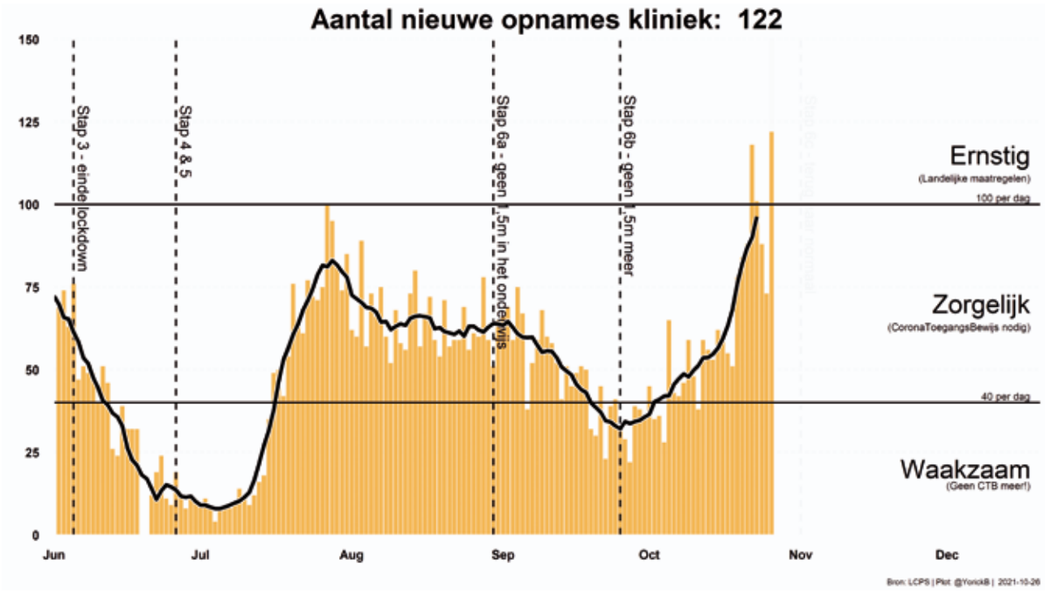
Wij achtten het van belang om deze informatievoorziening op een integrale wijze te publiceren via sociale media, aangezien dit een kanaal is waar op effectieve wijze korte samenvattingen gegeven konden worden van de epidemiologische data. Het officiële dashboard nam deze taak uiteraard ook op zich, maar het dashboard was gebonden aan een aantal instructies en het geven van duiding via het dashboard werd in het begin vrij complex geacht . Tevens was het doel van het dashboard om

een zo uitgebreid mogelijk overzicht te bieden van alle beschikbare data. Wij zijn echter van mening dat het verstandiger is om selectiever te rapporteren om zo de duidelijkheid van de rapportage te verhogen. Uiteraard hebben wij de vrijheid gehad om accenten te leggen op zaken die wij belangrijk achten omdat we die vrijheid hebben als burgerinitiatief.

Overwegingen bij datarapportage – Duiding en interpretatie

Wij zijn allebei niet opgeleid als epidemioloog maar zijn vooral geïnteresseerd in datavisualisatie en communicatie. Daarom hebben we er bewust voor gekozen om uitgebreid en herhaaldelijk te spreken met inhoudelijke experts, zoals virologen, epidemiologen, en zorgprofessionals om op die manier een beeld te krijgen van de meest relevante data. Dit bood ons de ruimte om de data zelf beter te 'verslaan' maar ook om vragen te beantwoorden van de lezer. De coronacrisis kenmerkt zich tenslotte door een veelvuldigheid aan desinformatie, waarbij het rapporteren van ruwe data ruimte biedt voor meerdere interpretaties. Aan de andere kant kunnen data over het algemeen meerledig worden geïnterpreteerd, waarbij bij meerdere interpretaties een gedeelte van de realiteit kunnen omvatten. Wij hebben daarom gekozen voor een tweesporenbeleid in de rapportage.

Enerzijds rapporteren we de ruwe data op een zo objectief mogelijke manier, voor zover dat mogelijk is. Hierin kiezen we ervoor om het aantal besmettingen, opnames, en sterfgevallen te rapporteren 'as is', zonder daar expliciete duiding bij te geven. Aan de andere kant gebruiken we deze rapportage ook om te wijzen op epidemiologische aandachtspunten: waar gaat de epidemie heen, vanaf welk moment is dat zorgwekkend, en welke beleidsoverwegingen worden dan relevant? Deze twee zaken proberen we zo strikt mogelijk te onderscheiden om een zo breed mogelijk publiek te bedienen en te informeren over de stand van zaken rondom de epidemie. Het rapporteren van de ruwe data gaat overigens wel gepaard met het inkaderen van deze interpretatie binnen het bestaande coronabeleid, om de lezer een houvast te bieden. Zo integreren we bijvoorbeeld de indicatoren uit de routekaart met de epidemiologische data, zodat de lezer een beter gevoel krijgt voor mogelijke beleidsontwikkelingen en het op- of afschalen van interventies. In figuur 1 worden het aantal ziekenhuisopnames per dag getoond alsook het zevendaags gemiddelde. Op de horizontale as worden tevens de risiconiveaus uit de routekaart getoond, inclusief bijbehorende interventies zoals het coronatoegangsbewijs. Voor een historische context zijn als laatste de interventies uit het verleden getoond. Zo kunnen we in een opslag de stand van zaken, de trend, en de mogelijke interventies tonen.



Figuur 1. Een weergave van het aantal ziekenhuisopnames en interventies bij een bepaald risiconiveau uit de routekaart

Beperkingen

Uiteraard brengt een grote datastroom, opgezet tijdens een wereldwijde crisis, ook de nodige problemen met zich mee. Een bekend buitenlands voorbeeld is het tijdelijk verliezen van cruciale data over besmettingen door het gebruik van een oud dataformat. De PHE (het Engelse RIVM) gebruikte namelijk .xls om informatie over besmettingen op te slaan, waardoor ongeveer 16.000 besmettingen en informatie rondom deze besmettingen tijdelijk verloren raakten. Hierdoor ontstond een achterstand in bron- en contactonderzoek, een cruciaal middel om de epidemie te bestrijden . Terwijl een probleem van dit formaat zich niet heeft voorgedaan in Nederland, hebben we toch veel problemen ervaren in de afgelopen 1,5 jaar. Door technische storingen waren er regelmatig onderrapportages in de dagelijkse meldingen, waardoor het soms lastiger werd om de trend te interpreteren. Alhoewel dit op zich niet zo'n groot probleem is, werd dit wel problematisch toen media dit interpreteerden als 'de epidemie is aan het dalen'.

Een ander voorbeeld is het achterstallig onderhoud in de datarapportage. Een aantal meldingen laat soms langer op zich wachten, zoals het doorgeven van de sterfgevallen door corona. Op 26 september 2020 werden er 38 sterfgevallen gemeld, wat een factor 3 hoger was dan men kon verwachten op basis van de epidemiologische curve. De koppen in de media waren te voorspellen. Het vereist inhoudelijk inzicht in de epidemiologische data om dit soort afwijkingen te interpreteren en te duiden. In zo'n situatie is het beter om uit te leggen wat de oorzaak is van deze afwijking, wat samenwerking vereist met inhoudelijke experts. In deze casus ontdekten we dat een GGD 21 sterfgevallen had gemeld die al meer dan drie maanden geleden hadden plaatsgevonden. Naar aanleiding van dit geval hebben we er voor gekozen om correcties op alle epidemiologische data te traceren zodat we zo snel mogelijk een beeld konden krijgen van de oorzaak van dit soort afwijkingen.

Aanpassen en samenwerking

We hebben in het afgelopen jaar 100% open-source gewerkt. Dat impliceerde dat we al onze data en scripts openlijk toegankelijk hebben gemaakt via Github en alle code in open-source programma's hebben geschreven (voornamelijk R). De motivatie hiervoor was tweeledig: ten eerste is transparantie tijdens een crisis essentieel, omdat je de afnemers van de informatie kunt meenemen in de overwegingen. Datapresentatie en interpretatie is niet waardenvrij, wat in een gepolariseerde crisis zoals de coronapandemie extra lastig is. Door maximale transparantie te betrachten waren we in staat om verantwoording af te leggen en afnemers in de gelegenheid te stellen om de data en analyses te hergebruiken op andere manieren. Ten tweede creëerde dit de mogelijkheid tot samenwerking met andere gebruikers van de data en geïnteresseerden. Vele gebruikers hebben ook bijgedragen aan de ontwikkeling van de datapresentatie en interpretatie: enerzijds door code te schrijven of te verbeteren of anderzijds door suggesties aan te leveren, die we daarna zelf konden verwerken. We waren allebei ook onderdeel van een data community waar we veel geleerd hebben en veelal snel zaken konden aanpassen. Dit heeft geleid tot een productieve samenwerking met afnemers, andere analisten, en instanties zoals het RIVM of het ministerie van VWS. Deze open samenwerking, mede mogelijk gemaakt door het hanteren van strikte principes rondom transparantie en evaluatie, hebben de rol die we hebben gespeeld zeker mogelijk gemaakt. Dit is daarmee een andere basisvoorwaarde geworden voor ons als data-analisten rondom deze coronacrisis en eventuele andere crisissen.

Conclusie

De coronacrisis is de eerste pandemie waarbij data en informatie zo toegankelijk zijn voor iedereen die geïnteresseerd is. De technologische ontwikkelingen hebben ons in staat gesteld om vele databronnen aan elkaar te koppelen en te presenteren op een begrijpelijke manier. In dit artikel hebben we getracht een kort overzicht te geven van onze geleerde lessen in het afgelopen jaar, waarbij we focussen op aandachtspunten voor datavisualisatie en interpretatie tijdens een crisis die zich razendsnel ontwikkelt. Duidelijke datavisualisatie en een begrijpelijke toelichting kan mensen een handelingsperspectief geven, waardoor een bijdrage kan worden geleverd aan het bestrijden van zo'n crisis.

Marino van Zelst is infectieziektenmodelleur bij Wageningen University & Research. Zijn onderzoek richt zich ondermeer op de effectiviteit van non-farmaceutische interventies bij uitbraken.
E-mail: marino.vanzelst@wur.nl

Yorick Bleijenberg is data-fan, ondernemer, en team manager op een klantenservice.
E-mail: yorick.bleijenberg@gmail.com

CODE ZWART

crisismanagement met slimme modellen

Er komen twee doodzieke patiënten binnen met COVID-19 op jouw intensive care. Alle overplaatsingslocaties zitten vol, dus je moet een keuze maken. Wie ga je helpen? Hoe maak je een keuze tussen mensen met een ziekte die je niet begrijpt? Voor veel artsen wereldwijd is dit dilemma werkelijkheid geweest, in Nederland waren we heel dichtbij. Om dit dilemma het hoofd te bieden heeft een multidisciplinair team van clinici en datawetenschappers de handen in elkaar geslagen om een data-gedreven model te ontwikkelen, zodat deze kritieke keuze niet zonder informatie gemaakt hoeft te worden. We nemen je mee door een beschrijving van keuzes en dilemma's om van idee naar resultaat te komen in onzekere tijden.

Maarten Ottenhoff

Nu het grootste deel van de bevolking is gevaccineerd lijken de strenge lockdowns van afgelopen jaar verleden tijd. Alle maatregelen waren essentieel om de instroom van COVID-19-patiënten in het ziekenhuis te beperken, zodat iedereen de zorg kan krijgen die zij nodig hebben. Het punt waarop zorg niet voor iedereen gegarandeerd kon worden, ook wel code zwart genoemd, was meerdere keren dichtbij. Met overplaatsingen naar onze Duitse buren kon deze situatie op het nippertje voorkomen worden.

Tijdens de eerste infectiegolf werden al snel ad-hoc protocollen gemaakt voor code zwart. In een ideale wereld zou je een afweging maken aan de hand van de risicofactoren van de ziekte en de huidige gezondheid van de patiënt. Echter, hiervoor moet er wel een duidelijke wetenschappelijke basis zijn van de risicofactoren. Voor het nieuwe coronavirus was die basis er nog lang niet, dus al snel werd er gekeken naar bredere risicofactoren, zoals leeftijd en bepaalde co-morbiditeiten zoals obesitas, diabetes of longaandoeningen. Toen de besmettingscijfers een paar maanden later waren gedaald en de collectieve paniek was gaan liggen, laaide de discussie op of het grondwettelijk was toegestaan om te selecteren op leeftijd.

Kortom, er was een grote vraag naar medische informatie over het coronavirus, maar voor zorgvuldig onderzoek was er tijd nodig. Om een betere keuze te kunnen maken tijdens code zwart moest er acuut informatie komen. Deze vraag naar informatie was het startpunt van een multidisciplinair onderzoek door een team van artsen en data-wetenschappers van het MUMC+ Maastricht en AMC Amsterdam. Het idee was simpel: verzamel de beschikbare data van opgenomen COVID-19-patiënten in Nederlandse ziekenhuizen en ontwikkel een model dat een inschatting maakt van de kans op overlijden. Bij een succesvol model heeft de arts tijdens code zwart extra informatie over de risico's per patiënt, gebaseerd op medische data. In dit artikel neem ik u graag mee in de verschillende ontwikkelingsstappen van dit project. Voor een volledige inhoudelijke beschrijving verwijs ik u graag door naar het inmiddels gepubliceerde artikel in het wetenschappelijke tijdschrift BMJ open (Ottenhoff et al. 2021).

Stap 1: Dataverzameling

Een week na het begin van de eerste lockdown in maart 2020 ging het project van start. Er waren zoveel mogelijk data nodig, dus werden alle mogelijke kanalen ingezet, van persoonlijke netwerken tot een oproep op televisieprogramma *Op1*. Uiteindelijk waren tien ziekenhuizen bereid om data aan te leveren. Ook bedrijven droegen graag op vrijwillige basis bij, zo konden we bijvoorbeeld de databasestructuur van CASTOR (data-management-platform voor klinische trials) kosteloos gebruiken. Ondanks dat projecten vaak gestroomlijnd lijken te verlopen op papier, zo anders was het in de praktijk: het online invoerformulier werd meerdere malen aangepast toen de data invoer al begonnen was; vanuit statistisch oogpunt een slecht idee. Ook werd er en passant hard gewerkt om alles juridisch in orde te krijgen. Ethische commissies moesten snel een voorstel krijgen om goed te keuren, ziekenhuizen moesten besluiten of er wel of geen informed consent nodig was. Desondanks stond er binnen een maand een database met data van meer dan tweeduizend patiënten van tien Nederlandse ziekenhuizen.

Stap 2: Data schoonmaken

Een goede datakwaliteit is cruciaal, dus werd de eerste maand na dataverzameling besteed aan het schoonmaken van de data. Vaak zijn dit vrij simpele bewerkingen zoals alle binaire variabelen dezelfde kant op laten wijzen (nee = 0, ja =1), alle waarden naar dezelfde eenheid omrekenen of tijdstippen omrekenen naar een relatieve tijd. Ook variabelen waarbij de vraagstelling tijdens data-in-

voer veranderde moesten worden aangepast . Niet alleen moet alles numeriek kloppen, maar ook vanuit medisch perspectief moet het logisch en verklaarbaar zijn. Staan er geen vreemde waarden in? Zijn gerelateerde medische variabelen op de juiste manier gecombineerd?

Hierdoor zijn er potentieel invloedrijke fouten hersteld, wat het belang van een goede samenwerking tussen de datawetenschapper en clinici onderstreept.

Stap 3: Modelontwikkeling

Een machine-learning algoritme genereert een model aan de hand van voorbeelden. Dat wil zeggen dat je het algoritme een set aan datapunten geeft met een bijbehorende uitkomst: overleden of niet. Het model leert dan welke onderliggende patronen bij welke uitkomst horen. Vervolgens kun je een vergelijkbare set aan datapunten invoeren in het model, die vervolgens een voorspelling teruggeeft over de uitkomst. Essentieel is een duidelijke definitie van de te voorspellen uitkomst. Wat valt er precies onder het overlijdensrisico en over welke tijdspanne gaat het? Dat laatste is een belangrijke afweging: een lange tijdspanne geeft genoeg tijd om het hele ziektebeloop te vangen, maar het kost meer tijd en de onzekerheid wordt groter. Er kunnen immers meer onverwachte gebeurtenissen voorkomen in zes weken dan in een week. Een te korte tijdspanne zorgt weer voor een zekere voorspelling, maar de kans is groter dat de patiënt daarna nog overlijdt. Het gevolg is dat een patiënt ten onrechte wordt gelabeld als 'overleefd'. Uiteindelijk is er gekozen voor de balans van 3 weken, wat resulteerde in de volgende definitie: De kans op overlijden binnen 21 dagen na ziekenhuisopname. Onder de kwalificatie 'overlijden' vallen ook patiënten die worden ontslagen naar palliatieve zorg. Van de initiële 2527 patiënten bleven er nu 2273 over.

Om meer informatie te krijgen over de risicofactoren zijn de meer dan 400 variabelen teruggebracht aan de hand van selectiecriteria opgezet door een team van clinici. Niet alleen de potentieel voorspellende waarde werd meegewogen, maar ook de uitvoerbaarheid werd bediscussieerd. Voor sommige voorspellende variabelen was er bijvoorbeeld een CT-scan nodig. Theoretisch waardevol, maar vanuit praktisch oogpunt is een CT-scan tijdrovend, duur en er is weinig capaciteit. Het heeft daarom weinig nut om deze variabele mee te nemen.

Uiteindelijk bleven er 80 variabelen over, die vervolgens weer zijn weer opgesplitst in drie categorieën:

1. Premorbiditeiten; leeftijd, geslacht, beroep en medische geschiedenis;
2. Klinische karakteristieken; bestaande uit simpele medische testen zoals hartslag, bloeddruk en ademhalingsritme;
3. Laboratorium en radiologie variabelen: bloedtesten en scans.

Ook de combinaties van categorie 1 en 2 en alle drie de categorieën werden meegenomen. Als laatste maken we ook gebruik van een data-gedreven test om de tien 'meest voorspellende' variabelen mee te nemen. Hierbij selecteerde we de tien variabelen met de hoogste F-waarde uit een Analysis of Variance (ANOVA). In totaal zijn er dus zes verschillende groepen van variabelen meegenomen in het uiteindelijke model.

Voor het modelleren hebben we gekozen voor twee fundamenteel verschillende soorten modellen: een lineaire logistische regressie (LR), en een non-lineair *tree-based gradient boosting* algoritme (TBGB).

Een LR is eenvoudig model dat goed te interpreteren is. In essentie het een binaire versie van een lineaire regressie. Bij een lineaire regressie wordt een lijn gezocht met de kleinste afstand tot alle datapunten. Om uitkomst te transformeren naar een waarde tussen 0 en 1 wordt deze in een logistische functie gegoten: $y = 1 / (1+\exp(-ax+b))$ waarbij ax+b de lijn is van de lineaire regressie. Vervolgens wordt er een grens gesteld om een voorspelling te maken: alle waardes boven die grens is overlijden, alles daaronder overleven.

Een TBGB-model is gebaseerd op een keuzeboom. Het model genereert een serie van keuzes met een uiteindelijke voorspelling: 'Als variabel groter dan 1, dan A, anders B'. Het TBGB-model controleert vervolgens welke voorspellingen hij fout heeft, en geeft deze datapunten een bepaalde weging aan de hand van de grootte van de fout. Hierdoor krijgen de moeilijke datapunten meer aandacht in het leerproces, waardoor de fout kleiner wordt. Dit proces wordt herhaald totdat de foutmarge geminimaliseerd is, en de optimale voorspelling gemaakt kan worden.

Idealiter gebruikt men twee onafhankelijke datasets om het model te trainen en te testen. Hiermee voorkomt je dat de modellen 'overfitten' op de data, dat wil zeggen dat de modellen irrelevante of dataset specifieke patronen vinden die niet generaliseerbaar zijn. Om dat te voorkomen hebben we gebruik gemaakt van een techniek genaamd cross-validatie. Daarbij splitsen we de data op per centrum. Vervolgens zijn negen ziekenhuizen gebruikt als 'training set' en een ziekenhuis als 'test set'. Dit herhaalde we dit tien keer, waarbij elk ziekenhuis een keer als test set gebruikt wordt. Uiteindelijk wordt de gemiddelde score gebruikt als eindscore.

Stap 4: Model validatie

Samenvattend hebben we twee modellen, LR en TBGB, gemodelleerd op 6 verschillende datasets, gebruikmakende van *leave-one-hospital-out cross-validation*. De uitkomst meten we onder meer met de *area under the receiver operator curve* (AUC).

De receiver operator curve is een plot waarbij de sensitiviteit (aantal correct positieve voorspellingen in verhouding met alle daadwerkelijk positieve gevallen plus alle fout negatieve voorspellingen) en specificiteit (aantal correct negatieve voorspellingen in verhouding met alle daadwerkelijke negatieve gevallen plus het aantal fout positieven) tegen elkaar uitgezet worden. Vervolgens wordt de beslisgrens (bijv. een geval is positief bij een waarde boven de 0,5, anders negatief) met kleine stapjes van 0 naar 1 verplaatst en wordt telkens de sensitiviteit en specificiteit berekend. Als je deze collectie aan punten tegen elkaar plot en de oppervlakte onder die lijn berekent, dan krijg je de AUC, een waarde tussen 0 en 1. Bij een AUC van 1 is de voorspelling perfect, bij een AUC van 0 is de voorspelling perfect fout, ofwel precies andersom. Een AUC van 0,5 betekent dat de voorspelling op kansniveau is. Je kunt dan net zo goed het model achterwege laten en gokken wat het juiste antwoord is.

Uiteindelijk haalde LR en TBGB een score van 0,82 AUC, een goede score. Ter vergelijking, als een simpel model wordt gehanteerd zoals 'iedereen ouder dan 70 of 80 zal overlijden', dan krijgt je een AUC van respectievelijk 0,69 of 0,61. Het werkt dus al beter dan simpele op leeftijd gebaseerde regels die in andere landen in de praktijk zijn toegepast.

Om de voorspellende variabelen zo inzichtelijk mogelijk te maken, hebben we gebruik gemaakt van SHAP-waarden. Deze methode berekent de gemiddelde relatieve bijdrage van een variabel ten opzichte van de voorspelling. Hierdoor krijgt men inzicht in a. de relatieve bijdrage ten opzichte van van de andere variabelen en b. de richting van de variabele. Met andere woorden: zorgen hogere waardes vaker voor een positieve uitkomst? Leeftijd leverde met afstand de belangrijkste bijdrage. Niet onverwacht zorgde een hogere leeftijd voor een hoger risico op overlijden. Verder vonden we ook dat het aantal verschillende medicaties, het stikstofgehalte in het bloedplasma, lactaatwaarden en het zuurstofgehalte in het bloed belangrijk waren voor een goede voorspelling. Een belangrijke kanttekening van contributie-analyses voor machine-learning modellen is dat je er geen conclusies uit kan trekken over het daadwerkelijk risico. Correlatie is immers geen causatie. Deze variabelen kunnen echter wel een goed startpunt vormen voor vervolgonderzoek.

In de zomer van 2020 barste er een flinke discussie los in de maatschappij over of leeftijd als selectiecriterium gebruikt mag worden indien er een gedwongen keuze gemaakt moet worden tussen patiënten. Veel medici zeggen van wel; leeftijd is een medisch relevant risico voor veel ziektes en aandoeningen. Denk bijvoorbeeld aan Parkinson of dementie. De politiek is het hier niet mee eens, omdat de grondwet voorschrijft dat er niet gediscrimineerd mag worden op leeftijd, geslacht, huidskleur of religie. Vanwege deze discussie hebben we het best presenterende model nog eens getraind, maar dan zonder leeftijd. Opvallend was dat de AUC slechts met 0,04 afnam naar 0,78, ondanks dat leeftijd met afstand het meeste bijdroeg aan de voorspelling. Dit is een indicatie dat de voorspellende waarde indirect ook aanwezig is in andere variabelen. Uit onze analyse blijkt dus dat het gebruik van leeftijd lijdt tot een betere voorspelling, maar dat het vanuit medisch perspectief te verantwoorden is om leeftijd niet meer te nemen.

Uiteindelijk is code zwart nooit bereikt. Er hoefde geen keuzes gemaakt te worden tussen twee patiënten die acuut hulp nodig hebben. Wel zijn we meerdere keren dichtbij geweest, en mocht dat nog voorkomen, dan hebben we in ieder geval een extra hulpmiddel. In de tussentijd rest het nog om het model in de praktijk te valideren. Bijvoorbeeld door schaduw te draaien en de uitkomst te vergelijken met de keuze die een arts zou maken. Ook is het belangrijk om te kijken of het model genoeg toevoegt; wellicht voorspelt het model uitstekend, maar is de keuze op basis van ervaring van de arts simpelweg beter. De beste uitkomst is in ieder geval als we altijd in het ongewis blijven, dat het model verstoft op de plank, zonder ooit gebruikt te worden.

Literatuur

Ottenhoff, M. C., Ramos, L. A., Potters, W. et al. (2021). Predicting mortality of individual patients with COVID-19: a multicentre Dutch cohort. *BMJ Open 11.* (e047347). doi: 10.1136/bmjopen-2020-047347

Maarten Ottenhoff is promovendus aan de Universiteit van Maastricht in de vakgroep Neurochirurgie. In zijn onderzoek binnen de neurotechnologie zoekt hij naar nieuwe signaalanalysemethoden voor de ontwikkeling van neuromodulators en brain-computerinterfaces. Door de kritieke COVID-toestand in het ziekenhuis heeft hij tijdelijk zijn aandacht gericht op acute problemen in de COVID-zorg.
E-mail: m.ottenhoff@maastrichtuniversity.nl

Chocoladelettermallen in het Bakkerijmuseum te Medemblik.
Foto: Arch (cc)

OVER NOBELPRIJZEN EN CHOCOLADE

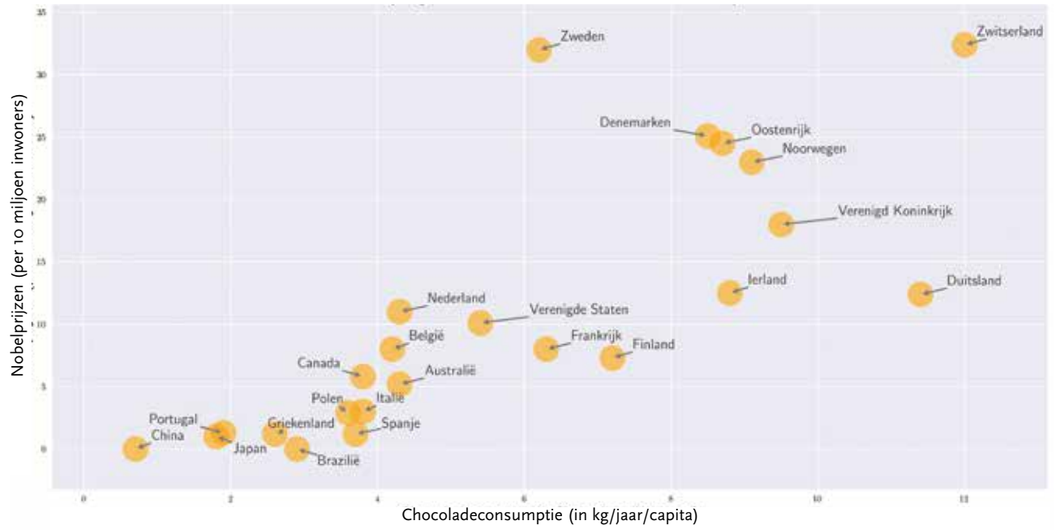
Wat een mooi begin van de week, een Nobelprijs met een Nederlands tintje. Guido Imbens, econometrist en Erasmus Alumnus, kreeg deze week samen met David Card en Joshua Angrist de Nobelprijs voor de Economie voor het baanbrekende onderzoek dat zij hebben verricht op het gebied van de causale inferentie. Daarmee is Imbens, na Jan Tinbergen en Tjalling Koopmans de derde (geboren) Nederlander die de Nobelprijs voor Economie ontvangt.

Conclusies uit gevonden data

Imbens' onderzoek heeft betrekking op het aantonen van causale verbanden met data uit natuurlijke experimenten en is daarmee van grote toegevoegde waarde voor onder andere economisch onderzoek. Bijvoorbeeld onderzoek naar de relatie tussen het aantal opleidingsjaren en inkomen. Het is algemeen bekend dat er een positieve correlatie is tussen de duur van het genoten onderwijs en inkomen. De vraag is of langer in de schoolbanken zitten een hoger inkomen oplevert of dat slimme mensen zowel op school als op het werk succesvol zijn. Het probleem om dit te onderzoeken is dat we nooit de situatie van een individu met minder scholing kunnen vergelijken met de situatie waarin hetzelfde individu meer scholing geniet. Dit is het fundamentele probleem van causale inferentie.

Een natuurlijk experiment is een observationele studie waarbij niet gecontroleerd wordt in de traditionele zin van een gerandomiseerd experiment. De data worden als het ware 'gevonden', en niet gegenereerd met een gecontroleerd experiment. De 'natuur' bepaalt of individuen worden blootgesteld aan de 'behandeling' of tot de controlegroep behoren, buiten de invloed van de onderzoeker. Zo is de invloed van de duur van het genoten onderwijs op inkomen onderzocht door gebruik te maken een bijzondere eigenschap van het Amerikaanse schoolsysteem. Kinderen kunnen in de VS hun opleiding stoppen op hun zestiende verjaardag. Als gevolg zullen kinderen die eind december geboren zijn, vergeleken met kinderen die een paar weken later zijn geboren (in januari), een heel jaar langer onderwijs hebben genoten als ze beide op hun zestiende stoppen. Of de kinderen stopte op hun zestiende bepaalde niet de onderzoeker, maar de 'natuur'. Onderzoek met de aldus verzamelde data toonde aan dat kinderen geboren in januari die stopte met school en dus een jaar minder onderwijs genoten hadden, meetbaar minder verdienden.

Het proberen uit te zoeken of een patroon (langere scholing → hoger salaris) wordt veroorzaakt door wat we denken dat het veroorzaakt is de kern van het werk van een econometrist. Als de data die voor een dergelijk onderzoek gebruikt worden niet zijn gegenereerd door een gecontroleerd proces vertroebeld dat de analyse. Een probleem dat dankzij het onderzoek van Guido Imbens kan worden aangepakt. Imbens ontwikkelde econometrische methoden om te helpen causale verbanden te vinden in data uit natuurlijke experimenten. Als gevolg is economisch onderzoek steeds meer een empirische wetenschap geworden, een vakgebied vol empirische bevindingen, waarvan een belangrijk aantal gevonden zijn dankzij het werk van Imbens.



Nobelprijzen vs chocoladeconsumptie

Kritiek

Toch is de aanpak van Imbens niet zonder kritiek. Het gebruik van data uit natuurlijke experimenten brengt het probleem van reproduceerbaarheid met zich mee. Ook is het proces waarin individuen door de 'natuur' worden toegewezen aan de 'behandeling'- of 'controle'-groepen zelden echt random. Niet waargenomen externe factoren kunnen de selectie beïnvloeden en gecorreleerd zijn met de uitkomst waarnaar onderzoek wordt gedaan. Los van de problemen die samenhangen met het gebruik van data uit natuurlijke experimenten kun je je ook afvragen of data alleen wel voldoende is om een causaal verband aan te tonen. Judea Pearl, schrijver van onder andere *Het boek waarom*, is een fervent tegenstander van de aanpak van Imbens. Het idee dat statistische data voldoende zijn om te bepalen hoe de werkelijkheid in elkaar steekt is in zijn optiek een grote misvatting. Om van een causaal verband te kunnen spreken moet je, aldus Pearl, aantonen dat het verband niet door andere factoren wordt verklaard en dat er een plausibel werkingsmechanisme is. Als voorbeeld noemt hij de correlatie tussen een kraaiende haan en de opkomst van de zon. Om te weten dat de haan kraait omdat de zon opkomt en niet andersom, heb je naast data een model nodig van de wereld. Dat model vertelt welke variabele (haangekraai) wordt beïnvloed door welke andere variabele (stand van de zon), het ordent de relaties tussen oorzaken en gevolgen. Je kunt zo'n model niet halen uit data volgens Pearl.

Chocolade

Als econometrist is het mijn dagelijks werk om chocolade te maken van de data van mijn klanten en beantwoord ik vragen over bijvoorbeeld de factoren die de status van een machine beïnvloeden, of wat de belangrijkste drijvers zijn van de CO_2 footprint van een product. Daarin maak ik gebruik van de technieken die ons vakgebied ons biedt, soms die van Imbens, soms die van Pearl. Gelukkig groeit dat instrumentarium en zorgen kruisbestuivingen tussen bijvoorbeeld machine learning en econometrie voor vernieuwing en verbetering. Of het tot nog een Nederlandse Nobelprijs leidt? Daartoe zal de chocoladeconsumptie in ons land flink omhoog moeten, als het onderzoek van Franz Messerli tenminste klopt. In 2012 schreef hij in de *New England Journal of Medicine* dat hoe hoger de chocoladeconsumptie van een land, hoe meer Nobelprijswinnaars het per hoofd van de bevolking voortbrengt. Wie weet, ik sla alvast wat extra chocoladeletters in.

John Poppelaars, Doing the Math
E-mail: john@doingthemath.nl

Foto: Denizzje cc

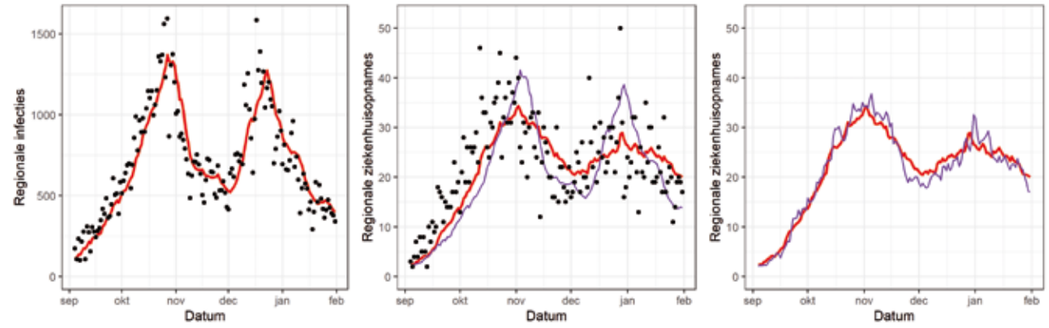
Eerlijke spreiding van coronapatiënten

Een dagelijks terugkerend journaalitem in coronatijd: 'de beschikbare capaciteit voor coronapatiënten in ziekenhuizen is gevuld'. Coronapatiënten worden daarom verplaatst: naar een ziekenhuis in de eigen zorgregio, naar een regio ver van huis en soms zelfs naar het buitenland. Verstandig verplaatsen is van belang voor coronapatiënten en voor eerlijke toegang tot reguliere zorg. Vanuit een accurate voorspelling van het aantal in ieder ziekenhuis opgenomen coronapatiënten ontwikkelden we een beslissingsondersteunend systeem om patiënten eerlijk te spreiden over ziekenhuizen binnen een zorgregio en het aantal bovenregionale verplaatsingen zoveel mogelijk te beperken.

Sander Dijkstra, Stef Baas, Aleida Braaksma & Richard Boucherie

Coronapatiënten worden opgenomen op speciale verpleeg- en intensivecareafdelingen (IC). Hier ingezette zorgverleners worden onttrokken aan de reguliere zorg, resulterend in afgeschaalde reguliere ziekenhuiszorg. Wanneer een corona-afdeling (bijna) vol ligt, worden coronapatiënten verplaatst naar andere ziekenhuizen. Door adequaat voorspellen van de bedbezetting kunnen we vooraf bepalen op welk moment verplaatsingen noodzakelijk zijn en daarbij een keuze maken naar welke ziekenhuizen of regio's patiënten verplaatst worden. Dit is van belang voor coronapatiënten en om beschikbare reguliere zorgcapaciteit te balanceren.

We beschrijven een methode die voor ieder ziekenhuis het aantal bezette coronabedden een aantal dagen vooruit accuraat voorspelt. De gebruikte data bevatten cijfers van het RIVM voor het dagelijkse aantal positieve coronatests per regio, regionale opnamecijfers van de Stichting NICE (Nationale Intensive Care Evaluatie) en



Figuur 1. NAZW. *Links:* regionale infecties (zwart), wekelijks voortschrijdend gemiddelde (rood); *midden*: regionale opnames (zwart), voortschrijdend gemiddelde (rood); vertraagde trend van regionale infecties (paars); *rechts*: voortschrijdend gemiddelde van regionale opnames (rood), voorspelling drie dagen vooruit (paars)

opname-, ontslag- en overplaatsingstijdstippen uit het ZIS (ziekenhuisinformatiesysteem). Hiermee bepalen we voor ieder tijdstip de parameters van een wachtrijmodel, waarmee we de bedbezetting in een ziekenhuis kunnen voorspellen. Deze voorspellingen voor de ziekenhuizen binnen een zorgregio (ROAZ-regio; Regionaal Overleg Acute Zorg) gebruiken we in een *load balancing* methode om coronapatiënten binnen de regio eerlijk te spreiden. Voor bovenregionale verplaatsingen, die nodig zijn als de regionale capaciteit onvoldoende is, beschrijven we een beslissingsondersteunend systeem op basis van een stochastisch *Mixed Integer Program* (MIP), dat patiënten eerlijk over zorgregio's spreidt, met voorkeur voor verplaatsingen over een korte reisafstand.

Regionale opnames

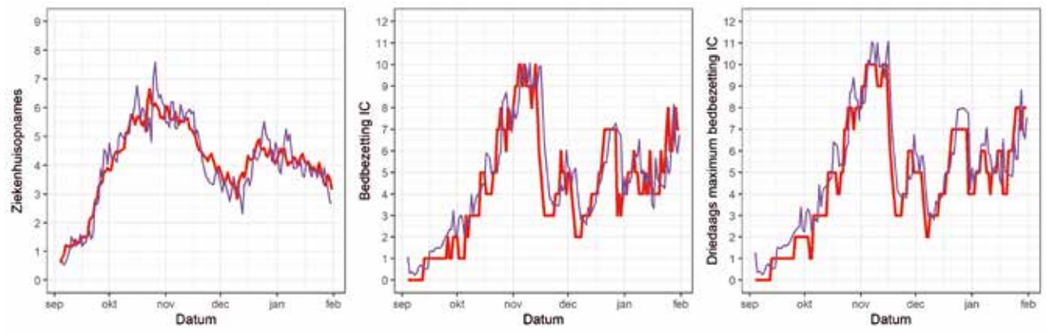
Voor de ROAZ-regio Netwerk Acute Zorg West (NAZW), met daarin het Leids Universitair Medisch Centrum (LUMC), HagaZiekenhuis en Groene Hart Ziekenhuis (GHZ), hebben we data voor de periode van 4 september 2020 tot 31 januari 2021 (begin tweede golf) gebruikt om onze methode te ontwikkelen en testen.

Figuur 1 (links) toont voor NAZW het dagelijkse aantal positieve tests (zwart), en het wekelijks voortschrijdend gemiddelde (rood). We zien duidelijke pieken eind oktober en medio december. Figuur 1 (midden) toont het aantal ziekenhuisopnames per dag (zwart), en het exponentieel gewogen voortschrijdend gemiddelde (rood). Een stijging van het aantal positieve tests resulteert enkele dagen later in een stijging van het aantal ziekenhuisopnames. We passen de kleinste-kwadratenmethode toe om de minimum afwijking te bepalen tussen het voortschrijdende gemiddelde aantal ziekenhuisopnames en verschoven grafieken van het wekelijks voortschrijdende gemiddelde aantal positieve tests. We vinden de minimum afwijking bij een verschuiving van zeven dagen (paarse lijn in figuur 1 midden). Het percentage opgenomen patiënten en een voorspelling bepalen we op basis van op dat moment beschikbare data. Figuur 1 (rechts) toont het drie dagen vooruit voorspelde (paars) en het gerealiseerde (rood) aantal ziekenhuisopnames; de voorspelling voor 3 oktober is dus gemaakt met beschikbare data tot 1 oktober. We concluderen dat onze methode het aantal regionale ziekenhuisopnames accuraat voorspelt.

Aantal bezette bedden per ziekenhuis

De corona-afdeling van een ziekenhuis modelleren we als een netwerk van twee *infinite-server queues* waarin coronapatiënten arriveren, worden opgenomen in één van de twee wachtrijen (verpleegafdeling en IC), daar hun ligduur verblijven, kunnen worden overgeplaatst tussen de afdelingen en worden ontslagen of overlijden. De kracht van dit model is dat accurate voorspelling van het aantal bezette bedden mogelijk is door simulatie van het onderliggende *Poisson Arrival Location Model* (PALM), zie kader.

Figuur 2 (links) toont voor Haga het gerealiseerde (rood) en voorspelde (paars) aantal opnames. De voorspelling is bepaald via de beste kleinste-kwadratenfit tussen de ZIS-opnamecijfers en de voorspelde regionale opnamecijfers. Het voorspelde aantal opnames bepaalt de in de tijd variërende intensiteit van een Poissonproces. Het ZIS bevat rechts-gecensureerde ligduurdata vanwege de onvoltooide ligduur van nog aanwezige patiënten. Voor de ligduurverdelingen gebruiken we hierom de Kaplan-Meier-schatter. Vanuit de ZIS-data schatten we ligduurafhankelijke overplaatskansen tussen



Figuur 2. Haga. *Links*: exponentieel gewogen voortschrijdend gemiddelde (rood) en voorspelde (paars) ziekenhuisopnames; *midden*: gerealiseerde (rood) en voorspelde (paars) IC-bezetting; *Rechts*: gerealiseerde (rood) en voorspelde (paars) maximum IC-bezetting

verpleegafdeling en IC. Hiermee zijn alle voor het PALM benodigde data beschikbaar.

Met het PALM voorspellen we een aantal dagen vooruit de bedbezetting op verpleegafdeling en IC op die dag om 10:00 uur (rapportagetijd van vrije beschikbare bedden door ziekenhuizen). Figuur 2 (midden) toont de gerealiseerde (rood) en drie dagen vooruit voorspelde (paars) IC-bezetting. Figuur 2 (rechts) toont de voorspelde en gerealiseerde maximum IC-bezetting gedurende de volgende drie dagen. Deze maat is beleidsondersteunend: indien het voorspelde maximum de capaciteit overschrijdt, kan het ziekenhuis meer bedden beschikbaar stellen of een patiëntenstop afkondigen. Onze voorspelmethode blijkt bijzonder accuraat, en is opgenomen in het NAZW-coronadashboard.

Eerlijk spreiden binnen een zorgregio

We combineren de voorspelde bedbezetting per ziekenhuis (figuur 2 rechts) en het voorspelde aantal benodigde regionale opnames (figuur 1 rechts) om patiënten eerlijk te spreiden over de ziekenhuizen binnen een zorgregio. Door simulatie van het PALM vergelijken we de spreiding van patiënten volgens onze dynamische *load balancing* regel met een statische regel die een vaste fractie patiënten toewijst aan elk ziekenhuis. Figuur 3 toont deze vergelijking voor NAZW.

Het historisch aandeel coronapatiënten opgenomen in LUMC is 30%, in Haga 43% en in GHZ 27%. Opnemen van patiënten volgens deze statische fracties geeft de bezetting op verpleegafdeling en IC weergegeven in figuur 3 (rechts). We zien voor LUMC en GHZ een onderbezetting en in Haga grote excursies van de IC-bezetting boven de capaciteit, terwijl de bezetting van de verpleegafdeling rond de capaciteit blijft. Patiënten worden niet tussen de ziekenhuizen verplaatst, maar verondersteld opgevangen te worden in zogenaamde *overbedden*, die tijdelijk extra beschikbaar gemaakt worden voor coronapatiënten.

Onze *load balancing* regel (figuur 3 links) neemt patiënten op in ziekenhuizen gebaseerd op veiligheidsniveaus, zie kader. We zien dat excursies boven de aanwezige capaciteit in een ziekenhuis worden opgevangen door opname van patiënten in andere ziekenhuizen. De korte excursies op de IC in GHZ in december wordt bijvoorbeeld gedempt door opname van meer patiënten in LUMC. *Overbedden* worden vrijwel niet ingezet. Onder de *load balancing* regel kunnen vrijwel alle patiënten worden opgenomen binnen de regio.

Poisson Arrival Location Model (PALM)

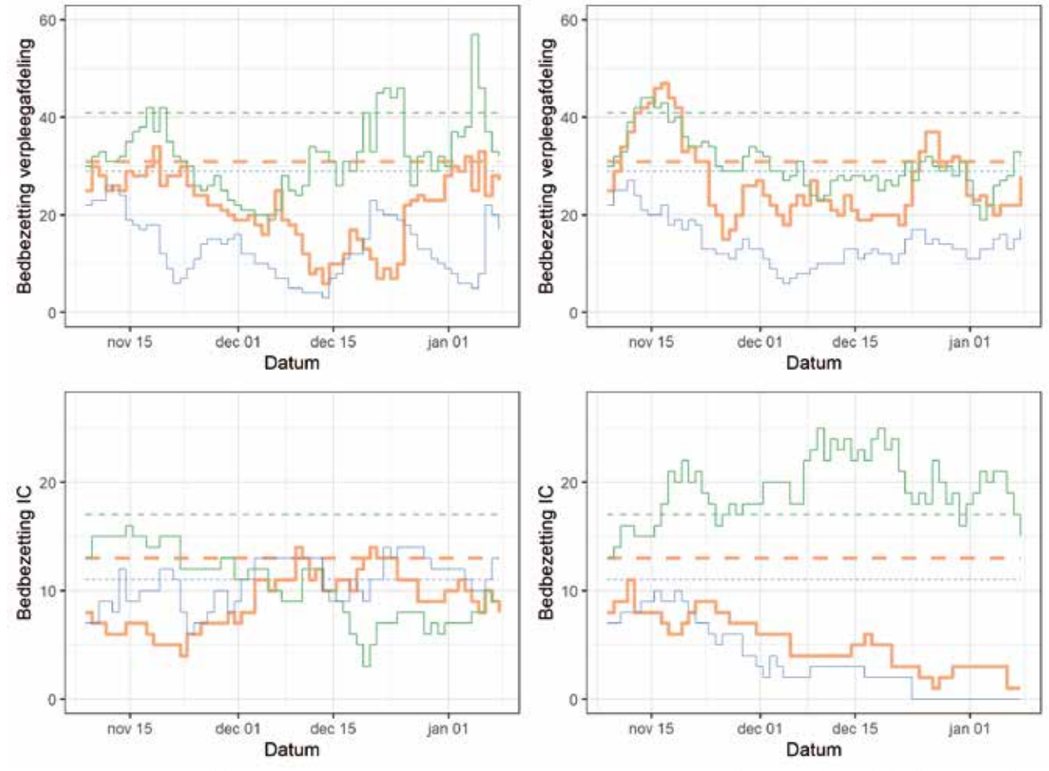
Een netwerk van *infinite-server queues* met in de tijd variërende intensiteit van het Poisson aankomstproces en algemene ligduurverdeling heeft als belangrijke eigenschap dat het aantal aanwezige patiënten, zeg op verpleegafdeling en IC, op ieder tijdstip s Poisson verdeeld is. Het PALM beschrijft de *sample-paden* van dit proces. Vanuit iedere huidige toestand (aangeduid met $l(s)$ op tijdstip s, waarin opgenomen het aantal patiënten, hun gerealiseerde ligduur en eventueel gerealiseerde overplaatsingen) van het netwerk kan het PALM voor empirische verdelingen van de onderliggende stochastische variabelen middels simulatie worden gegenereerd. Hiermee kunnen op ieder tijdstip s relevante prestatiematen worden bepaald, zoals het verwachte aantal aanwezige patiënten t dagen vooruit en de verwachte maximale bedbezetting in $[s, s + t]$:

$$\mathbb{E}[N_V(s+t) \mid \boldsymbol{\ell}(s)], \qquad \mathbb{E}[N_I(s+t) \mid \boldsymbol{\ell}(s)],$$

$$\mathbb{E}\left[\max_{u\in[s,s+t]} N_V(u) \mid \boldsymbol{\ell}(s)\right], \qquad \mathbb{E}\left[\max_{u\in[s,s+t]} N_I(u) \mid \boldsymbol{\ell}(s)\right].$$

Dynamische load balancing voor regionale spreiding

Voor ziekenhuis h in een regio bepalen we de maximaal op te nemen fractie $\Theta_h(s,t)$ van de regionaal benodigde opnames (rechter grafiek figuur 1), zodat de kans op voldoende capaciteit tussen tijdstippen s en $s + t$ minimaal veiligheidsniveau α_{hV} is op de verpleegafdeling en α_{hI} op de IC. Indien $\sum_h \Theta_h(s,t) \geq 1$, dan verwachten we dat de regio geen *overbedden* nodig heeft in $[s,s + t]$ en neemt ziekenhuis h het proportionele deel $\Theta_h,(s,t)/\sum_h \Theta_h(s,t)$ op. Indien $\sum_h \Theta_h(s,t) < 1$, dan is de regionale capaciteit onvoldoende en neemt ziekenhuis h zijn eigen fractie $\Theta_h(s,t)$ op. In dit geval wordt de resterende fractie $1-\sum_h \Theta_h(s,t)$ geplaatst op een *overbed* en daarmee aangemerkt voor bovenregionale verplaatsing.



Figuur 3. NAZW. LUMC (*blauw*), Haga (*groen*), GHZ (*oranje*); aanwezige capaciteit (*stippellijn*), gerealiseerde bedbezetting (*doorgetrokken lijn*); load balancing (*links*), statische fracties (*rechts*)

Bovenregionale spreiding

In de praktijk is langdurig inzetten van *overbedden* onmogelijk. Patiënten worden verplaatst binnen een regio, bovenregionaal of naar het buitenland. Ons beslissingsondersteunend systeem voor bovenregionale spreiding op basis van een stochastisch MIP houdt rekening met bezetting van de verpleegafdelingen en IC's op het beslismoment en de voor de komende dagen voorspelde (maximum) bezetting op basis van het PALM. Dit laatste is van belang om te voorkomen dat patiënten tussen regio's heen en weer verplaatst worden, bijvoorbeeld doordat verplaatsing van een patiënt naar een regio enkele dagen later resulteert in overbezetting van die regio. De te minimaliseren doelfunctie bevat reisafstand, transportkosten en balans in het aantal vrije bedden over de regio's.

Figuur 4. Aantal verplaatsingen op verpleegafdelingen (*boven*) en IC's (*onder*); bedden vrij (*blauw*), beslissingsondersteunend systeem (*rood*)

Figuur 4 toont *boxplots* voor de vergelijking van onze verplaatsingsstrategie met de strategie 'bedden vrij', die zonder rekening te houden met reisafstand patiënten verplaatst naar regio's waar na het moment van beslissing nog minimaal twee bedden vrij zijn. Een cluster bestaat uit een aantal nabijgelegen regio's. De prestatiematen totaal aantal verplaatsingen, verplaatsingen naar het buitenland, binnen de clusters, en tussen de clusters, zijn bepaald op basis van het PALM. Het totaal aantal bovenregionale verplaatsingen verschilt niet significant. We zien een significante reductie in het aantal verplaatsingen naar verder gelegen regio's.

Conclusie

Statistiek en Operations Research bieden ondersteuning om patiënten tijdens een pandemie eerlijk te spreiden en reguliere zorg gebalanceerd te leveren.

Informatie over benodigde data en software is beschikbaar in de referenties en via www.utwente.nl/en/choir/research/Covid19-wardICU/.

Literatuur

Baas, S., Dijkstra, S., Braaksma, A., Van Rooij, P., Snijders, F. J., Tiemessen, L., & Boucherie, R. J. (2021) Real-time forecasting of COVID-19 bed occupancy in wards and Intensive Care Units. *Health Care Management Science* 24, 402–419.

Dijkstra, S., Baas, S., Braaksma, A., & Boucherie, R.J. (2021) Dynamic fair balancing of COVID-19 patients over hospitals based on forecasts of bed occupancy. Onder review.

Sander Dijkstra (s.dijkstra-1@utwente.nl), Stef Baas (s.p.r.baas@utwente.nl), Aleida Braaksma (a.braaksma@utwente.nl) en Richard Boucherie (r.j.boucherie@utwente.nl) zijn onderdeel van het Center for Healthcare Operations Improvement and Research (CHOIR) aan de Universiteit Twente. Op basis van toonaangevend wetenschappelijk onderzoek en in nauwe samenwerking met de praktijk ontwikkelt en implementeert CHOIR logistieke oplossingen in de zorg.

Jelke Bethlehem

column

PEILINGPRAKTIJKEN

Wilde de meerderheid wel thuisblijven met oud en nieuw?

Er wordt veel gepeild in Nederland. Dat zie je vooral tijdens verkiezingscampagnes. Snelle politieke peilingen van marktonderzoekbureaus volgen elkaar dan in hoog tempo op. Ook andere organisaties, zoals de landelijke overheid (CBS, SCP), peilen. Die peilingen zijn vaak groter en ingewikkelder dan de kleine en snelle politieke peilingen van de marktonderzoekbureaus. Ze worden meestal enquêtes of surveys genoemd.

Goed en slechte peilingen

Er zijn goede en slechte peilingen. Een goede peiling is gebaseerd op een aselecte steekproef. Dat levert een representatieve steekproef op, je schattingen zijn valide en je kunt ook de precisie van je schattingen bepalen (in de vorm van betrouwbaarheidsintervallen of onzekerheidsmarges). Dus je kunt de uitkomsten van je steekproef generaliseren naar de doelgroep van je onderzoek.

Er zijn helaas ook slechte peilingen. Die zijn niet gebaseerd op een aselecte steekproef maar op een vorm van zelfselectie. Zelfselectie betekent dat je het aan de personen in de doelgroep overlaat om zichzelf wel of niet aan te melden voor de peiling. Via berichten in de media maak je ze attent op de peiling. Je hoopt dat ze zich daardoor laten overhalen om deel te nemen aan de peiling. In de steekproef zitten dan vooral mensen die het leuk vinden om regelmatig aan peilingen mee te doen en geïnteresseerd zijn in het onderwerp van het onderzoek. Dat levert meestal geen representatieve steekproef op. En dus trek je verkeerde conclusies uit je onderzoek. De peilingen van *Stand.nl* zijn een duidelijk voorbeeld van hoe het niet moet.

De peilingen van *Standpunt.nl*

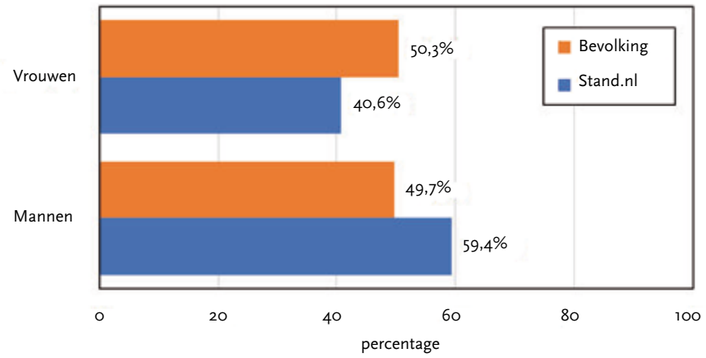
Stand.nl is een radioprogramma dat dagelijks (op werkdagen) te beluisteren valt op NPO Radio 1. In het programma komt dagelijks een actuele kwestie aan de orde met een actuele stelling. De luisteraars kunnen over de stelling online of per telefoon stemmen ('eens' of 'oneens'). Als peiling zit *Stand.nl* niet goed in elkaar. Het is geen representatieve peiling en dus kun je de uitkomsten niet generaliseren naar de hele bevolking. We illustreren de tekortkomingen van deze peiling aan de hand van een voorbeeld van zo'n peiling.

Op 30 december 2020 was de stelling 'Iedereen moet uit solidariteit thuisblijven met oud en nieuw'. Nadat bijna 1.300 respondenten gereageerd hadden, bleek dat 56% voor de stelling was. De vraag is nu of je deze uitslag kunt generaliseren naar de hele bevolking. Kun je zeggen dat 56% van de bevolking voorstander is van thuisblijven met oud en nieuw? Helaas, dat kan niet want er is geen sprake van een representatief onderzoek. De deelnemers aan deze peiling vormen geen goede afspiegeling van de bevolking.

De representativiteit van de peiling

De peilers van *Stand.nl* vragen de respondenten niet alleen naar hun mening over de stelling. De respondenten moeten ook hun geslacht en leeftijd opgeven. Er zijn vier leeftijdsklassen: 0-20 jaar, 20-40 jaar, 40-60 jaar en 60 jaar en ouder. Je kunt dus vaststellen hoeveel mannen en vrouwen er in de peiling zitten en of de percentages mannen en vrouwen in de peiling overeen komen met de percentages mannen en vrouwen in de hele bevolking.

In figuur 1 en 2 is die vergelijking gemaakt. De oran-



Figuur 1. De percentages mannen en vrouwen in de bevolking en in de peiling; vrouwen ondervertegenwoordigd in de peiling van *Standpunt.nl*

je staven in figuur 1 geven de percentages mannen en vrouwen in de bevolking weer. Duidelijk is te zien dat er ongeveer evenveel mannen als vrouwen zijn. De blauwe staven staan voor de percentages mannen en vrouwen in de peiling. Er is iets mis hier. Er zitten te weinig vrouwen in de peiling (40,6% in plaats van 50,3%) en er zitten teveel mannen in (59,4% in plaats van 49,7%). In beide gevallen is er sprake van een afwijking van ongeveer 10 procentpunten.

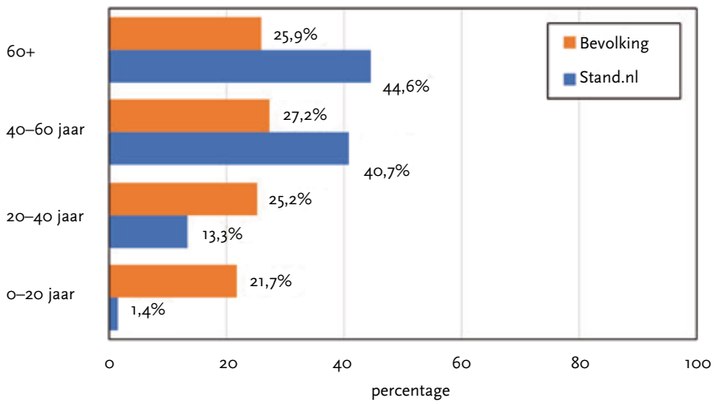
We kunnen ook voor het kenmerk leeftijd de verdeling in de bevolking vergelijken met die in de peiling. Het resultaat staat in de figuur 2 hieronder. De problemen met de representativiteit blijken ook hier groot te zijn.

De ouderen zijn zwaar oververtegenwoordigd in *Stand.nl*. In de peiling is 44,6% 60 jaar of ouder en in de bevolking als geheel is dit maar 25,9%. Een afwijking van 18,7 procentpunten dus! Ook de middelbaren tussen de 40 en 60 jaar zijn behoorlijk oververtegenwoordigd. Er is hier sprake van een afwijking van 12,8 procentpunten.

Bij de jongeren zie je het omgekeerde effect. Die jongeren (0-20 jaar en 20-40 jaar) zijn juist ondervertegenwoordigd in de peiling. De situatie is vooral ernstig voor de jongeren in de groep 0-20 jaar. De peiling zou voor 21,7% (één op de vijf) uit deze jongeren moeten bestaan, maar het is slechts een minimaal percentage van 1,4%.

Mannen zijn dus oververtegenwoordigd in de peiling en vrouwen zijn ondervertegenwoordigd. Dat betekent dat de mening van vrouwen onvoldoende doorklinkt in de resultaten. Ook zijn ouderen oververtegenwoordigd en jongeren ondervertegenwoordigd. De mening van jongeren klinkt dus onvoldoende door. We moeten concluderen dat deze peiling niet representatief is en we dus de uitkomsten niet kunnen vertalen naar de hele bevolking.

Waarom wijkt de samenstelling van de peiling zo af van die van de populatie? Dat komt omdat de steekproef niet op correcte wijze is getrokken. Bij voorkeur trek je als onderzoeker een aselecte steekproef. Daarbij heeft iedereen dezelfde kans om in de steekproef te komen. Je steekproef is dan dus representatief. Bij Stand.nl wordt gebruikt gemaakt van zelfselectie. Iedereen die dat wil, kan meedoen. Je krijgt alleen mensen in de steekproef die spontaan besluiten mee te doen. Die dus gemotiveerd en geïnteresseerd zijn. Het zijn vooral mensen die elke ochtend naar het radioprogramma luisteren. Dat leidt niet tot een representatieve steekproef.



Figuur 2. De verdeling van de leeftijden in de bevolking en in de peiling; jongeren ondervertegenwoordigd in de peiling van *Standpunt.nl*

Correctie door weging

Als een peiling niet representatief is, kun je proberen een correctie uit te voeren. Dat kan via een procedure die we wegen noemen. De mensen van *Stand.nl* hebben dat helaas niet gedaan. Het zou echter wel kunnen. We schetsen hier hoe je dit zou kunnen doen.

Omdat we verdeling naar geslacht en de leeftijdsverdeling in de peiling en in de bevolking kennen, kunnen we gewichten bepalen die corrigeren voor onder- of oververtegenwoordiging van bepaalde groepen. Als we dat doen, dan krijgen bijvoorbeeld jonge vrouwen (0-20 jaar) een groot gewicht van ruim 22,91. En oude mannen (60+) krijgen juist een klein gewicht van 0,43. Zo wordt de zaak, althans voor wat betreft geslacht en leeftijd weer rechtgetrokken.

Als je nu opnieuw het percentage voorstanders van thuisblijven uitrekent (maar dan gewogen), kom je uit op 39%. Dat is een stuk lager dan het ongewogen percentage van 56%! We kunnen verwachten dat 39% een betere benadering is van het percentage voorstanders in de bevolking. Het oorspronkelijke percentage was dus veel te rooskleurig. Er is geen meerderheid van 56% voorstanders maar een meerderheid van 61% tegenstanders!

De mensen van *Stand.nl* vroegen alleen naar geslacht en leeftijd. Daarom kunnen we alleen voor deze twee variabelen het gebrek aan representativiteit analyseren en corrigeren. Je kunt echter verwachten dat nog meer kenmerken een rol spelen. Je zou bijvoorbeeld kunnen denken aan kenmerken als opleidingsniveau en stemgedrag bij de vorige verkiezingen. Je kunt die kenmerken echter alleen gebruiken als je ze meet in de peiling. Helaas is dit niet gebeurt.

We kunnen concluderen dat de uitkomsten van deze peiling geen goede afspiegeling zijn van wat er leeft in de bevolking. Het zou goed zijn als de redactie van *Stand.nl* daarvan melding maakte bij de presentatie op de radio en de verslaggeving op de website. Ook zou de redactie kunnen overwegen om gewogen uitkomsten te publiceren in plaats van ongewogen uitkomsten. Kortom, er is ruimte voor verbetering!

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

IM Constance van Eeden

Op 21 september 2021 is Constance van Eeden op 94-jarige leeftijd overleden. Zij was erelid van onze vereniging en een van de Nederlandse grondleggers van de statistiek.

Professor Van Eeden werd in 1927 in Delft geboren. Haar carrière begon bij het CWI in Amsterdam bij David van Dantzig en Jan Hemelrijk. In 1958 behaalde zij haar doctorstitel met van Dantzig als promotor. Van Eeden was de eerste vrouw die in Nederland in de statistiek promoveerde, en het duurde maar liefst 29 jaar voordat een volgende vrouw dit zou doen.

Na haar promotie verhuisde ze naar de Verenigde Staten, waar ze aan de University of Minnesota werkzaam was, en daarna aan de Canadese Université de Montréal, tot aan haar pensioen in 1989. In Canada heeft ze een grote rol gespeeld in de ontwikkeling van de statistiek in het Franstalige deel van dat land.

Haar onderzoek richtte zich op niet-parametrische statistiek en beslistheorie. Het boek *A nonparametric introduction to statistics*, dat zij in 1968 samen met haar man Charles Kraft schreef, is voor veel VVSOR-leden een eerste introductie in niet-parametrische technieken geweest.

In 2013 werd Van Eeden erelid van de VVSOR. Hiernaast was zij fellow van de American Statistical Assocation en het Institute of Mathematical Statistics, ontvanger van de Henri Willem Methorst medal van het International Statistical Institute en de Gold Medal van de Statistical Society of Canada, waarvan ze ook erelid was. Het CWI zal in haar naam een Constance van Eeden PhD Fellowship oprichten voor talentvolle vrouwelijke wetenschappers.

Na haar pensioen keerde ze terug naar Nederland. Ondanks haar hoge leeftijd, bleef ze trouw de *annual meetings* van de VVSOR bezoeken. De VVSOR is Constance van Eeden dankbaar voor haar grote bijdragen aan de statistiek. In het komende nummer van *STAtOR* zal uitgebreid bij haar werk en haar leven worden stilgestaan.

Patronen in protest tijdens pandemieën

COVID-19 zorgt voor veel onzekerheid en onrust en maakt kwetsbaarheden in de samenleving beter zichtbaar. Onzekerheden over het verloop van pandemie, over het eigen inkomen en de eigen gezondheid en over de effectiviteit van de maatregelen kunnen voor veel spanningen zorgen en een toename van het aantal protesten tegen overheidsmaatregelen. In dit artikel wordt onderzocht hoe data-analyse kan worden gebruikt om de relatie tussen regelgeving, sociaaleconomische dynamiek en het ontstaan van protestbewegingen te duiden.

Koen van der Zwet, Bas Keijser

De weerbaarheid van samenlevingen is door de COVID-19-pandemie op de proef gesteld. De pandemie heeft uiteenlopende maatschappelijke effecten, waaronder op de economie, sociale omgang en mentale gesteldheid van de bevolking. Hierbij spelen een rol: onzekerheid over de epidemische verspreiding, de ernst van de ziekte, en de effectiviteit en legitimiteit van de maatregelen om de verspreiding te voorkomen. Deze onzekerheden veroorzaken spanningen die de stabiliteit van de samenleving onder druk zetten. De verstoring van deze stabiliteit is te zien in de toename van het aantal protesten tegen overheidsbeleid (Kishi, 2021). De toename van sociale onrust tijdens pandemieën kent historische parallellen (Censolo & Morelli, 2020): ook tijdens recente uitbraken van Ebola en SARS traden rellen, stigmatisering en plunderingen op in bijvoorbeeld West-Afrika en Canada.

De toename van protest in Nederland en andere landen tijdens de huidige COVID-19-pandemie biedt een venster om het ontstaan van sociale onrust beter te begrijpen. De drijfveren die ten grondslag liggen aan deze toename lijken direct te relateren aan de indammende maatregelen en sociaaleconomische effecten van de pandemie. Het moment waarop maatschappelijke effecten van de pandemie en protesten optraden, kan mogelijk meer vertellen over de relatie tussen deze factoren. Daarnaast dient de ontwikkeling van kennis over drijfveren van protest een maatschappelijk belang: in een volgende pandemie kan de impact van de pandemie door overheidsingrijpen adequater worden beperkt.

Patronen in protest: moment en intensiteit

Ons onderzoek richt zich op twee specifieke patronen in protest. Op welk moment treedt protest op? En: wat is de intensiteit van protest tijdens pandemieën? Om deze vragen te beantwoorden zijn kwalitatieve en kwantitatieve methoden gecombineerd. In een voorbereidende literatuurstudie zijn verklarende factoren voor het optreden van sociale onrust geïdentificeerd. De factoren zijn ten eerste geclassificeerd op basis van drie fundamentele theorieën in de studie van sociale bewegingen. De eerste theorie beschrijft het ontstaan van grieven op individueel en groepsniveau als oorzaak van de beslissing deel te nemen aan een protestbeweging (Cederman et al., 2010). De tweede theorie beschrijft politieke gelegenheid als voorwaarde voor succesvolle actie om politieke invloed te verkrijgen (Meyer, 2004). Ten slotte beschrijft de theorie van mobilisatie hoe het proces tot verkrijgen van potentieel voor een sociale opstand werkt (Klandermans & Oegema, 1987).

De gevonden factoren zijn ook geordend op basis van de tijdschaal waarop verandering optreedt. Een duidelijk voorbeeld hiervan is het verschil tussen de tijdschaal waarin mitigerende interventies zoals een *lockdown* worden toegepast ('snel', verandering in dagen of weken) en de tijdschaal waarop de demografische opbouw van de populatie verandert ('langzaam', verandering in jaren). Er zijn uiteenlopende datasets geïdentificeerd als *proxy* variabele voor verklarende factoren uit de literatuur.

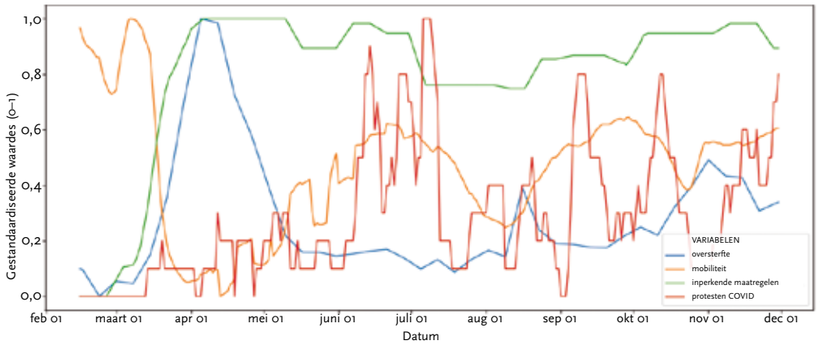
In dit artikel lichten we toe hoe met statistische analyse een deelantwoord op de onderzoeksvragen is gevonden. De beschikbare data over langzaam en snel veranderende factoren en optreden van protest werpen twee interessante vragen op. Enerzijds: met welk statistisch model kunnen relaties tussen voorgaande factoren worden beschreven? Anderzijds: welk type factoren kan het optreden van protesten over de tijd het beste verklaren? Verschillende statistische modellen zijn hiertoe met elkaar vergeleken. Ten slotte schetsen we in de afsluitende discussie hoe in aanvulling op statistische analyse theorie- en modelvorming moet worden ingezet om protest en drijfveren van protest nauwkeurig te begrijpen.[1]

Databronnen

Om de hierboven beschreven analyse uit te voeren zijn data van *The Armed Conflict Location & Event Data Project* (ACLED) gebruikt. Deze organisatie biedt een gevalideerde dataset van alle protestevenementen die wereldwijd

VARIABELE	IN – OUT	MODEL	DYNAMIEK	TIJDSCHAAL	RESOLUTIE	BRON
Protesten COVID	Output	1 - 8	-	Snel	Dagelijks	ACLED
Oversterfte	Input	1 - 8	-	Gemiddeld	Wekelijks	The Economist
Populatie	Input	1 - 8	-	Langzaam	Jaarlijks	UN
Werkloosheid	Input	1 , 4	Grieven	Gemiddeld	Maandelijks	OECD
Grieven in de samenleving	Input	5 , 8	Grieven	Langzaam	Jaarlijks	FSI
Staatsgeweld tegen burgers	Input	3	Mobilisatie	Snel	Dagelijks	ACLED
Populatie tussen 15-24 jaar, %	Input	7 , 8	Mobilisatie	Langzaam	Jaarlijks	UN
Populatie woonachtig in steden, %	Input	7	Mobilisatie	Langzaam	Jaarlijks	World Bank
Mobiliteitsdata	Input	2 , 4	Gelegenheid	Snel	Dagelijks	Google
Inperkende maatregelen (COVID)	Input	2 , 4	Gelegenheid	Snel	Dagelijks	OxCGRT
Economische steun maatregelen	Input	2 , 4	Gelegenheid	Snel	Dagelijks	OxCGRT
Investeringen gezondheidszorg (COVID)	Input	2	Gelegenheid	Snel	Dagelijks	OxCGRT
Bruto nationaal product	Input	6	Gelegenheid	Gemiddeld	Per kwartaal	OECD
Politieke rechten	Input	6	Gelegenheid	Langzaam	Jaarlijks	FH
Burgerlijke vrijheden	Input	6	Gelegenheid	Langzaam	Jaarlijks	FH
Legitimiteit van de staat	Input	6 , 8	Gelegenheid	Langzaam	Jaarlijks	FSI

Tabel 1. Variabelen in de dataset, afkortingen: United Nations (UN), Organisation for Economic Co-operation and Development (OECD), Failed State Index (FSI), Oxford COVID-19 Government Response Tracker (OxCGRT), Freedom House (FH)



Figuur 1. Exploratieve vergelijking van selectie van variabelen voor Nederland

plaatsvinden. Met de *COVID-19 Disorder Tracker* heeft ACLED specifiek gecodeerd welke protesten gerelateerd zijn aan de pandemie (Kishi, 2021). We willen patronen in het optreden van protest over de tijd in specifieke landen analyseren. Om optreden van protesten te kunnen verklaren zijn omgevingsfactoren uit de initiële literatuurstudie genomen die het aantal protesten over de tijd kunnen verklaren, het gaat hierbij zowel om algemene drijfveren van protest als drijfveren die specifiek met de COVID-19-pandemie te maken hebben. Zie tabel 1 voor een compleet overzicht van variabelen in de statistische modellen.

Er is gekozen voor een cross-nationale vergelijking om onderscheid te kunnen maken tussen sociaaleconomische condities en mitigerende maatregelen in verschillende landen. Voor 27 landen[2] konden deze condities worden onderscheiden. Doordat niet in elk land de maatregelen op hetzelfde moment ingesteld zijn, zijn er kleine verschillen in de beschikbare hoeveelheid dagen voor analyse. De samengestelde dataset bevat in totaal 7545 datapunten (gemiddeld 280 dagen per land).

Figuur 1 is een exploratieve visualisatie over de tijd van een selectie van de data voor Nederland met gebruikmaking van gestandaardiseerde waarden tussen 0 en 1. De toename van de oversterfte markeert de impact van de pandemie op de volksgezondheid, de mobiliteit in termen van reisbewegingen laat de impact zien van de inperkende maatregelen op het gedrag van de bevolking. Alle waardes zijn opgenomen als voortschrijdend gemiddelde van 7 dagen. In het figuur is te zien dat het aantal protesten tijdens de uitbraak van de pandemie eerst lager wordt, wanneer uitgebreide restricties zijn ingesteld om verspreiding van de epidemie tegen te gaan. Wanneer de door de pandemie veroorzaakte oversterfte afneemt, nemen de protesten toe.

Toepassing van statistische analyse

Met een op regressie gebaseerde analyse is getracht de relatie tussen de omgevingsfactoren en het aantal protesten dat over de tijd plaatsvindt te analyseren. Voor de verklarende, statistische analyse van het aantal protesten kan een model in de klasse van *generalized linear models* worden toegepast (Hilbe, 2011). De protestdata laten een verdeling met een gemiddelde zien dat sterk van de variantie afwijkt – dit is een reden om een regressiemodel met een negatief-binomiale verdeling (NBM) toe te passen. Ook is in de protestdata sprake van oververtegenwoordiging van dagen zonder protesten – dit is een reden om een *zero-inflated* model (ZINB) toe te passen. Dus is een vergelijking gemaakt tussen de twee modellen. Voor het fitten van de modellen is de *iteratively reweighted least squares*-methode toegepast. De fit van de modellen is ten slotte vergeleken aan de hand van Akaike's informatiecriterium en de *log-likelihood* (AIC en LL, Hilbe, 2011). Beide de waardes zijn zo te interpreteren dat een lage waarde duidt op een betere modelfit.

De twee modellen worden met elkaar vergeleken op basis van de eerste vier sets van inputvariabelen. De inputvariabelen zijn geclusterd op basis van tijdsschaal waarop wijzigingen plaatsvinden en op basis van het theoretisch kader (grieven, gelegenheid en mobilisatie). Daarnaast zijn modellen 4 en 8 geoptimaliseerde modellen gecorrigeerd voor multicollineariteit. Tabel 1, kolom drie, geeft een overzicht van variabelen opgenomen in ieder van de modellen. Tabel 2 geeft een overzicht van de modellen en de bijbehorende scores voor de modelfit. Omdat voor de langzame variabelen maar één of een beperkt aantal meetmomenten beschikbaar zijn, kunnen voor model 5 tot en met 8 alleen NBM-modellen worden gefit.

Er is een aantal interessante observaties te doen in de modelfitscores van de verschillende modellen. Ten eerste leveren van de eerste vier modellen zowel ZINB-model 4 als NBM-model 4 de beste waardes op. ZINB-model 4 scoort beter dan NBM-model 4. Dit kan gezien worden als een bevestiging van de analyse dat twee verschillende processen verantwoordelijk zijn voor enerzijds het uitbreken van protesten en anderzijds de intensiteit van protesten. Daarnaast valt op dat NBM-model 8 beter scoort dan

MODEL			SCORE	ZINB	NBM
SNEL	1	Grieven	AIC	15782,1	16074,4
			LL	-7888,0	-8034,2
	2	Gelegenheid	AIC	14287,3	15018,0
			LL	-7133,7	-7499
	3	Mobilisatie	AIC	15881,4	16021,6
			LL	-7937,7	-8007,8
	4	Combinatie	AIC	14227,4	14654,9
			LL	-7104,7	-7318,5
LANGZAAM	5	Grieven	AIC	-	15240,7
			LL	-	-7615,4
	6	Gelegenheid	AIC	-	14802,9
			LL	-	-7395,4
	7	Mobilisatie	AIC	-	15179,5
			LL	-	-7584,7
	8	Combinatie	AIC	-	13872,2
			LL	-	-6431,1

Tabel 2. Vergelijking van fit van de modellen op basis van AIC en log-likelihood

ZINB-model 4, in termen van modelfit afgelezen door het AIC. Dit kan erop duiden dat langzame processen zoals het ontstaan van grieven en demografische verandering meer invloed hebben op de intensiteit van de protesten dan snelle verandering in *lockdown*-maatregelen. Tenslotte hebben alle modellen een relatief lage score. Dit duidt erop dat een groot deel van de variantie in de data niet kan worden verklaard met dergelijke statistische modellen.

Belang van multimethode-aanpak

Uit bovenstaande analyse blijkt dat met geavanceerde statistische analyse slechts een gedeeltelijke verklaring voor de relatie tussen omgevingsfactoren en patronen in optreden van protesten tijdens pandemieën kan worden gevonden. Bovendien wordt uit een input-outputanalyse met een statistisch model niet duidelijk hoe onderliggende processen van grieven, politieke gelegenheid en mobilisatie precies bepalen op welk moment welke protesten optreden. Ook spelen waarschijnlijk interactie-effecten tussen omgevingsfactoren, al bestaande grieven, nieuwe grieven over de pandemie en overheidsoptreden, nieuwe politieke gelegenheid en mobilisatie van specifieke groepen een rol in het ontstaan van protest. Precies toen de strengste *lockdown*-maatregelen in juni 2020 werden afgeschaald, ontstond in Nederland protest. De theorieën over sociale bewegingen uit de literatuur bieden aanvullende componenten van begrip van het pad naar protesten – echter is ook niet *a priori* duidelijk welke van de genoemde theorieën de grootste verklarende kracht heeft.

Daarom roepen we op complexe maatschappelijke en sociale processen te onderzoeken met een multimethode-aanpak. Deze aanpak moet elementen van statistische analyse en data-analyse combineren met theorievorming en modelvorming. Zo kunnen bestaande theorieën en modellen van sociale bewegingen rond grieven, gelegenheid en mobilisatie worden getoetst met empirische data. En kunnen empirische observaties in samenhang worden geduid met eerder uitgevoerde theorie- en modelstudies. Daarnaast is het van belang elementen van psychologische theorieën rond emotionele besmetting en groepsdynamiek te betrekken. Met een multimethode-aanpak is tastbaar begrip op te bouwen van het pad naar protest tijdens pandemieën. Dat begrip is nodig om patronen in protest te onderkennen en te mitigeren.

Noten

1. In een separaat *working paper* wordt een uitgebreide versie van de hier beschreven analyse gegeven. Dit artikel is op aanvraag beschikbaar bij de auteurs.
2. België, Chili, Colombia, Denemarken, Duitsland, Estland, Finland, Frankrijk, Griekenland, Hongarije, Italië, Letland, Luxemburg, Mexico, Nederland, Noorwegen, Oostenrijk, Polen, Portugal, Slovenië, Slowakije, Spanje, Tsjechië, Verenigde Staten, Zuid-Korea en Zweden.

Literatuur

Cederman, L. E., Wimmer, A., & Min, B. (2010). Why do ethnic groups rebel? New data and analysis. *World Politics, 62*(1), 87–119.

Censolo, R., Morelli, M. (2020). COVID-19 and the potential consequences for social stability. *Peace Economics, Peace Science and Public Policy, 26*(3). 20200045.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

Kishi, R. (2021). *A Year of COVID-19: The pandemic's impact on global conflict and demonstration trends*. Beschikbaar via https://acleddata.com/acleddatanew/wp-content/uploads/2021/04/ACLED_A-Year-of-COVID19_April2021.pdf, geraadpleegd op 3 augustus 2021.

Klandermans, B., & Oegema, D. (1987). Potentials, networks, motivations, and barriers: Steps towards participation in social movements. *American Sociological Review, 52*(4), 519–531.

Meyer, D. S. (2004). Protest and political opportunities. *Annual Review of Sociology, 30*, 125–145.

Koen van der Zwet is PhD-student bij het Computational Science Lab van de Universiteit van Amsterdam, hij is ook werkzaam bij TNO Defensie & Veiligheid. Zijn onderzoek richt zich op emergent gedrag in conflict.
E-mail: koen.vanderzwet@tno.nl

Bas Keijser is werkzaam bij TNO Defensie & Veiligheid, hij werkt aan de analyse van complexe maatschappelijke problemen waaronder conflict, criminaliteit en sociale onrust.
E-mail: bas.keijser@tno.nl

Meer operaties bij OLVG tijdens corona door slimme *forecasting*

Wineke van Lent, Jasper Buil & Rob Vromans

Het OLVG zet zich als stadsziekenhuis in om de gezondheid en zorg voor iedereen te verbeteren in de stedelijke omgeving Groot-Amsterdam; 'voor een beter leven in een gezonder Amsterdam'. Het verbeteren van zorg vraagt naast aandacht voor zorginhoudelijke aspecten ook aandacht voor de wijze waarop de zorg wordt geleverd. Het continu optimaliseren van het capaciteitsmanagement krijgt vanuit die visie binnen het OLVG veel aandacht.

Deze aandacht kreeg bovendien extra focus op de inzet van capaciteiten in de Snijdende Zorgketen, in het bijzonder het OK-centrum en de kliniek, door de druk op behoud van reguliere zorg naast het leveren van COVID-zorg.

Al voordat COVID zich aandiende stond het OLVG voor uitdagingen ten aanzien van schaarste in OK-tijd en schaarste in de kliniek, waardoor er wachtlijsten voor OK bestonden naast het regelmatig weigeren (doorschuiven) van geplande patiëntopnames. De komst van COVID verergerde deze situatie. Personeel moest worden vrijgespeeld om COVID-zorg te kunnen leveren en wachtlijsten op de reguliere (electieve en semi-spoed) zorg namen toe. Een sterkere afstemming tussen OK-centrum en kliniek werd meer dan ooit een vereiste om de zorg te leveren die werd gevraagd.

De wens van het OLVG was en is om zo productief mogelijk te zijn op het OK-centrum, zowel onder als zonder druk van COVID, en tegelijkertijd een efficiënte inzet van capaciteiten te realiseren. Om dat te bereiken heeft het OLVG stappen gezet ten aanzien van integratie in besturingswijze van OK-centrum en kliniek.

Wijze van besturen van OK-centrum en kliniek

Voorafgaand aan COVID maakte en publiceerde het OLVG per 3 maanden een nieuw of aangepast OK-schema, dat de toewijzing van sessies (d.w.z. OK-ruimtes per kalenderdag) aan specialismen specificeerde. Dit schema was relatief heilig: na de publicatie kende het schema een zeer beperkt aantal wijzigingen. Bij het opstellen van het OK-schema werd weinig afstemming gehouden met de kliniek, in feite diende de kliniek te voorzien in zorg voor de bedbezetting die een consequentie was van het OK-schema ongeacht of deze last enigzins werkbaar was of niet.

Dit stramien schoot tekort tijdens de eerste COVID-golf. De impact van deze en daaropvolgende COVID-golven was een sterke variatie met een grote mate van onvoorspelbaarheid ten aanzien van de voor reguliere zorg beschikbare OK- en klinische capaciteit. De plancyclus van 3 maanden kon geen recht doen aan deze variatie. De noodzaak was geboren om veel dynamischer te zijn in het toewijzen van OK-capaciteit. Dienovereenkomstig paste het OLVG haar werkwijze aan naar een kortcyclische publicatie van OK-schema's, die bovendien sterk wisselen in aantal en type OK-sessies. Tegelijkertijd werd ook een frequente communicatie tussen OK-centrum en Kliniek geïnitieerd om afstemming te zoeken tussen beschikbare klinische capaciteit en de toewijzing van OK-capaciteit naar specialismen in het OK-schema. Tijdens de eerste COVID-golf zag het OLVG, door het ontbreken van een beter instrumentarium, zich genoodzaakt om telkens te 'gokken' over het aantal OK-sessies dat moeten worden afgeschaald om voldoende personeel vrij te spelen en de patiënten te kunnen huisvesten in de (eveneens) afgeschaalde kliniek. Geen optimale werkwijze, maar wel een waardoor met piepen en kraken een deel van de reguliere zorg overeind bleef.

Het was het OLVG duidelijk dat de onder druk tot stand gekomen werkwijze tijdens de (plotselinge) eerste COVID-golf verre van optimaal was. De drastische afschalingen van reguliere zorg waren onder tijdsdruk destijds noodzakelijk, maar met een betere methodiek vermoedelijk deels ook vermijdbaar. De stap die het OLVG wenste te zetten was om bij het ontwerp van een specifiek OK-schema inzicht te hebben in de te verwachten bedbezetting op de kliniek, om vervolgens verschillende concept OK-schema's te evalueren op die te verwachten bedbezetting en te komen tot een OK-schema dat maximale service op het OK-centrum combineert met een uitvoerbare bedbezetting op de kliniek.

Deze wens en de mogelijkheid dat een beter instrumentarium tot een beter resultaat, dat wil zeggen een groter behoud van reguliere operatieve zorg, electief en semi-spoed, kan leiden was voor het OLVG aanleiding om op zoek te gaan naar hulp via het kenniscentrum CHOIR (Center for Healthcare Operations Improvement and Research) van de Universiteit Twente.

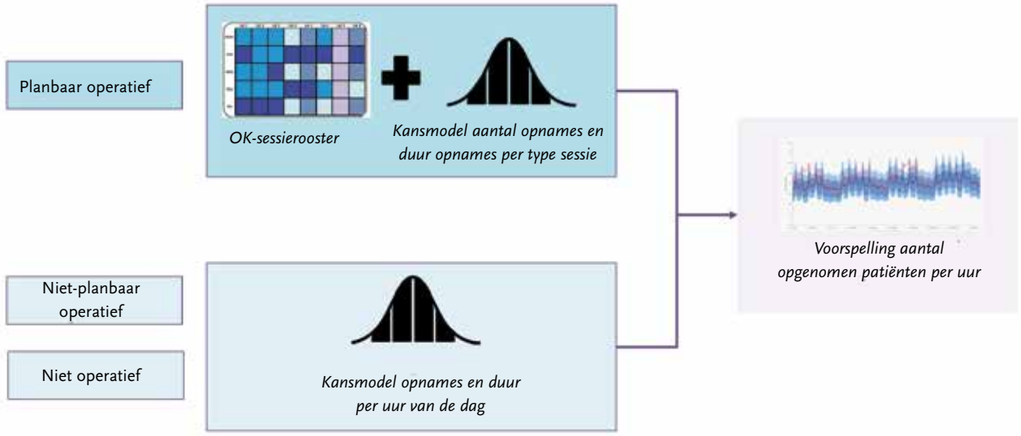
Vanuit eerder, eigen onderzoek kon CHOIR de wiskundige modellen aandragen die het OLVG helpen bij de invulling van haar wens (zie ook Vanberkel et al., 2011; Kortbeek et al., 2015a; Kortbeek et al., 2015b). Deze modellen zijn tussentijds in een Patient Flow Forecasting software-oplossing gevat, zodat de intelligentie van de modellen ook voor niet-wiskundigen op praktisch toepasbare wijze beschikbaar is en in een breed domein kan worden toegepast (zie ook Vromans et al., 2021; Braaksma, 2021). Het OLVG is deze modellen gaan toepassen vanaf de tweede COVID-golf. En met succes!

Verbeteringen in proces en resultaat

Door toepassing van het voorspellende model (zie het kader) is inzicht verkregen in beschikbare versus te verwachten benodigde verpleegkundige capaciteit. De voorspellingen die gemaakt worden met het model zijn gevalideerd op de data van het OLVG om te borgen dat betrokkenen van OK-centrum en kliniek vertrouwen in de voorspelde waarden kunnen hebben. Dankzij dit vertrouwen en de kwaliteit van de voorspellingen is het planproces van het maken en publiceren van OK-schema's sterk verbeterd. Er is een betere samenwerking tussen OK-centrum en kliniek gerealiseerd, hetgeen tot uiting komt in betere gesprekken over ongebruikte capaciteit, commitment om het OK-schema gezamenlijk te realiseren, kortere plancycli en doorlooptijd van besluitvorming over het OK-schema, en het meegeven van planregels ten aanzien van ingreep- en opnameplanning die ondersteunend zijn aan de uitvoerbaarheid van het OK-schema en de daaruit voortvloeiende verpleegbehoefte.

Deze procesverbeteringen leiden op hun beurt tot interventies op aantallen (extra) OK-sessies, type en vulling van OK-sessies (zoals sessies met alleen dagbehandeling of vermindering van het aantal spoed-OK-sessies), verschuivingen in de Kliniek (zoals het tijdelijk samenvoegen van afdelingen en het op voorhand afspreken van achtervang), en interventies op verpleegkundige inzet (zoals adviezen ten aanzien van verminderen van inzet).

Het gevolg van deze procesverbeteringen en interventies is meervoudig. Het OLVG kon gedurende 2020 en 2021 haar *fair share* van COVID-zorg continu leveren. Tevens heeft het OLVG ongeveer 10% meer OK-sessies uitgevoerd dan oorspronkelijk (zonder toepassing van de



PATIENT FLOW FORECASTING MODEL

Een OK-schema of OK-sessierooster specificeert per dag en per OK-kamer welk specialisme op die dag van die OK-kamer gebruik zal maken. Vanuit data over het verleden wordt voor ieder type OK-sessie (bepaald door karakteristieken zoals het specialisme en de weekdag) geleerd wat de bedbezettingverwachting van dat type OK-sessie is (de uitstroom vanuit de OK naar de kliniek volgend op een OK-sessie van het gegeven type). Via een convolutie van de kansen van de individuele sessies in een (eventueel nieuw ontworpen) OK-schema kan daarmee een kansverdeling worden berekend van de bedbezetting volgend op dat specifieke OK-schema, voor de electieve (planbaar) operatieve zorg. Door deze voorspelling aan te vullen met (tijdreeks-) voorspellingen van de bedbezetting voor de niet-planbare operatieve zorg en de niet-operatieve (beschouwende) zorg wordt een voorspelling van het totaal aantal op te nemen patiënten verkregen.

wiskundige modellen) was gepland. Daardoor heeft het OLVG duidelijk meer reguliere operaties uitgevoerd dan tijdens de eerste COVID-golf. Tegelijkertijd waren er, door de inzet van verpleegkundigen te baseren op de voorspelde bedbezetting, aan de kliniekzijde minder diensten nodig om voldoende zorgcapaciteit beschikbaar te hebben. Deze besparing op benodigde verpleegkundige diensten is ingezet om COVID-zorg te leveren, mensen op te leiden en te werken aan kwaliteit van zorg middels zogenaamde kwaliteitsprojecten.

Vanuit deze mooie resultaten heeft het OLVG besloten de overlegstructuur te continueren en hierbij de inzichten vanuit de modelleringen te gebruiken.

Literatuur

Braaksma, A., Kortbeek, N., & Boucherie, R. J. (2021), Bed census predictions and nurse staffing. Chapter in *Springer Handbook Healthcare logistics – Bridging the gap between theory and practice*, eds. M. Zonderland, R. Boucherie, E. Hans, N. Kortbeek.

Kortbeek, N., Braaksma, A., et al. (2015a). Integral resource capacity planning for inpatient care services based on bed census predictions by hour. *Journal of the Operational Research Society, 66*.

Kortbeek, N., Braaksma, A., Burger, C. A. J., Boucherie, R. J., & Bakker, P. J. M. (2015b). Flexible nurse staffing based on hourly bed census predictions. *International Journal of Production Economics, 161*.

Vanberkel, P. T., Boucherie, R. J., Hans, E. W., Hurink, J. L., Lent, W. A. M. van, & Harten, W. H. van. (2011). An exact approach for relating recovering surgical patient workload to the master surgical schedule, *Journal of the Operational Research Society, 62*.

Vromans, R., Kortbeek, N., Schoonhoven, L., Bosch, B. van den, & Houdenhoven, M. van. (2021), Workload forecasting and demand-driven staffing: a case study for post-operative physiotherapy. Chapter in Springer *Handbook Healthcare logistics – Bridging the gap between theory and practice*, eds. M. E. Zonderland R. J. Boucherie, E. W. Hans, N. Kortbeek.

Wineke van Lent is senior stafadviseur capaciteitsmanagement bij het OLVG.
E-mail: w.a.m.vanlent@olvg.nl

Jasper Buil is senior expert healthcare operations management bij Rhythm. Rhythm maakt opgedane kennis op gebied van zorglogistiek breder toegankelijk door begeleiding van zorgaanbieders in capaciteitsmanagement en opleiden van zorgprofessionals in zorglogistiek.
E-mail: jasper.buil@rhythm.nl

Rob Vromans is senior expert healthcare operations management bij Rhythm en PhD-kandidaat bij CHOIR/UT.
E-mail: rob.vromans@rhythm.nl

Gedragswetenschappelijk onderzoek naar coronagedragsmaatregelen en hun invloed op het dagelijks leven en welzijn van Nederlandse burgers

Wijnand van den Boom & Mart van Dijk

Preventief gedrag speelt een belangrijke rol bij het tegengaan en onder controle krijgen van de verspreiding van SARS-CoV-2, het virus dat de ziekte COVID-19 veroorzaakt. Ten tijde van de coronacrisis heeft het kabinet maatregelen afgekondigd en (dringende) adviezen afgegeven om de verspreiding van dit virus af te remmen. Deze adviezen zijn erop gericht om het aantal contactmomenten tussen burgers te beperken en het transmissierisico tijdens contactmomenten te reduceren. De effectiviteit van dit beleid hangt grotendeels af van de mate waarin men de coronamaatregelen naleeft. Vanuit de overheid kan naleving gestimuleerd worden door ongewenst gedrag op risicolocaties te verminderen (zoals sluiting van horeca, handhaving van de avondklok), of het gewenste gedrag te stimuleren en motiveren (zoals inrichten van de omgeving, heldere risicocommunicatie, faciliteren van testen).

Invloed van corona op het dagelijks leven

Voor effectief pandemiebeleid is het van belang te begrijpen waarom mensen wel of juist niet, ondanks handhaving of stimulering vanuit de overheid, de coronamaatregelen naleven. Daarom heeft het RIVM (Rijksinstituut voor Volksgezondheid en Milieu) in het voorjaar van 2020, tijdens de eerste golf van de COVID-19-pandemie, de Corona Gedragsunit opgericht. Het doel van de Corona Gedragsunit is om de kennis en expertise van de gedragswetenschappen beschikbaar te stellen voor overheidsbeleid en communicatie. Hiermee helpt de Corona Gedragsunit om preventiegedrag en welzijn van de bevolking maximaal te ondersteunen en uiteindelijk om de coronacrisis, en mogelijk andere crisissen in de toekomst, zo effectief mogelijk te bestrijden.

Een van de activiteiten van de Corona Gedragsunit is een grootschalig 6-wekelijks vragenlijstonderzoek, waar-

mee beter inzicht wordt verkregen in de invloed van de coronamaatregelen op het dagelijks leven, het welzijn van Nederlandse burgers en de mate waarin het ze lukt om de gedragsregels op te volgen. Zie het kader voor een overzicht van de kennisproducten.

Het RIVM heeft samen met GGD-GHOR NL en de 25 GGD'en alle deelnemers (16 jaar en ouder) van al bestaande panels van de GGD'en benaderd. Het design van het onderzoek is een 'dynamisch cohort': een groep deelnemers wordt een onbepaalde tijd gevolgd, waarbij ook nieuwe deelnemers op latere tijdstippen instromen in het cohort. De nieuwe instroom wordt onder andere geworven via sociale media kanalen. Deelnemers kunnen ook aangeven of zij benaderd willen worden voor een diepte-interview of deelname aan een focusgroep. Interviews en focusgroepen bieden de mogelijkheid om verdieping te zoeken op specifieke thema's en om inzicht te krijgen in de motivaties en drijfveren van gedrag.

Alle respondenten krijgen iedere zes weken vragen over naleving van de coronamaatregelen, sociale activiteiten, gezondheid en een aantal demografische kenmerken. Om de duur van het invullen van de vragenlijst te beperken zijn deelnemers naar drie sub-cohorten gerandomiseerd: psychosociale determinanten van gedrag, draagvlak voor de coronamaatregelen en welbevinden. In tabel 1 staan enkele voorbeelden van variabelen die voor de verschillende thema's worden uitgevraagd.

THEMA	ENKELE VOORBEELDEN
Demografie	Geslacht, leeftijd, gemeente, opleidingsniveau, geboorteland, leefsituatie, werk, financiële situatie, kwetsbare gezondheid
Gedrag	• Naleven van coronamaatregelen: hygiëne maatregelen (bijv. handen wassen), contact-beperkende maatregelen (bijv. beperkt aantal bezoekers thuis ontvangen), dragen van mondkapjes, quarantaineplicht • Testen en COVID-19 gerelateerde klachten: het hebben van klachten en een test doen (waar, waarom wel/niet), vragen omtrent vaccinaties • Sociale activiteiten: aantal keer het huis verlaten voor verschillende redenen (bijv. om te werken), bezoekers thuis hebben, ventileren van het huis, op vakantie gaan, bezoeken van hoog-risico landen
Draagvlak	Steun voor de coronamaatregelen zoals die gelden op moment van afname en in het hypothetische geval dat ze nog zes maanden zouden gelden; procedurele rechtvaardigheid, vertrouwen in de overheid
Psychosociale determinanten van gedrag	• Risico perceptie (o.a., inschatting van kans en ernst om zelf besmet te raken) • Emotionele respons ten aanzien van het virus (perceptie van hoe snel het virus zich verspreidt en hoeveel zorgen, stress, angst mensen ervaren). Per coronamaatregel: • *Self-efficacy* (hoe moeilijk/makkelijk mensen het vinden dat gedrag uit te voeren) • *Response efficacy* (hoe effectief mensen denken dat gedrag is in het voorkomen van verspreiding van het virus) • Descriptieve norm (in welke mate mensen ervaren dat anderen in de nabije omgeving desbetreffende maatregel naleven) • *Cue to action* (bijv. COVID-19-besmetting in de nabije omgeving)
Welzijn	Algemene gezondheid, mentale gezondheid, eenzaamheid, leefstijl (bijv. alcohol inname)
Media	Gebruik van media in relatie tot toegang COVID-19-informatie

Tabel 1. Een selectie van variabelen binnen het vragenlijstonderzoek van de Corona Gedragsunit, april 2020 – september 2021

Een selectie van de resultaten

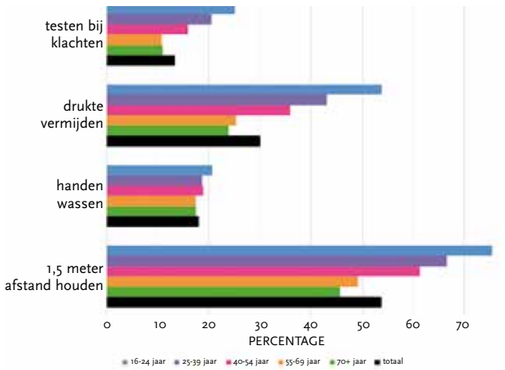
Tot en met september 2021 zijn gegevens verzameld van meer dan 180.000 mensen die samen meer dan 800.000 vragenlijsten hebben ingevuld (tabel 2). Per vragenlijstronde doen ongeveer 50.000 deelnemers mee.

In dit artikel wordt een selectie van de resultaten van het vragenlijstonderzoek besproken. Een periode van anderhalf jaar maakt duidelijk dat de deelnemers het steeds moeilijker zijn gaan vinden om zich aan de coronamaatregelen te houden. Dit geldt niet voor hygiënemaatregelen zoals handen wassen, maar wel voor contact-beperkende maatregelen (bijvoorbeeld, niet op een drukke plek komen) en coronamaatregelen in het geval van COVID-19-specifieke klachten (bijvoorbeeld, testen en thuisblijven bij klachten). De huidige vaccinatiegraad lijkt vooral een relevante factor voor contact-beperkende maatregelen: een deel van de deelnemers geeft aan dat ze zich minder aan de coronamaatregelen zijn gaan houden na vaccinatie (zie figuur 1). Ook zijn er duidelijke verschillen tussen de leeftijdsgroepen: de jongste leeftijdsgroep (16–24 jaar) houdt zich na vaccinatie minder aan de coronamaatregelen dan de oudere leeftijdsgroepen, met name bij 1,5 meter afstand houden en drukte vermijden.

Er is ook gemeten in hoeverre de deelnemers de coronamaatregelen steunen (draagvlak). Voor contact-beperkende maatregelen is een duidelijke afname te zien van draagvlak. Deze daling heeft ingezet in januari 2021 en is ook waarneembaar voor andere maatregelen zoals het houden van 1,5 meter afstand en thuiswerken. Na wat schommelingen in het vertrouwen in het beleid van de Nederlandse overheid met betrekking tot het terugdringen van corona, is ook dit vertrouwen aan het dalen. Dit gaat gepaard met meer mensen die het beleid oneerlijk vinden en een forse toename in mensen die de coronamaatregelen verwarrend vinden, ondanks dat er sinds september 2021 minder coronamaatregelen zijn. Ook komt uit de resultaten naar voren dat gevoelens van eenzaamheid toenemen wanneer er strengere contact-beperkende maatregelen gelden.

RONDE	COHORT (N)	INSTROOM (N)	AFNAME PERIODE
1		89.943	17 – 24 april 2020
2	52.950		7 – 12 mei
3	47.475	16.522	26 mei – 1 juni
4	50.290		17 – 21 juni
5	44.982	5.483	8 – 12 juli
6	44.466	17.185	19 – 23 augustus
7	47.740		30 september – 4 oktober
8	44.546	19.625	11 – 15 november
9	51.468		30 december – 3 januari 2021
10	45.741	8.622	10 – 14 februari
11	47.254		24 – 28 maart
12	44.159	11.283	5 – 9 mei
13	42.550	6.678	16 – 20 juni
14	37.874		28 juli – 1 augustus
15	36.680	7.686	8 – 12 september

Tabel 2. Deelnemers van het vragenlijstonderzoek van de Corona Gedragsunit tussen ronde 1 en ronde 15, april 2020 - september 2021



Figuur 1. Percentage van de deelnemers dat aangeeft (veel) minder bepaalde coronamaatregelen na te leven na vaccinatie, naar leeftijd, RIVM Corona Gedragsunit vragenlijstonderzoek ronde 15 (8-12 september 2021)

Vaccinatie-intenties

In november 2020 konden deelnemers van het vragenlijstonderzoek in open antwoorden aangeven welke overwegingen zij hadden om zich wel of niet te laten vaccineren tegen het coronavirus. Deze redenen werden vervolgens onderzocht in telefonische interviews met een select aantal deelnemers. Uit de thematische analyse van de interviews kwamen in totaal zes overtuigingen naar voren, waaronder 'Als ik het vaccin krijg, ben ik beschermd tegen het coronavirus' (Sanders et al., 2021). Deze werden in de eerstvolgende vragenlijst opgenomen als stellingen, ieder met een 5-punts antwoordschaal van 'helemaal mee oneens' tot 'helemaal mee eens'. Het RIVM was geïnteresseerd of mensen die switchten van 'ik twijfel of ik een vaccin wil' naar 'ja, ik wil een vaccin' of 'nee, ik wil géén vaccin' verschilden van hen die tussen twee meetmomenten nog steeds twijfelden. De resultaten van een multinomiale logistische regressie lieten zien dat onder deelnemers die twijfelen over vaccinatie, de overtuiging dat het coronavaccin veilig is samenhing met de keuze om zich wel of niet te laten vaccineren: twijfelaars die zich wel lieten vaccineren waren het eens met dat het vaccin veilig is, terwijl twijfelaars die zich niet laten vaccineren het niet eens waren dat het vaccin veilig is. Uit zowel de interviews

als de regressies kwam naar voren dat mensen die twijfelen over vaccinatie vaak pas de keuze maken wanneer zij de uitnodiging voor vaccinatie ontvangen.

Methodologische uitdagingen

Een groot deel van de respondenten is benaderd als lid van een al bestaand GGD-panel, een ander deel is uitgenodigd via sociale media. Beide manieren hebben echter een selectie-effect. Hierdoor zijn de respondenten demografisch niet volledig representatief voor de Nederlandse bevolking. Daarom is voorzichtigheid geboden bij het extrapoleren van de resultaten naar de Nederlandse bevolking. Overigens wordt een deel van de vragen binnen de vragenlijst (rondom draagvlak en naleving) geverifieerd met het trendonderzoek voor het landelijk coronadashboard (https://coronadashboard.rijksoverheid.nl/landelijk/gedrag). Hieruit blijkt dat de resultaten grotendeels aansluiten. Daarnaast hebben de onderzoekers te maken met een aantal vormen van ontbrekende (*missing*) data. Zo zijn er deelnemers die besluiten na een eerste of een vervolgmeting niet meer mee te doen (*lost to follow-up*). Daarnaast krijgen deelnemers, door het bovengenoemde design, vragen uit één bepaald thema (bijvoorbeeld welzijn), en standaard niet de vragen van een ander thema. Dit bemoeilijkt het analyseren van verbanden tussen de antwoorden van verschillende thema's. Een andere uitdaging zijn externe factoren zoals vakantieperiodes, seizoensinvloeden en het steeds (op korte termijn) veranderende overheidsbeleid. Dit zijn factoren die mogelijk een rol kunnen spelen bij het gedrag van deelnemers of (mede) een verklaring zijn voor de verschuivingen in cijfers, maar die moeilijk te meten zijn. Tevens kunnen onderwerpen die rondom het afnamemoment vaker in de media zijn geweest van invloed zijn op bijvoorbeeld het draagvlak en vertrouwen in de overheid, bijvoorbeeld de media aandacht rondom de vaccinatiestrategie. Een ander voorbeeld is het moment van vaccinatie: deze hangt vrijwel altijd samen met leeftijd, waardoor het effect van vaccineren (op bijvoorbeeld naleven van de gedragsregels) lastig is los te koppelen van leeftijd.

Conclusies

De resultaten van het onderzoek van de Corona Gedragsunit, en die van het vragenlijstonderzoek in het bijzonder, hebben ervoor gezorgd dat adviezen van gedragswetenschappers nadrukkelijker worden meegewogen in de besluitvorming en communicatie over de coronamaatregelen. Het vragenlijstonderzoek geeft elke zes weken opnieuw inzicht in de naleving van en het draagvlak voor de coronamaatregelen en het welbevinden van Nederlandse burgers. Daarnaast bieden de data mogelijkheden voor verdere analyses, waarbij bijvoorbeeld effecten over tijd of verschillen tussen groepen geanalyseerd worden. De gedragswetenschappelijke inzichten die hiermee verkregen worden kunnen van toepassing zijn bij toekomstige pandemieën, of nationale gezondheidscrisissen in het algemeen.

Met dank aan onze collega's bij het RIVM: Jan Brouwer de Koning, Guus Luijben, Margriet Melis en Marcel Scholten

Voor vragen of opmerkingen, mail de Corona Gedragsunit, coronagedragsunit@rivm.nl

Kennisproducten van de RIVM Corona Gedragsunit

Op basis van kennis uit onder meer literatuur en de 6-wekelijkse onderzoeksresultaten van het vragenlijstonderzoek, schrijft de Corona Gedragsunit ook geïntegreerde memo's: onder andere over testen en thuisblijven, veilige feestdagen en vakantie, hoe om te gaan met een opleving van het virus, risicogedrag in relatie tot de stijgende vaccinatiegraad en de introductie van het coronapaspoort, en vaccinatiebereidheid bij verschillende doelgroepen. In aanloop naar besluitvorming over aanpassingen in het maatregelenpakket, levert de Corona Gedragsunit op verzoek van de Nationaal Coördinator Terrorismebestrijding en Veiligheid (NCTV) en het ministerie van Volksgezondheid, Welzijn en Sport (VWS) ook gedragsreflecties op, zoals gedragsreflecties op het mogelijk loslaten van 1,5m en brede inzet van coronatoegangsbewijs. Praktijkvoorbeelden met verhalen over onder meer school, sport, jeugd en evenementen worden ook gepubliceerd. Opgedane kennis wordt gedeeld via de RIVM website, tijdens online duidingssessies en via de nieuwsbrief Corona & Gedrag, die ongeveer 10 keer per jaar verschijnt en waarop men zich kan abonneren. Voor meer informatie: https://www.rivm.nl/gedragsonderzoek/publicaties

Literatuur

Sanders, J. G., Spruijt, P., van Dijk, M., Elberse, J., Lambooij, M. S., Kroese, F. M., & de Bruin, M. (2021). Understanding a national increase in COVID-19 vaccination intention, the Netherlands, November 2020-March 2021. *Euro Surveill*, *26*(36). https://doi.org/10.2807/1560-7917.ES.2021.26.36.2100792

Wijnand van den Boom (wijnand.van.den.boom@rivm.nl) en Mart van Dijk (mart.van.dijk@rivm.nl) werken beide bij het Rijksinstituut voor Volksgezondheid en Milieu (RIVM). Sinds de coronacrisis zijn ze actief in de Corona Gedragsunit van het RIVM.

De COVID Radar is een smartphone app, gelanceerd in Nederland op 2 april 2020, waarmee gebruikers zelf hun COVID-19 gerelateerde symptomen en social-distancing gedrag kunnen rapporteren. Zij kunnen dit doen door in de app regelmatig een korte vragenlijst in te vullen, bij voorkeur dagelijks. De app gebruikt dynamische plattegronden om gebruikers feedback te geven over hun symptomen en gedrag vergeleken met nationale en regionale gemiddelden. Data worden anoniem verzameld. De enige persoonsgegevens die de app verzamelt zijn iemands viercijferige postcode, geslacht, en leeftijdscategorie (in blokken van tien jaar). Sinds de lancering, heeft de app al meer dan 7 miljoen ingevulde vragenlijsten van meer dan 285.000 unieke gebruikers verzamel

COVID RADAR *Populatie-gerichte monitoring gedurende de COVID-19-crisis*

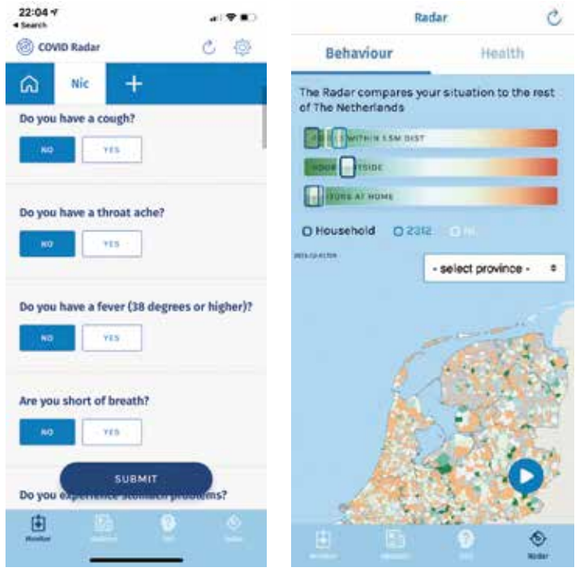
Nic Saadah & Michelle Brust

Vlak na de invoering van de intelligente lockdown in Nederland, was er bij Leiden Universitair Medisch Centrum (LUMC) en ORTEC, een data-analysebedrijf, de wens om grootschalig wetenschappelijk onderzoek te doen naar de geografische verspreiding en context van het virus dat COVID-19 veroorzaakt. Dit resulteerde in de COVID Radar. In een sindsdien wekelijks stuurgroepoverleg wordt de koers van de app en potentiële wijzigingen van de app besproken en doorgevoerd. De stuurgroep bestaat uit betrokkenen van zowel LUMC als ORTEC, zoals epidemiologen, infectiologen, huisartsen, onderzoekers, data analisten en specialisten op het gebied van communicatie en digitale media.

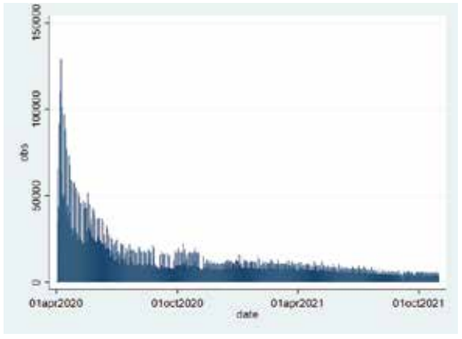
Een persbericht van het LUMC direct na de lancering van de app werd opgepakt door verschillende lokale en nationale media, wat al snel resulteerde in een grote groep gemotiveerde deelnemers. In de loop van de tijd zijn er regelmatig vragen gewijzigd, toegevoegd, of juist verwijderd. Door te vragen naar symptomen die specifiek zijn voor COVID-19 en gedragsinformatie werd er geprobeerd te voorspellen waar, wanneer en in welke populaties zich nieuwe uitbraken voordoen. Het doel was om deze gegevens te gebruiken voor populatiegerichte beleidsadvisering en wetenschappelijk onderzoek.

COVID Radar sneller dan de officiële testuitslag

Inmiddels hebben wij een databestand van meer dan 16 maanden aan ingevulde COVID Radar-vragenlijsten, verzameld vanuit het hele land. Analyse op deze data laat zien dat zelf-rapporteren van symptomen en gedrag geassocieerd is met het door RIVM gerapporteerde officiële aantal COVID-19-cases. Van Dijk et al. (2021) laten zien dat zelfgerapporteerde symptoom- en gedragsinformatie in de COVID Radar verband heeft met het aantal COVID-



Figuur 1. Vragenlijst en kaart met symptomen

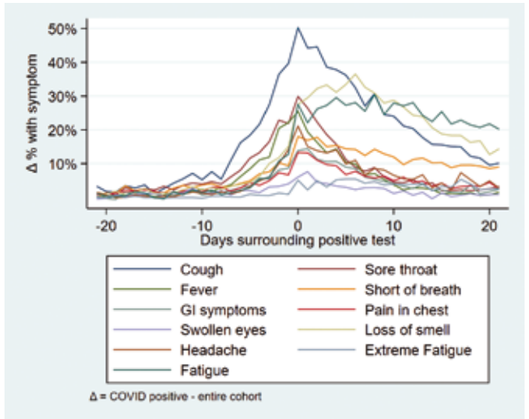


Figuur 2. Aantal observaties per dag

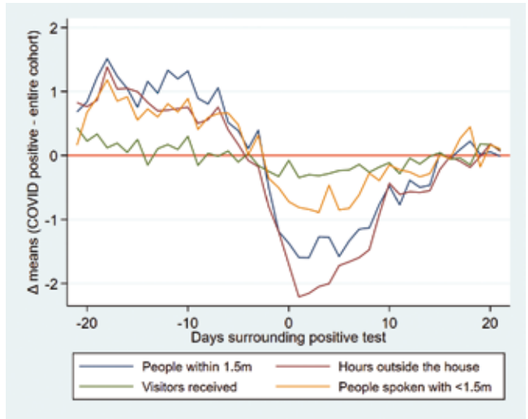
19-cases op regionaal niveau, waarbij er een sterker verband is in regio's met veel gebruikers dan in regio's met minder gebruikers. Gezien het normale traject van (1) besmetting, (2) symptomen, (3) testen, (4) confirmatie, zou het voorspellen van een COVID-19-besmetting enkele dagen voor een officiële testuitslag de tijd waarin iemand besmettelijk rondloopt kunnen verminderen.

Figuur 1 laat de interface van de app zien, met hierin een screenshot van de vragenlijst en de kaart met symptomen. Gebruikers van COVID Radar laten weten dat zij deze kaart gebruikt hebben om beslissingen te maken. Zij kozen bijvoorbeeld aan de hand van de kaart naar welke supermarkt of park ze wilden gaan, op basis van waar symptomen het minst voorkomend waren. Figuur 2 laat het aantal dagelijkse ingevulde vragenlijsten in de loop van tijd zien. De periodiciteit die hier te zien is heeft te maken met een push-bericht welke om de twee dagen als herinnering op het toestel van de gebruiker verschijnt. Figuur 3 laat zien hoe zowel symptomen (figuur 3a) als gedrag (figuur 3b) geassocieerd zijn met een positieve COVID-19-test. Te zien is dat voorafgaand aan een positieve uitslag alle symptomen stijgen. Voornamelijk hoesten werd door de helft van positief geteste gebruikers gerapporteerd. Figuur 3b laat zien dat mensen die positief testen gedurende de periode voor én na hun uitslag bovengemiddeld risicovol gedrag vertonen. Dit effect is niet te zien in de weken rondom hun testuitslag. Waarschijnlijk omdat ze hebben besloten dat ze getest moeten worden, doen zij het dan juist beter dan gemiddeld.

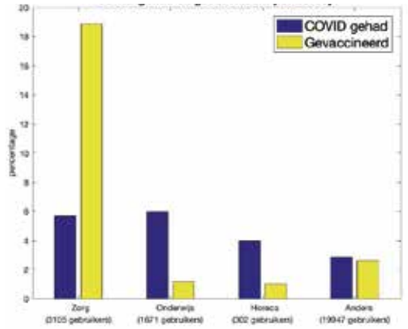
Ondanks dat meer dan 285.000 unieke gebruikers ooit de app ingevuld hebben, zien wij recentelijk een



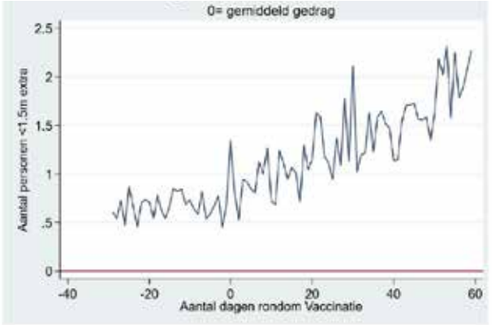
Figuur 3a. Symptomen rondom positieve COVID-19-test



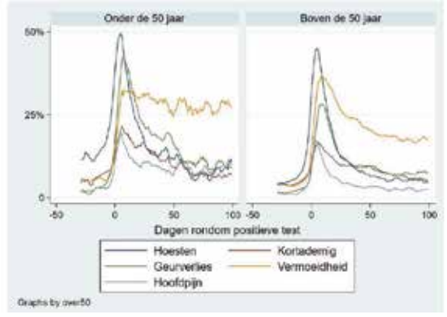
Figuur 3b. Gedrag rondom positieve COVID-19-test



Figuur 4a. Vergelijking vaccinatiegraad tussen zorg- en onderwijspersoneel



Figuur 4b. Risicogedrag neemt toe na vaccinatie



Figuur 4c. Proportie COVID-patiënten met klachten 3 maanden na diagnose

Figuur 4d. Versoepelingen maatregelen terug te zien in de data van COVID Radar

enorme daling in het aantal dagelijks ingevulde vragenlijsten. Van dagelijks boven de 100.000 vragenlijsten in de beginfase naar rond de 10.000 à 15.000. De cohort die de app nog regelmatig trouw invult, is over het algemeen ouder, werkt twee keer vaker binnen de zorg en komt grotendeels uit de regio Leiden, waar de app ontwikkeld is en beheerd wordt.

Surveillance doeleinden

In de loop van de afgelopen anderhalf jaar hebben wij een uniek blik gehad op de pandemie dankzij de analyses die wij uit kunnen voeren met de COVID Radar-data. In april, toen de vaccincampagne pas begonnen was, hebben wij kunnen laten zien dat de vaccinatiegraad binnen onderwijsmedewerkers veel lager was dan binnen de zorgmedewerkers, ondanks het feit dat de kans op corona bij deze twee groepen redelijk gelijk was (figuur 4a). In mei hebben wij kunnen laten zien dat het risicogedrag van mensen toeneemt na vaccinatie, ondanks het feit dat er toen geen versoepeling voor gevaccineerden aangekondigd werd (figuur 4b). Met de loop van tijd, toen het helder werd dat COVID ook tot langdurige klachten leidt, was het met onze dataset mogelijk te laten zien hoe veel COVID-patiënten nog klachten hadden maanden na hun diagnose (figuur 4c). Ook in deze meeste recente fase van de pandemie konden wij het effect van het loslaten van maatregelen in onze data zien. Zo zagen wij bijvoorbeeld een stijging van het aantal mensen binnen de 1,5 meter bij onze COVID Radar-gebruikers (figuur 4d).

De technologie van de COVID Radar laat zien dat het haalbaar is om op korte termijn data over gedrag en symptomen van de populatie te verzamelen voor surveillance-doeleinden gedurende de COVID-19-crisis. Tegelijkertijd vergt het continue aandacht om gebruikers gemotiveerd te houden om de app in te blijven vullen. Wij deden dit door hen bij de dataverzameling te betrekken, regelmatig resultaten via nieuwsberichten terug te koppelen alsmede de app aan te passen op basis van nationale ontwikkelingen zoals de uitrol van vaccinatie en veranderende *social distancing*-maatregelen.

Literatuur

Van Dijk, W.J., Saadah, N. H., Numans, M. E., Aardoom, J. J., Bonten T. N., & Brandjes M., et al. (2021) COVID RADAR app: Description and validation of population surveillance of symptoms and behavior in relation to COVID-19. *PLoS ONE, 16*(6): e0253566. https://doi.org/10.1371/journal.pone.0253566

Nic Saadah (n.h.saadah@lumc.nl) en Michelle Brust (M.Brust@lumc.nl) zijn werkzaam bij het Leids Universitair Medisch Centrum en lid van de Stuurgroep COVID Radar.

Bewijs het maar eens!

hoe wiskundig denken soms fout uit kan pakken

Het is alweer ruim 17 jaar geleden. Mijn oudste zoon werkte in Engeland aan zijn PhD en had daar een leuke Griekse ontmoet die ook aan een PhD werkte. Zij besloten te trouwen in Athene en mij werd gevraagd alle Nederlandse papieren daarvoor in orde te maken.

Nou, dat heb ik geweten, vele dagen en honderden euro's gingen er mee heen. Ik dacht heel naïef dat het een kwestie was van even naar de Burgerlijke Stand gaan. Ten slotte zou de plechtigheid zich afspelen in een EU-land, beide partners waren afkomstig uit een EU-land en woonden in een toen nog EU-land. Nu zou ik eindelijk eens aan den lijve de zegeningen van een verenigd Europa ervaren. Maar dat viel tegen!

Het grootste struikelblok was dat er, naast geboorte- en doopcertificaten, een 'verklaring van huwelijkse bevoegdheid' moest komen voor mijn zoon. Onderdeel daarvan is dat men aantoont dat men niet getrouwd is. In Nederland vertrouwt men erop dat als zoiets niet in de Bevolkingsadministratie voorkomt het ook niet aanwezig is. Maar omdat het Napoleon indertijd niet lukte het Kanaal over te steken heeft men in het VK geen Bevolkingsadministratie zoals wij die kennen. Er is daar een Belastingregister en daar moest ik het mee doen.

Ik heb eerst geprobeerd de ambtenaren in Wageningen, zijn laatste Nederlandse woonplaats, ervan te overtuigen dat men een onmogelijk bewijs eiste. Met mijn wiskundig gedeformeerde geest legde ik uit dat men per definitie niet kan bewijzen dat men *niet* getrouwd is. Immers, dat is de *default* toestand die iedereen heeft vanaf de geboorte. Pas als men getrouwd is kan men een bewijs dáárvan overleggen en tot die tijd is men gewoon ongetrouwd. Tja, met dit soort formalistische redeneringen moet je niet aankomen bij de gemiddelde Nederlandse ambtenaar. Het scheelde maar weinig of men had mij uit het stadhuis verwijderd.

Gelukkig wist ik op tijd mijn mond te houden en een opmerking in te slikken waarmee ik de absurditeit van de eis wilde illustreren. Ik stond namelijk op het punt te stellen dat hij best in een ver buitenland getrouwd kon zijn tijdens een afstudeerstage zonder dat we daar iets van wisten. Als ik dat had gezegd waren denk ik de rapen pas goed gaar geweest.

Er zat dus niets anders op dan te proberen in Engeland een document te verkrijgen waarmee 'bewezen' werd dat mijn zoon niet getrouwd was. Hij heeft uitgebreid overlegd met het stadhuis, de afdeling personeelszaken en de juridische afdeling van de universiteit en met de Nederlandse ambassade in Londen. Die zagen in wat het probleem was en werkten allemaal van harte mee. Gezamenlijk heeft men een indrukwekkend document opgesteld en dat van veel stempels en het grootzegel van de universiteit voorzien. Hierin verklaarden de belastingambtenaren en de universiteit dat, voor zover uit hun administraties bleek, de heer Stemerdink niet gehuwd was. De ambassade verklaarde aanvullend dat de handtekeningen op dat document van officiële bij hen bekende instanties afkomstig waren en voegde er nog een aantal stempels aan toe. Tegen zoveel gewichtigheid was de Wageningse ambtenaar niet bestand en hij gaf de gewenste verklaring af. Daarna moest het hele dossier door een beëdigd vertaler naar het Grieks worden omgezet. Als finishing touch moest de rechtbank waar die vertaler ooit beëdigd was nog een gestempelde verklaring toevoegen dat hij inderdaad bevoegd was die vertaling te produceren.

Bruidspaar Chris en Maria Stemerdink-Panagos met de ouders van Chris. Athene, 3 juli 2004.

En de bruid? Die had zich, met Griekse nonchalance jegens regels, niet laten uitschrijven toen ze naar Engeland vertrok. Die had nog gewoon een Grieks adres en ondervond geen enkel probleem. Soms kan het toch wel handig zijn je niet netjes aan de regels te houden.

Gerrit Stemerdink is eindredacteur *STAtOR*.
E-mail: gjstemerdink@hotmail.com

PANDEMIC PREPAREDNESS IN DATA SHARING

lessons learned from collaborating in a live meta-analysis

Judith ter Schure, Peter Grünwald & Alexander Ly

The scientific response to the COVID-19 pandemic was far from perfect. Conflicting results on hydroxychloroquine made officials in the U.S. first recommend the anti-malaria drug and then warn against it. Similarly, systematic reviews on ivermectin could not draw robust conclusions when they first included results in the meta-analysis and had to exclude them later after their retraction from a preprint server. How different was the response of the research community studying the Bacillus Calmette-Guérin (BCG) vaccine! No BCG researcher went on television to state that they single-handedly proved that the BCG vaccine – originally developed to protect against tuberculosis – makes us invincible. On the contrary, the BCG community worked closely together and remained cautious until this day. The results will be published later this year; here we want to simply chronicle how it all started and what we learned along the way.

Early 2020, BCG researchers from the university medical centers of Utrecht and Nijmegen were among the first to announce their clinical trial (newspaper Trouw, March 18, 2020). Not only were they early, they were also generous and shared their protocol when other researchers around the world started similar trials. Already from the beginning these trials had much in common and great potential to be analyzed together. The chaos surrounding hydroxychloroquine shows how important coordination can be. The story of ivermectin illustrates the risks of a meta-analysis that waits for summary estimates to appear in (preprint) publications.

Even if trials are performed perfectly, however, unreliable results can arise due to multiple testing when many trials address the same question simultaneously. Fortunately, the BCG researchers were warned of this risk by their trial statistician dr. Henri van Werkhoven. A conse-

quence of this risk is that the first trial to find an effect could be an outlier, but still be published quickly and threaten the continuation of the other trials. In the urgency of a pandemic, a disagreeing meta-analysis might come too late to start the trials up again.

ALL-IN meta-analysis

When we offered the Dutch BCG researchers a solution to this problem they had already contacted many of the trials around the world. For statistical validity, the BCG trials needed coordination and needed to be analyzed together. For efficiency's sake, we should start the statistical analyses as soon as possible. Our plans got a name: ALL-IN meta-analysis, for Anytime Live and Leading Interim meta-analysis. We provided the statistical methodology to analyze all these BCG trials together on a continuous basis, while they were still ongoing.

The BCG researchers courageously embraced our novel methods and ALL-IN-META-BCG-CORONA was born (Van Werkhoven et al, 2020). It became a collaboration by two groups of clinical studies, of 7 and 4 each, that decided on trial selection together, shared their data at interim stages and monitored the results *live* in a dashboard. We will focus here on the 7 trials that studied healthcare workers (the others study the elderly) and on the outcome measure of COVID-19 infections (the other being severe COVID-19 infections requiring hospitalization).

The main goal was to find out whether an immune response to BCG provides indirect protection against COVID-19. If a beneficial effect were to be confirmed quickly, this could save many lives since BCG is widely available around the world, which was not the case for many other possible treatments at the time. COVID-19 specific vaccines, of course, just started development. On the other hand, if futility or harm could be confirmed, studies could be stopped early and resources saved and put to better use elsewhere in the scientific response to the pandemic.

Lessons learned

Working in a large-scale multi-center collaboration across the globe comes with many lessons. First, we learned the intricacies of time in sequential time-to-event analysis in our development of the *safe* logrank test and *anytime-valid* confidence sequences for the hazard ratio (Ter Schure et al., 2020a). Second, we realized the need for a software package *safestats* (Ly et al., 2020) with transparent tutorials and a webinar (Ter Schure et al., 2020b). Third, we were confronted with meta-analysis issues that arise from a bottom-up collaboration, like heterogeneity in trials, (dis-)agreement on decision rules and interpretation of results. The experience does make us hopeful about the benefits of the approach in general, in terms of efficiency, collaboration and communication (Ter Schure & Grünwald, 2021). Finally, we learned about data sharing, which is what we would like to discuss here. We came up with solutions to the issues at hand, but also have questions that still remain.

Crucial is that the statistical test and confidence intervals that we developed and applied are valid at any time. Figure 1 shows the dashboard that we used to communicate interim results to the participating trials. (The dashboard is in a demo mode and based on synthetic data: 'fake' values for each trial based on public trial characteristics.) We kept track of an *e*-value (Grünwald et al., 2019), a measure of evidence and test statistic that we compared to the threshold $1/\alpha = 400$. Whenever the cumulative meta-analysis *e*-value would cross this threshold, we could declare statistical significance – you can think of an *e*-value as $1/p$-value, with the crucial difference that it keeps its validity for testing irrespective on when we stop data collection. This procedure guarantees type-I error control at level $\alpha = 0.0025$ regardless of the sampling plan, the number of analysis or their timing. Importantly we chose a strict alpha of 0.0025 because we wanted to reserve maximum power for the co-primary endpoint Covid-19 related hospitalization while maintaining overall type-1 error <0.05. Apart from hypothesis testing, we also kept track of confidence intervals that were anytime-valid. In Figure 3 we show examples of these.

Data sharing

Practical hurdles arise when data are shared. This is true in general, but in our ambition to do a *live* analysis this was even more pronounced. Our intention was to retrospectively process each newly updated trial data set to show how the evidence since the last upload had changed; not only to find a conclusion of benefit as early as possible, but also to show whether the evidence was moving in the right direction and facilitate preparations for a future conclusion. By showing an *e*-value for each calendar day, our dashboard allowed users to spot trends in the evidence very easily.

Sharing live results through a dashboard while keeping researchers blinded

In clinical trials like the BCG trials most involved researchers and doctors are blinded to the allocation of treatment and placebo, as well as to any results. In general, these two are connected: if you know that the results start pointing towards an effective treatment, this might also indicate whether a participant that is improving has received treatment or placebo. For our analysis, however, we needed at least one person to handle the trial data fully unblinded. This turned out to be no problem, since most trials had a trial statistician that would also perform interim analyses and/or provide interim data to a data safety and monitoring board. We asked for this person to be the data uploader for the meta-analysis.

These data-uploaders were the first to get access to the dashboard, each with personal login details. To others, the dashboard could reveal a participant's allocation when they still needed to remain blinded. Figure 1 shows that the line of *e*-values goes up if an event occurs in the control group (evidence against the null brings us closer to the threshold at 400 for benefit) and that the line goes down if an event occurs in the BCG group. This level of detail in the dashboard reveals both the time and place of occurrences of COVID-19 by the calendar date and trial. If you observe that a sequence of *e*-values goes up at a certain calendar date, and you know the person that tested

Figure 1. Dashboard used to communicate interim results in ALL-IN-META-BCG-CORONA to all data uploaders with a login. The involved trials were performed in the Netherlands (NL), Denmark (DK), the United States (US), Hungary (HU), Brazil (BR), France (FR) and Guinea-Bissau/Mozambique (AF). The dashboard is in demo mode with "fake" values. Note that the y-axis is on the log scale (https://cwi-machinelearning.shinyapps.io/ALL-IN-META-BCG-CORONA/ username = demo, password = show)

intervention	dataRand	hospital	COV19	dateCOV19	COV19hosp	dateCOV19hosp	dateLastFup
control	2020-05-07	A	yes	2020-05-11	yes	2020-05-15	2020-06-23
control	2020-05-04	B	yes	2020-05-08	yes	2020-05-12	2020-06-23
BCG	2020-05-08	A	yes	2020-05-21	yes	2020-06-01	2020-06-23
control	2020-05-07	B	yes	2020-05-25	no	NA	2020-06-23
BCG	2020-05-05	A	yes	2020-05-24	no	NA	2020-06-23
BCG	2020-05-10	B	yes	2020-06-03	no	NA	2020-06-23
control	2020-05-14	A	yes	2020-06-23	no	NA	2020-06-23
control	2020-05-10	B	no	NA	no	NA	2020-06-23
BCG	2020-05-08	A	no	NA	no	NA	2020-06-23
BCG	2020-05-04	B	no	NA	no	NA	2020-06-23

Figure 2. Example (fake) data set from the working instructions to data-uploaders (Ter Schure et al., 2020b)

positive for COVID-19 that day, you can deduce with certainty that the person was randomized to placebo (and similarly to BCG if the *e*-values go down).

In the early stages of the meta-analysis, these logins only gave permission to view the overall meta-analysis *e*-values and the data uploader's own trial contribution. This set-up ensured that no one could access the dashboard that needed to stay blinded to interim results and that data uploaders could not access privacy sensitive data from other trials. Observed COVID-19 infections from other trials were bundled together in the meta-analysis *e*-values such that the location of those events could not be derived from the dashboard.

Remaining questions

If we would group more than one observation of COVID-19 together by calculating an *e*-value by week or month, instead of by day, would that make it sufficiently hard to deduce randomization from observed events? Would that be enough to allow data-uploaders (not blinded to their own trial results) to inspect other trial's *e*-values? Or even to allow all participating researchers (blinded to their own trial results) to inspect all trial results except their own?

A central analysis

In collecting the data from the trials we had two options for an analysis by calendar date (as in the dashboard in Figure 1). Either do a meta-analysis based on summary statistics (so-called *two-stage* meta-analysis) or do a meta-analysis on the raw data (so-called *IPD* meta-analysis, for *Individual Patient Data*). While this decision is familiar in meta-analysis, for the first option, we had to ask the data-uploaders something completely unfamiliar. We did not only want them to share new summary statistics at each data upload, but to share a sequence of summary statistics by calendar date each time they uploaded new data. On the other hand, for the *IPD*-analysis of the second option, there was nothing special and we needed all trials to simply upload their data so far to an upload-only folder that only we could access. We chose that second option.

Figure 2 shows the type of data we requested for our BCG analysis. The analysis was stratified by hospital, so for each healthcare worker randomized in the trial, we had to receive information about their location. We allowed the trials to label the hospitals ('A', 'B') without actually naming them, since by knowing the hospitals, the data would become more privacy sensitive. If you know the hospital and the calendar date that someone entered the study (dateRand), you could recognize that person in the raw data and identify whether the person had COVID-19. The approach did not mitigate this risk entirely, though, since some trials were performed in a single hospital so their participants could still be recognized in this way.

Remaining questions

- Would it even be possible to collect summary statistics by calendar date from trials? Maybe if we would write an R script that each data-uploader could run locally? Or would too many problems arise for the time we had to overcome them and would this not be any faster than sharing the raw data?
- Would it even be possible to ask all involved trials to work in R? By allowing the use of other software packages, we risk that trials share incorrect summary statistics. What we asked for was not trivial, since we analyze the data as left-truncated calendar time data. Even in R, no standard software outputs the right logrank statistic and we had to write our own.

Data transfer agreements

Sharing clinical trial data such as in Figure 2 involves agreeing on a Data Transfer Agreement (DTA) and signing it. These agreements protect the privacy of the participants in the trial but also cause an enormous delay. For some of the trials in our meta-analysis it took months for the lawyers on both ends to agree on the terms in the DTA. Interestingly, halfway this process a lawyer commented that we might not even have needed DTAs for data of the structure described in Figure 2.

Remaining questions

- Were these DTAs really necessary given the limited amount of data we asked for (Figure 2)?
- How does our requested data compare to full Kaplan-Meier plots, which are routinely included in medical publications?
- How do we convince trials in the future to share limited raw data without DTAs?

Estimation

So far, we have focused our discussion on testing the null hypothesis. This was the main aim in our ALL-IN meta-analysis: rejecting the null hypothesis in favor of an alternative

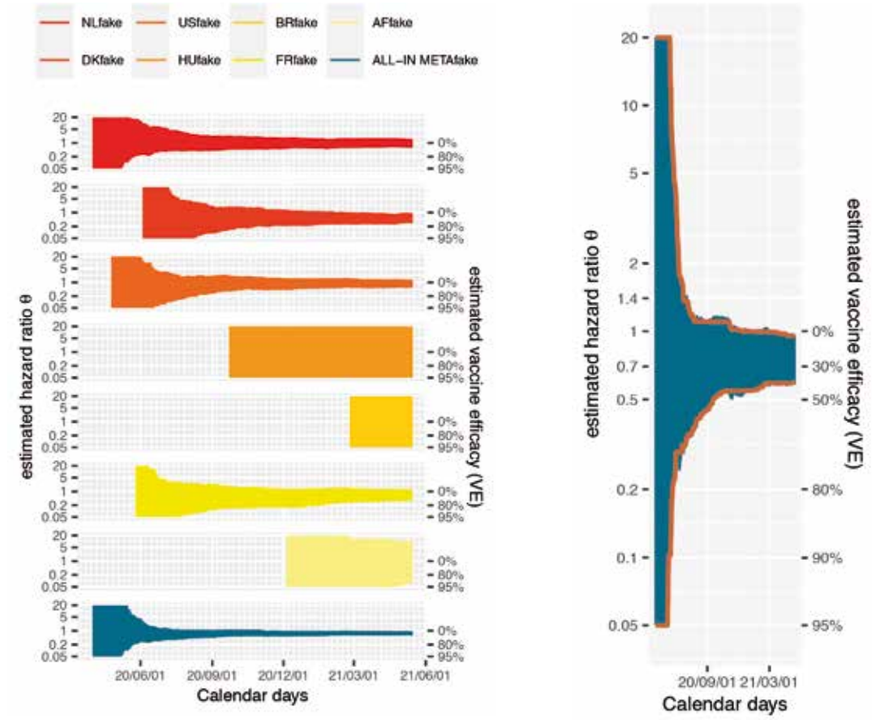


Figure 3. Anytime-valid confidence sequences corresponding to the fake data in Figure 1 with the running intersection for the meta-analysis sequence. Data generated according to the meta-analysis design with effect just slightly larger (hazard ratio = 0.7) than that of minimal interest (hazard ratio = 0.8)

hypothesis of minimal clinical relevance set at hazard ratio of 0.8 (see Safe design in Figure 1). Such a rejection could lead the conclusion of the meta-analysis and advice to stop the individual trials. A second aim is of course estimation. Here, two more disadvantages arise for meta-analysis on summary statistics that we fortunately did not encounter since we analyzed the raw data. First, a meta-analysis of time-to-event summary statistics is biased. This a technical point that we will not discuss in detail here. For practical purposes, it is common to settle for biased estimates (Simmonds et al., 2011). For our purposes, however, there is a second disadvantage of summary statistics. If we cannot collect the summary statistics for each calendar day, they produce wider confidence intervals.

Remember that a $(1-\alpha)$-confidence interval is a collection of parameter values that, if taken as the null hypothesis, each cannot be rejected at level α. This means that if we have a sequence of such confidence intervals, each of which is valid at any time, we can take its running intersection. Once a value for the parameter is rejected by an anytime-valid test, it never has to be re-included in the interval. A running intersection often achieves a narrower interval over time, as is shown in Figure 3

In summary, we believe that it was wise to not only collect summary statistics. Summary statistics are also known to be much more prone to mistakes and manipulation than IPD meta-analysis (Lawrence et all, 2021). Collecting the raw data indeed allowed us to turn data cleaning into a collective effort between the data uploader and the meta-trial statistician and to spot the inadvertent mistakes. We could also confirm a suspicion of insufficient randomization in one trial which let to its exclusion due to increased risk of bias.

The main question is still whether we could have done the same approach without the delay of data transfer agreements. Crucial here is maybe how different this approach was to usual research. The involved trials put the research line before their own publication; the collaboration before expanding their own academic CVs. If this becomes more commonplace, we might also view the need for DTAs very differently. Or the opposite is true and DTAs can serve a role in this transition from a science of individual interests to a science of live collaboration.

References

Grünwald, P., De Heide, R., & Koolen, W. (2019). *Safe Testing*. arXiv:1906.07801.

Lawrence, J. M. et al. (2021). The lesson of ivermectin: meta-analyses based on summary data alone are inherently unreliable. *Nature Medicine*.

Ly, A., Turner, R., Pérez-Ortiz, M. F., Ter Schure, J., & Grünwald, P. (2021). Rpackage safestats, 2021. Maintainer: Alexander Ly <a.ly@jasp-stats.org>, install in R by devtools::install_github("AlexanderLyNL/safestats", ref = "logrank", build_vignettes = TRUE).

Ter Schure, J., Pérez-Ortiz, M. F., Ly, A., & Grünwald, P. (2020a). The safe logrank test: Error control under continuous monitoring with unlimited horizon. *arXiv preprint* arXiv:2011.06931.

Ter Schure, J., Ly, A., & Grünwald, P. (2020b). *Safestats and ALL-IN meta-analysis project page*. https://projects.cwi.nl/safestats/.

Ter Schure, J., & Grünwald, P. (2021). ALL-IN meta-analysis: breathing life into living systematic reviews. *arXiv preprint* arXiv:2109.12141. 2021.

Simmonds, M. C., Tierney, J., Bowden, J., & Higgins, J. P. T. (2011). Meta-analysis of time-to-event data: a comparison of two-stage methods. *Research synthesis methods, 2*(3), 139–149.

Van der Wier, M. (2020). Helpt TBC-vaccin in the strijd tegen het coronavirus? *Trouw*, 18 maart 2020.

Cornelis H Van Werkhoven, C. H., Ter Schure, J., Bonten, M., Netea, M., Grünwald, P., & Ly, A. (2020). *Anytime Live and Leading Interim meta-analysis of the impact of Bacillus Calmette-Guérin vaccination in health care workers and elderly during the sars-cov-2 pandemic* (ALL-IN-META-BCG-CORONA).

Judith ter Schure is finalizing her Ph.D. on live meta-analysis at the Machine Learning group of CWI. She has been dividing her time between her Ph.D. research, freelance statistical consultancy (significanthelp.nl) and board membership (treasurer) of the VVSOR. She will start working as a clinical biostatistician next year.
E-mail: Judith.ter.Schure@cwi.nl

Alexander Ly is CTO at JASP (https://jasp-stats.org/) and a postdoctoral researcher in the Machine Learning group at CWI and the Psychological Methods group of the University of Amsterdam. He studies Bayes factors and sequential analysis.
E-mail: Alexander.Ly@cwi.nl

Peter Grünwald heads the Machine Learning group at CWI and is full professor of Statistical Learning at the Mathematical Institute of Leiden University. He founded the research line *safe statistics* that aims to make statistical inference more flexible while maintaining error control.
E-mail: Peter.Grunwald@cwi.nl

VU-hoogleraar Business Analytics Sandjai Bhulai en hoogleraar Toegepaste Wiskunde Rob van der Mei (Centrum Wiskunde & Informatica / VU) hebben de Huibregtsenprijs 2021 gewonnen met hun onderzoeksproject 'Wiskunde voor een veiliger en gezonder Nederland'. De prijs is 4 oktober uitgereikt op de Avond van Wetenschap & Maatschappij in de Nieuwe Kerk in Den Haag

Huibregtsenprijs 2021 voor Sandjai Bhulai en Rob van der Mei

Caroline Jagtenberg

Sandjai Bhulai en Rob van der Mei ontwikkelden elegante en efficiënte wiskundige oplossingen voor te late aankomsten van ambulances, wachttijden in de ouderenzorg, zelfmoordpreventie en het snel herkennen van nieuws op social media. De jury waardeerde dat hun onderzoek, ver buiten het eigen vakgebied, nadrukkelijk aan de weg timmert en ook letterlijk de straat weet te vinden.

De Huibregtsenprijs is in 2005 ingesteld de Stichting De Avond van Wetenschap & Maatschappij en vernoemd naar ir. Wouter Huibregtsen. De prijs is bestemd voor een recent onderzoeksproject dat wetenschappelijk kwaliteit en vernieuwing combineert met een bijzondere maatschappelijke meerwaarde of outreach. De winnaar wordt jaarlijks gekozen door een vakjury en de prijs bestaat uit een sculptuur, een workshop en een geldbedrag van € 25.000 voor onderzoeksactiviteiten. Caroline Jagtenberg interviewde de prijswinnaars.

Namens de STAtOR-redactie van harte gefeliciteerd met deze prachtige prijs. Was het een groot vooropgezet doel om wiskunde in te zetten voor een veiliger en gezonder Nederland, of hoe is dat balletje ooit gaan rollen?
Rob van der Mei: In 2007 kwam het toenmalige hoofd van de GGD-ambulancedienst op bezoek op het CWI en vroeg of we konden helpen want, zo zei hij: 'elke seconde telt'. Naast hem stond een ambulanceverpleegkundige in vol

Sandjai Bhulai en Rob van der Mei. Foto: Suzy Bhulai

ornaat en samen maakten ze veel indruk. Dit was een startschot voor het project dat we REPRO genoemd hebben, waaraan zes ambulance-aanbieders hebben meegedaan.
Sandjai Bhulai: Dus nee het was niet gepland. Op een gegeven moment kijk je terug en zie je veel projecten over gezondheidszorg, zoals ouderenzorg en vaccinatielogistiek, en we werken veel met ziekenhuizen. En van het een komt het ander. Doe je bijvoorbeeld een project met ambulances, dan komen de brandweer en politie zichzelf melden.

Hulpverlenende organisaties komen dus bij jullie met een probleem?
RvdM: Ja, vooral in meetings met niet-academici. Er zit dan bijvoorbeeld iemand in de zaal als je ergens spreekt, Dan komen mensen na afloop naar je toe en willen eens verder met je praten.
SB: Dat onderscheidt ons ook, dat we ook veel op niet-academische congressen spreken. In het begin moesten we wel echt op zoek naar projecten, maar later draaiden de rollen om.

Ambulances vaker op tijd laten rijden is één van de onderwerpen waar jullie deze prijs voor gewonnen hebben. Kunnen jullie iets meer vertellen over de andere onderzoeken?
RvdM: We hebben een project genaamd Dolce Vita: data-gedreven optimalisatie van acute ouderenzorg. Beeld je in: een oudere valt thuis van de trap en breekt een heup, komt zo bij de eerste hulp en moet daarna naar een verzorgingstehuis, krijgt langdurige zorgindicatie, er moet thuishulp komen, etc. Zo bevind je je in een complex systeem met allerlei aanbieders met elk hun eigen doelen en organisatie. De patiënt beweegt door zo'n systeem heen en met name op de transitiepunten tussen zorginstellingen gaat het vaak mis. Wij zijn bezig met het maken van een macromodel hoe je *what-if*-scenario's kan doorrekenen van beleidsbeslissingen. Als er bijvoorbeeld ergens geld wordt uitgetrokken om een deel van het systeem te verbeteren, waar zal dan een nieuwe bottleneck ontstaan? Een ander model is een slimme toewijzing van zorgcentra aan patiënten op basis van voorkeuren. Nu praten we ook met psychiatrie en jeugdzorg, want een gedeprimeerde jongere komt ook in een complex GGZ-systeem.
SB: Neem bijvoorbeeld ons onderzoek op het gebied van suïcidepreventie. We kijken naar welke combinaties van factoren in iemands leven ervoor zorgen dat iemand aan zelfdoding denkt. Bijvoorbeeld of iemand zijn baan kwijtraakt of een echtscheiding aanvraagt, of allebei, kun je daar patronen in vinden? Daarnaast komen er ook chatgesprekken binnen van mensen die om hulp vragen. Die kan je anonimiseren en risico's analyseren. Ons idee is om hulpverleners real-time te gaan helpen met suggesties van antwoorden. Dat is minder OR en meer statistiek, Natural Language Processing (NLP) en machine learning.
RvdM: Dat is 113 hè, ik vind het super interessant om daar rond te lopen en te zien hoe die hulpverleners echt de chats uitvoeren.
SB: Als je meeluistert met zo'n gesprek, dan doet het je wel wat. Dat iemand, vrij jong, zegt 'ik wil er morgen niet meer zijn'. Dat doet je echt wel wat.

De Huibregtsenprijs staat in het teken van wetenschap en maatschappij. In voorgaande jaren werd deze vaak uitgereikt aan onderzoekers in de geneeskunde, of soms natuurkunde; jullie zijn de eerste wiskundigen die hem winnen. Dit heeft ongetwijfeld te maken met de onderwerpen die je uitkiest, maar misschien ook wel met jullie vermogen om de kracht van OR

uit te leggen aan een breder publiek. Is er speciale aandacht voor wetenschapscommunicatie in jullie projecten?
SB: De jury zei expliciet dat als ze denken aan wiskunde, ze de toepassing meestal niet direct evident vinden. Bij onze projecten vonden ze dat juist sterk, dat we echt ons best hebben gedaan om die toepassing op te zoeken, en het echt te laten draaien in de praktijk. En daar hoort natuurlijk wel bij dat we er ook bekendheid aan geven. Als je kijkt hoeveel communicatie er rond het ambulance project is geweest, dat is fantastisch. Rob heeft bij Universiteit van Nederland gesproken over ambulancelogistiek, en ik heb op een TEDx event iets verteld over social media en wiskunde. Dat helpt wel.
RvdM: Wat ook helpt is dat we dat allebei leuk vinden. Je moet dat niet met tegenzin doen omdat het goed is voor acquisitie.

Welke momenten uit jullie onderzoek staan jullie nog het meeste bij? Wanneer voelde je écht dat je belangrijke dingen voor de maatschappij deed?
RvdM: Op het moment dat onze ambulancesoftware draaide in een pilot bij GGD Flevoland, toen de eerste cijfers binnenkwamen. Voorheen reden ze geloof ik 94% van hun ritten op tijd, en met onze software 96.8%. Die cijfers waren eigenlijk nog beter dan we gedacht hadden, dat was echt een yes-moment.
SB: Ik heb dat niet zo... Nou ja, met het Twitterproject gaan mensen twitteren en het verschijnt op het scherm. Dat is ook Yes! Maar meer zo van: dat mijn algoritme werkt. Een paar dagen nadat we de prijs wonnen moet je nog beseffen wat er gebeurd is, je gaat nadenken waarom we dit nou hebben gekregen. En dan pas ga je beseffen dat je toch wel wat gedaan hebt. Zo komen er buren naar je toe: 'Ik wist niet dat je zo'n belangrijk werk deed!'

Jullie weten met de gekozen onderwerpen veel maatschappelijke relevantie en zingeving te creëren binnen het beroep van OR-expert. Hebben jullie het idee dat dat is waar ons vakgebied naartoe gaat, of zou moeten gaan?
Beiden: Nee dat vind ik niet. Er moet ruimte blijven voor beide takken van sport. Als iedereen dit gaat doen is het ook niet meer bijzonder. Nederland heeft een groot aantal geweldige OR groepen die vooral fundamenteel onderzoek doen, en ook dat is heel belangrijk en moet blijven worden ondersteund.
SB: Ik vind het wel belangrijk dat de industrie/buitenwereld zich realiseert dat OR een belangrijke rol kan spelen. Dat de drempel lager wordt om samen te werken met industrie en dat er veel barrières weggenomen kunnen worden voor jong aanstormend talent.

Maatschappelijk relevante onderwerpen hebben een aantrekkingskracht op de jeugd. Daarmee kunnen projecten zoals die van jullie de populariteit van ons vakgebied vergroten. Hebben jullie de indruk dat dit inderdaad het geval is? Maken de onderwerpen waarmee jullie de Huibregtsenprijs wonnen het gemakkelijker om jong talent te recruiten?
RvdM: Ja, ik denk het wel. We geven hier op de universiteit ook voorlichting, en daarin hebben voorbeelden en succesverhalen – met name de toegepaste – een enorme aantrekkingskracht. Scholieren en hun ouders zitten dan enthousiast in de zaal. Dat zien we ook aan de studentenaantallen.
SB: Voor de popularisering van wiskunde is dit geweldig. Als je zo'n prijs wint is dit hét moment om wiskunde op de kaart te zetten. We zijn in zekere zin verplicht wiskundedocenten op middelbare scholen hierover te informeren. Zij kunnen dan hun leerlingen vertellen wat er allemaal kan met een wiskundestudie. Studenten moeten het ook horen, om ze te motiveren. En overheden moeten dit weten want er moet geld naar de wiskunde gaan.

Zou het een ding van deze tijd zijn, of zien jullie binnen de OR-wereld in de generatie boven jullie net zulke maatschappelijk relevante onderwerpen?
RvdM: Die geest is uit de fles, dit gaat niet meer terug.
SB: Er is meer data, data-analyse is prominenter geworden bij veel bedrijven. Dus het is wel een heel opportuun moment voor de OR om de expertise in te zetten die we al jarenlang hebben opgebouwd zonder data, om daar nu een data-gedreven versie van te maken.
RvdM: Dat is ook een kans voor ons vakgebied om nu de combinatie te zoeken met AI en snelle rekenkracht.

Naast de eer hebben jullie een geldbedrag voor onderzoek gewonnen, hebben jullie al besloten waar dat naartoe gaat? Staat er nog een mooi onderwerp in de wacht?
Beiden: Nee, we hebben nog niet de tijd en rust gehad om daarover na te denken. Maar we willen er wel een mooie bestemming voor vinden.
SB: Ik ben dankbaar aan mijn collega's dat wij dit kunnen doen.
RvdM: De vrijheid die we krijgen van onze werkgevers, en de waardering middels zo'n prijs is wel een stimulans om in die richting verder te gaan.

Caroline Jagtenberg is universitair docent aan de VU, School of Business and Economics. Van 2012 tot 2016 promoveerde ze onder begeleiding van Rob van der Mei en Sandjai Bhulai op het onderwerp ambulancelogistiek. Caroline is ook redacteur van *STAtOR*.
E-mail: c.j.jagtenberg@vu.nl

Sifan Hassan. Foto: Filip Bossuyt

Olympische medailleklassementen hebben inderdaad enig nut

De Olympische Winterspelen van 2022 staan voor de deur. Verwachtingen alom over het aantal medailles dat de Nederlandse delegatie zal behalen. Voor de 2020-Zomerspelen, vanwege COVID-19 gehouden in 2021, had Maurits Hendriks, technisch directeur van NOC*NSF, een olympische doelstelling: bij de eerste tien in het Olympisch Medailleklassement. En dat is gelukt, Nederland staat bij de eerste tien in dat klassement! Maar wat betekent dit eigenlijk? Dat wordt behandeld in deze als column begonnen, maar tot artikel uitgegroeide bijdrage.

Gerard Sierksma & Elmer Sterken

'We' staan op plek zeven en dat is nooit eerder vertoond. Nederland is een echt topsportland, ronkt het door de media. En net als bij andere grootschalige sportevenementen verbindt de natie zich aan dit succes. Hoe hoger in het Olympisch Medailleklassement (in Nederland ook wel Medaillespiegel genoemd), hoe groter de nationale trots. Dat is natuurlijk heel mooi, maar 'klopt' die lijst eigenlijk wel, en wat betekent zo'n zevende plaats eigenlijk? In de euforie van het succes komt deze vraag wellicht wat zurig over, maar is daarom niet minder relevant. Omdat dit medailleklassement zo prominent en breed internationaal wordt gehanteerd en gewaardeerd, doen we een

poging de volgende vragen te beantwoorden. Welk doel dient zo'n klassement precies? Waarom is 'bij-de-eerste-tien' een zinvolle beleidsdoelstelling van het NOC*NSF? En, waarom hanteert ons eigen NOC*NSF het eigenlijk als beleidsdoelstelling, terwijl het Internationaal Olympisch Comité officieel geen enkel medailleklassement hanteert, maar wel juist deze ranglijst op de website laat zien? En tot slot, is het niet zinvoller om ook te kijken naar de omvang van de atletenequipe waarmee de medailles zijn gewonnen?

Ieder zijn eigen olympisch medailleklassement

Het zou mooi zijn als er één uniek olympisch totaalklassement zou bestaan. Maar dat is helaas niet het geval. Vrijwel onmiddellijk na afloop van elke editie van de Olympische Spelen ontstaat discussie over de geschiktheid van landenklassementen met als centrale vraag: Wat brengen ze nou eigenlijk in kaart? De werkelijke prestaties van de landen die hebben deelgenomen? Overigens zijn het formeel geen landen maar Nationaal Olympisch Comités (NOC's), die aan Olympische Spelen deelnemen. In ieder geval valt op dat in vrijwel alle klassementen de rijke en grote landen in de top van de klassementen te vinden zijn. Maar hebben die 'rijken' ook beter gepresteerd dan de minder rijke en kleinere landen? Daarmee begint de discussie. En dus verschijnen er ranglijsten met voor elk land correcties op onder meer: het aantal medailles per inwoner, het aantal medailles gecorrigeerd voor het reële inkomen per hoofd van de bevolking en het aantal medailles per geschat reëel vermogen. Met dergelijke correcties is het geen verrassing dat de kleine landen door het 'noemereffect' in de top van zulke klassementen komen, met landen als San Marino, Bahama's en Bermuda als lichtende voorbeelden. Hoewel Nederland de top tien dan meestal niet haalt, is Tokio 2021 een uitzondering met een negende plaats in de bevolkingsomvang-gecorrigeerde ranglijst. Dit type correcties legt impliciet een relatie tussen het aantal behaalde medailles van een land en (slechts) een van de (mogelijk vele) onderliggende zogenaamde 'productiefactoren' van dat land.

Naast het bij ons populaire Olympisch Medailleklassement (tabel 1) hanteert de Verenigde Staten een heel eigen ranglijst gebaseerd op voor elk land het totaalaantal door dat land behaalde medailles, zonder onderscheid te maken tussen goud, zilver en brons. De reden van deze keuze is eenvoudig te raden: De VS is olympisch grootverdiener en in dat totaalklassement vrijwel altijd koploper.

Het Olympische Medailleklassement is opgesteld volgens de zogenaamde lexicografische rangschikking. Daarbij wordt voor elk land eerst het aantal gouden medailles geteld, vervolgens -bij gelijke stand- het aantal zilveren en bij een gelijk aantal zilveren medailles wordt gekeken naar het aantal keren brons. De Spelen van Tokio brachten TeamNL een totaal van 36 medailles (10 goud, 12 zilver, 14 brons), waarmee we volgens de VS-systematiek op plek negen staan. Naast deze klassementen bestaan nog een aantal minder bekende ranglijsten, waarbij andere wegingen worden toegekend aan de drie kleuren medailles. In de literatuur (zie bijvoorbeeld Sergeyev, 2015) zien we onder meer de zogenaamde Fibonacci-weging met de verhouding 3:2:1 voor goud:zilver:brons, de exponentiële weging met 4:2:1 en de Londen 1908-weging met 5:3:1. Als we die wegingen toepassen op TeamNL, komen wij op plaats acht volgens zowel de Fibonacci- als de exponentiële weging. Volgens de Londen 1908-weging komt Nederland op plaats zeven (gedeeld met Italië). Duidelijk is dat de VS de weging 1:1:1 hanteert, met alle 'kleuren' even zwaar gewogen. Merk op dat de zevende plek in het in Nederland gehanteerde medailleklassement gezien kan worden als het gevolg van een weging a:b:c, waarbij a veel groter is dan b en b weer veel groter is dan c.

Interessant is ook het model van Bredtmann *et al.* (2016) dat de *Financial Times* hanteert voor haar medaille-ranking. Eerst worden prognoses, op basis van onder meer bevolkingsomvang, inkomen en thuisvoordeel, van de medaillewinsten van elk land afzonderlijk berekend.

		GOUD	ZILVER	BRONS	TOTAAL
1	Verenigde Staten	39	41	33	113
2	China	38	32	18	88
3	Japan	27	14	17	58
4	Groot Brittanië	22	21	22	65
5	ROC	20	28	23	71
6	Australië	17	7	22	46
7	Nederland	10	12	14	36
8	Frankrijk	10	12	11	33
9	Duitsland	10	11	16	37
10	Italië	10	10	20	40

Tabel 1. Eerste tien landen (lexicografisch) Olympisch Medaille-klassement Tokio 2021

Vervolgens wordt voor elk land gekeken naar de grootte van de afwijking van de gerealiseerde medailleoogst ten opzichte van de geprognotiseerde medaillewinst. Deze afwijking bepaalt dan het relatieve prestatieniveau van dat land en wordt vervolgens gebruikt voor de ranking. De *Financial Times* raamde Nederland in Tokio op 21 medailles. Met 15 medailles meer dan die raming zet de *Financial Times* Nederland op plaats drie, achter het team van het Russisch Olympisch Comité (ROC) en Australië, maar wel vóór China, Japan en de VS. Duitsland en vooral Frankrijk zijn de grootste verliezers volgens deze ramingen.

Nog 'bonter' maakt *Google Trends* (2021) het met het hanteren van een 'media-aandachtvariabele' per land om een correctie door te voeren op de medailleoogst van dat land. *Google Trends* bepaalt voor elke variabele het aantal malen dat op het internet gezocht wordt naar de betreffende olympische sportprestaties. Nederland zakt met *Google Trends* naar plaats 46; Japan staat bovenaan.

De voorsprong van rijke landen

Met bovenstaande overwegingen is het overduidelijk dat er meerdere aspecten van belang zijn en meegenomen moeten worden om de positie van een NOC te bepalen in een totaalklassement (zie bijvoorbeeld De Hoog, 2021). De zogeheten *univariate* correcties gaan ervan uit dat slechts één eenvoudige en overheersende variabele de prestatie verklaart. Bij het meten van relatieve olympische landenprestaties is dat vanzelfsprekend een tamelijk ongelukkige veronderstelling. Natuurlijk betekent dit niet dat zoiets als 'bevolkingsomvang van het land' geen positieve determinant kan zijn van de medaillewinst van dat land. Enerzijds hebben landen met relatief veel actieve sporters in de regel meer keuze om ervaren sporters af te vaardigen naar de Spelen, terwijl anderzijds er ook bevolkingsrijke landen zijn, zoals India en Bangladesh, waar olympisch succes zeldzaam is.

Maar toch. Vrijwel altijd hebben de rijke landen, gemeten aan de hand van hun reële inkomen per hoofd van de bevolking (een variabele die overigens niet sterk correleert met de bevolkingsomvang) meer olympisch succes dan de armere. Rijke landen kunnen zich een vergaande arbeidsspecialisatie veroorloven en 'vrije tijd' met sport invullen. Ook kunnen de rijke landen meer investeren in training, technologie en lobbyactiviteiten (om bijvoorbeeld nieuwe voor hen kansrijke sportonderdelen op het programma te krijgen). Vanzelfsprekend is ook de omvang van de ploegen van de rijke landen meestal groter.

Er is meer dan rijk alleen

De analyse over de rol van rijk-zijn is evenwel niet het hele verhaal. Het ooit zo uiterst succesvolle, relatief rijke, Finse NOC moet het de laatste edities van de Zomerspelen doen met een paar schamele bronzen medailles. In het algemeen blijken bevolkingsomvang en rijkdom slechts ongeveer voor de helft de olympische medaillewinst te verklaren (zie Bredtmann *et al.*, 2016). Ook politieke en culturele invloeden, en -niet te vergeten- 'thuisvoordelen', spelen vaak belangrijke rollen. In de periode van de Koude Oorlog, direct na de Tweede Wereld Oorlog, waren de toenmalige Sovjet-Unie (USSR) en Oost-Duitsland (DDR), als sterk centraal geleide landen, grootverdieners op de olympische medaillemarkten en was sportsucces een belangrijk politiek wapen. Tegenwoordig zijn het landen met een sterk ontwikkelde vrouwenemancipatie die relatief succesvol zijn, ook omdat het aantal sportevenementen voor vrouwen en gemengde teams een sterke toename kent. Landen waar de vrouwenemancipatie nog in ontwikkeling is blijven daardoor achter in de race om olympisch succes (zie ook hier Bredtmann *et al.* 2016).

Ook het zogenaamde thuisvoordeel speelt een rol. Dit voordeel houdt niet alleen het bekende succes tijdens het thuisevenement in, maar ook dat een NOC met een succesvolle editie van de Spelen achter de rug, opvallend vaak dit succes weet door te zetten naar daaropvolgende Olympische Spelen en tussenliggende wereldkampioenschappen. De preolympische investeringen van organiserende landen betalen zich gedurende enige jaren uit.

Dus nogmaals, er zijn altijd meerdere verklarende variabelen van medaillesuccessen en elke partiële analyse mist een deel van het verhaal. In plaats van de veel gehanteerde *univariate* benadering is dus een *multivariate* weging van de medaille-aantallen evident geschikter. Jammer is dat de bijbehorende statistische berekeningen daarvan tamelijk ingewikkeld kunnen zijn, met als mogelijk gevolg dat de resultaten minder eenvoudig te communiceren zijn. Bovendien, als we de top-10 van het klassement van de *Financial Times* van Tokio 2021 bezien staan daar landen als Taiwan (3 medailles verwacht, 12 gekregen), Oekraïne (13, 19), Zwitserland (7, 13), Oostenrijk (1, 7) en Noorwegen (3, 8) op de plaatsen zes tot en met tien. De vraag is of het brede publiek dit ook als echte olympische prestaties beschouwt.

Het huidige medailleklassement van het NOC*NSF is zo slecht nog niet

Bestaat er dan geen 'eerlijk' klassement? Misschien wel niet. Simpel de gewonnen medailles van elk land tellen en rangschikken snapt iedereen, zeker met de weging 1:1:1 of met 'onze' lexicografische ordening. Dat landen als de VS, China en Rusland altijd bovenaan staan en vele andere landenprestaties ver ondergewaardeerd worden moet dan op de koop toegenomen worden. Die simpelheid van het 'lexicografische' klassement kan ook de reden zijn dat het NOC*NSF het hanteert voor haar olympisch selectiebeleid. Als afgeleide van haar doelstelling 'een plek bij de eerste tien van dit klassement', hanteert het NOC*NSF als globale selectienorm voor de deelnemende atleten: 'een grote kans op een plaats bij de laatste acht op de Spelen'. Voor de operationele vertaling ervan hanteert het NOC*NSF juridisch verantwoorde protocollen, hoewel de veronderstelde relatie tussen de top 10-doelstelling voor het land als geheel en de top 8-doelstelling voor de individuele atleten en teams nauwelijks onderbouwd is. Mooi van de keuze van het NOC*NSF voor dit landenklassement is dat de positie van Nederland tussen landen als Duitsland, Frankijk, Spanje en Australië een goede maat is voor de kracht en de prestatie van TeamNL. Dat de VS en China in dit klassement altijd van een 'andere planeet' zijn wordt gemakshalve dan terecht *for granted* genomen. Maar hiermee is, wat ons betreft, niet alles gezegd.

Rendement als basis voor een nieuw klassement

Een berekening die meer recht doet aan de echte landprestatie zal vrijwel zeker dus ten koste gaan van die 'simpelheid'. De moed om tot verbetering te komen opgeven betekent dat we blijven zitten met de onbevredigende situatie dat rijke landen heel veel goede atleten kunnen afvaardigen en dus de medaille-spekkopers zijn en dat ook blijven. Om het scherp te stellen het volgende voorbeeld. Stel dat land A met een afvaardiging van 60 atleten vijf gouden medailles gewonnen heeft en land B met 30 atleten vier keer goud. Dan komt land A boven land B in het Olympisch Medailleklassement. Maar is het dan echt wel terecht dat de prestatie van land A als beter wordt beschouwd dan die van land B? Of zouden we moeten 'corrigeren' voor de omvang van de atletenafvaardiging en dus het medaillerendement van de afvaardiging moeten berekenen? Hoe effectief was een NOC in de medailleruif gezien de omvang van haar afvaardiging? Een effectiviteits- of rendementsbenadering dient een relatie te leggen tussen de totale medailleoogst en de omvang van de atletenequipe. Hierbij moet vanzelfsprekend rekening gehouden worden met de situatie bij teamsporten, waar bijvoorbeeld een hockeyteam met 16 atleten maar één medaille kan winnen. Ook zijn er vechtsporten (boksen, judo, karate, taekwondo en worstelen: vrije-stijl en Grieks-Romeins) waarbij twee bronzen medailles te winnen zijn. De deelname van een hockeyteam beschouwen we als één 'startdeelname', terwijl bijvoorbeeld een team van vier wielrensters van hetzelfde land in de wegwedstrijd voor vier 'startdeelnames' telt, immers alle vier wielrensters komen in de uitslag van de wedstrijd voor. Zo krijgt een hockeyteam één plek in de eindrangschikking van het hockeytoernooi, ondanks het feit dat alle speelsters bij winst een gouden medaille krijgen. Zo kunnen we van alle NOC's uitrekenen hoeveel gouden, zilveren en bronzen medailles er behaald zijn als (rendements-) percentage van het totale aantal startdeelnames van dat NOC. Tabel 2 geeft de lexicografische top-20 aangevuld met de zeven landen met tenminste 250 atleten in haar afvaardiging. In kolom 2 geven we aan op welke plaats het NOC in het Olympisch Medailleklassement (tabel 1) staat. In kolom 4 staat het aantal startdeelnames en in kolom 5 het aantal startdeelnames in de onderdelen met twee bronzen medailles. Het goudrendement is het percentage gouden medailles van alle startdeelnames. Voor de berekening van het bronsrendement gebruiken we het aantal startdeelnames plus het aantal dubbel-brons evenementen als noemer.

In, wat we noemen, het Olympisch Rendementsklassement verschijnt Bermuda als lijstaanvoerder. Het fabelachtige goudrendement van 50% van Bermuda is afkomstig van twee startdeelnames en één gouden medaille: roeister Dara Alizadeh en de gouden triatleet Flora Duffy. In de top-5 staan verder nog de kleine landen Kosovo, Qatar, Fiji en Bahama's. De nummer twee van tabel 1, China, verschijnt op plaats zes; de VS nog lager op plaats acht. Nederland staat op plek 20, maar wel weer voor de grote Europese landen.

De 'vechtsport' NOC's, met veel dubbelbrons en nauwelijks teamsporten, vinden we hoofdzakelijk terug op die plaatsen in tabel 2 waar het aantal startdeelnames groter is dan de omvang van de atletenequipe. Tussen de Aziatische, Midden-Amerikaanse en Oost-Europese vechtsportlanden staat verrassend genoeg ook het wintersportland Oostenrijk op plaats 59 met 75 atleten en 76 startdeelnames.

Zoomen we wat verder in op het verschil tussen de omvang van de atletenafvaardiging en het aantal startdeelnames van die afvaardiging, dan vallen de volgende vier zaken op. Eén: De VS had in Tokyo niet alleen de grootste atletenafvaardiging (615), maar ook het hoogste aantal startdeelnames (472). Twee: Hoewel het verschil 143 (615 – 472) groot is vergeleken met andere NOC's, is de VS hier toch niet de allergrootste: Australië met 169 (477 – 308) en Japan met 201 (556 – 355) scoren duidelijk hoger, wat hier dus niet betekent 'beter'. Drie: Een dergelijk (groot) verschil duidt natuurlijk op de aanwezigheid van meerdere teams in de afvaardiging. Het thuisland Japan had met het verschil van 201 kennelijk veel teams in de geledingen, zeker als we daarbij in acht nemen dat het aantal dubbelbrons-startdeelnames 43 bedroeg. Vier: Alleen het Russische ROC had, met 47 verschil, meer dubbelbrons-startdeelnames. Ondanks dit grote aantal

POSITIE	RANG IOC	LAND	STARTDEELNAMES	DUBBEL BRONS	GOUDRENDEMENT	ZILVERRENDEMENT	BRONSREDEMENT
1	63	Bermuda	2	0	50,00	0,00	0,00
2	43	Kosovo	11	7	18,18	0,00	0,00
3	41	Qatar	15	1	13,33	0,00	6,25
4	60	Fiji	8	1	12,50	0,00	11.11
5	42	Bahamas	16	0	12,50	0,00	0,00
6	2	China	313	31	12,14	10,22	5,23
7	14	Cuba	72	26	9,72	4,17	5,10
8	1	VS	472	35	8,26	8,69	6,51
9	19	Kenia	52	5	7,69	7,69	3,51
10	4	VK	288	22	7,64	7,29	7,10
11	3	Japan	355	43	7,61	3,94	4,27
12	37	Uganda	28	3	7,14	3,57	3,23
13	21	Jamaica	57	2	7,02	1,75	6,78
14	30	Bulgarije	44	14	6,82	2.27	3,45
15	13	Nieuw Zeeland	106	3	6,60	5,66	6,42
16	5	ROC	309	47	6,47	9,06	6,46
17	31	Slovenië	47	6	6,38	2,13	1,89
18	27	Iran	50	19	6,00	4,00	2,90
19	33	Georgië	35	20	5,71	14,29	1m82
20	7	Nederland	179	15	5,59	6,70	7,22
22	6	Australië	308	13	5,52	2,27	6.85
34	10	Italië	289	20	3,46	3,46	6,47
35	8	Frankrijk	290	25	3,45	4,14	3,49
36	12	Brazilië	204	27	3,43	2,94	3,46
37	9	Duitsland	313	29	3,19	3,51	4,68
45	11	Canada	257	18	2,72	2,33	4,00
58	22	Spanje	195	17	1,54	4,10	2,83

Tabel 2. Het Olympisch Rendementsklassement Tokyo 2021

'dubbelbronzers' van ROC, valt hier ook het lage verschil tussen de omvang van de Russische equipe en hun aantal startdeelnames op, namelijk 335 – 309 = 29, wat duidt op relatief weinig teams. Vijf: Nederland zit met het verschil 278 – 179 = 99 tussen de grote Europese landen. Zes: Mexico kunnen we zien als de grote verliezer van de Spelen in Tokyo. Met 164 atleten en 112 startdeelnames scoort het middelgrote Mexico het schamele aantal van vier keer brons, met dus een bronsrendement van 0,33%. Goed voor de laatste plaats in het Olympisch Rendementsklassement.

Een disclaimer

Nu maakt het wel een verschil of een NOC met slechts maximaal 20 startdeelnames een hoog medaillerendement haalt of dat landen als China en de Verenigde Staten dit met enige honderden startdeelnames doen. De onzekerheid rond de schatting van het 'echte rendement' is voor veelverdienende NOC's geringer dan voor de kleine landen met heel weinig medailles. Een manier om deze gevoeligheid in kaart te brengen is door de noemer van het rendement met een uniform vast aantal op te hogen. Zo'n scenario-analyse toegepast op tabel 2, houdt Bermuda met een ophoogfactor 5 nog steeds op plaats één. China komt pas met een factor 10 weer op de topposition. Nederland komt met een ophoogfactor 10 op plek 17 terecht. Dat is nog steeds heel mooi, maar voldoet niet aan de Hendriks-norm.

Tot slot. Het Olympisch Rendementsklassement als toegevoegde waarde

Het Olympische Rendementsklassement geeft een mooi beeld van de effectiviteit van de inzet van de peperdure 'middelen', de atleten, waarmee de medailles worden verworven. Omdat dit klassement een wissel kan gaan trekken op het zorgvuldig selecteren van deelnemende atleten, krijgen juist de minder rijke NOC's de mogelijkheid zich te onderscheiden door met weinig atleten relatief veel medailles te scoren. Overigens is het daarbij voorstelbaar dat de rijkere landen zich weinig gelegen zullen laten aan rendementsoverwegingen en altijd veel atleten blijven sturen. Andere landen, waaronder Nederland, hanteren al een uiterst kritisch selectiebeleid en zullen derhalve vast wel geïnteresseerd zijn in de positie op de rendementslijst. Zo hanteert Nederland met haar selectiecriterium van 'een grote kans hebben bij de laatste acht op de Spelen te eindigen' feitelijk reeds de rendementsgedachte. Ook het IOC kan niet anders dan een dergelijk kritisch selectiebeleid toejuichen, niet alleen vanwege de kwaliteitsgarantie en de kostenbesparing die dat oplevert, maar ook vanwege de ruimte die die houding kan scheppen om nog meer landen te kunnen uitnodigen mee te doen aan de Olympische Spelen. Al met al redenen genoeg toch om het bekende Olympische Medailleklassement te combineren met ons Olympisch Rendementsklassement. Bovendien zou een hoge rendementsscore van minder rijke NOC's door het Internationaal Olympisch Comité beloond kunnen worden met extra financiële middelen ter stimulering van de sport in die landen. En zo krijgt een olympisch landenklassement, binnen de oude olympische gedachte van 'meedoen is belangrijker dan winnen', een hedendaagse competitie-component van 'effectief en efficiënt strijden om de top van de Olympus'.

De volledige ranglijsten en rendementsgrafieken zijn opvraagbaar bij e.sterken@rug.nl.

Literatuur

Bredtmann, J., Crede, C. J., & Otten, S. (2016). Olympic Medals: Does the Past Predict the Future? *Significance* 13(3), 22–25.

Google Trends (2021). *Tokyo 2020; Alternative Olympics Medal Table*, https://olympics-trends. appspot.com.

Hoog, M. de. (2021). Deze Medaillespiegel is Niks. Wat is een Betere? *De Correspondent,* 7 augustus 2021.

Sergeyev, Y. (2015). The Olympic Medals Ranks, Lexicographic Ordering and Numerical Infinities, *The Mathematical Intelligencer* 37(2), 4–8.

Sierksma, G., & Talsma, B.G. (2021). The Dutch Approach in Selecting Olympic Speed Skaters. *OR/MS Today* 48(2), 43–47.

Talsma. B. G., Sierksma, G., & Turkensteen, M. (2017). The Growing Problem of Comparing Elite Sport Performances: The Olympic Speed Skating Case. *Journal of Human Sport and Exercise* 13, S892–S907.

Sterken, E. (2017). Performance Development at the Olympic Games. In: Albert, J., Glickman, M.E., Swartz, T.B. & Koning, R.H. (eds.). *Handbook of Statistical Methods and Analyses in Sports*. CRC Press, 461–484.

Gerard Sierksma is emeritus-hoogleraar Operations Research aan de Rijksuniversiteit Groningen. Sierksma adviseert, in samenwerking met ORTEC-Sports, het NOC*NSF over olympische selectieprocedures.

Elmer Sterken is hoogleraar Monetaire Economie aan de Rijksuniversiteit Groningen. Sterken heeft gepubliceerd over olympische medaillevoorspellingen en ontwikkelingen van wereldrecords in de atletiek en de schaatssport.
E-mail: e.sterken@rug.nl