



EEN AANPASSING OP DE NON-PARAMETRISCHE EFFECTMAAT CLIFFS DELTA VOOR DE VERGELIJKING VAN GEDRAGSPROFIELEN

Cliffs Delta is een non-parametrische effectmaat die is gebaseerd op data-observaties. We passen Cliffs Delta aan om er gedragsprofielen mee te vergelijken. Gedragsprofielen zijn dichtheidsverdelingen waarin antwoordgedrag van surveys is samengevat voor specifieke groepen respondenten of items. De aangepaste Cliffs Delta kan twee groepen respondenten vergelijken (bijvoorbeeld hoog- versus laagopgeleide respondenten) op de frequentie van specifiek antwoordgedrag (bijvoorbeeld het antwoorden van ‘weet niet’). De aangepaste Cliffs Delta is een solide en conservatieve effectmaat die zowel bruikbaar als gunstig is om gedragsprofielen met elkaar te vergelijken.

FRANK BAIS & JOOST VAN DER NEUT

Kwaliteit van surveydata kan worden beïnvloed door respondentkenmerken en itemkenmerken. Zulke kenmerken, bijvoorbeeld opleidingsniveau van de respondent of moeilijk taalgebruik in een item, kunnen leiden tot ongewenst antwoordgedrag, zoals het antwoorden van ‘weet niet’. We nemen aan dat elke individuele respondent een

latente of ‘ware’ kans (uitgedrukt in proportie) heeft om bepaald antwoordgedrag te laten zien. Een uitdaging is hoe om te gaan met de onzekerheid die gepaard gaat met de relatieve gedragsfrequentie of gedragskans van respondenten. Hoe kleiner het aantal items dat een respondent invult voor een schatting van bepaald antwoordgedrag,

hoe onzekerder het is dat de resulterende gedragskans naar de ‘ware’ gedragskans van die respondent verwijst. Daarnaast bestaan veel surveys deels uit filtervragen, waardoor sommige respondenten minder items invullen dan andere respondenten. Dit betekent dat respondenten onderling veelal niet makkelijk te vergelijken zijn op hun gedragskans vanwege een verschillend aantal ingevulde items. Om met deze onzekerheden rekening te houden, zodat we groepen respondenten kunnen vergelijken, kunnen we gedragsprofielen construeren.

Een gedragsprofiel is een dichtheidsverdeling die een visuele en statistische samenvatting van antwoordgedrag geeft voor een groep respondenten of een groep items met bepaalde kenmerken, respectievelijk het respondentprofiel en het itemprofiel. Een gedragsprofiel representeert de relatieve kans van een groep respondenten (bijvoorbeeld hoogopgeleide respondenten) of een groep items (bijvoorbeeld items die moeilijk taalgebruik bevatten) op het vertonen van specifiek antwoordgedrag (bijvoorbeeld het antwoorden van ‘weet niet’) voor alle mogelijke kansen van 0 tot 1. Zie figuur 1 voor enkele voorbeelden van gedragsprofielen. Behalve de technische notatie en de uitwisselbaarheid van de termen ‘respondent’ en ‘item’ verloopt het construeren van respondentprofielen en itemprofielen hetzelfde. Voor de overzichtelijkheid behandelen we daarom vanaf hier alleen respondentprofielen.

Het optreden van antwoordgedrag voor een enkele respondent r wordt bepaald door het aantal ingevulde items I_r waarvoor dat gedrag had kunnen worden vertoond en het aantal items G_r waarvoor het gedrag daadwerkelijk is vertoond:

$$\lambda_r(p) = \binom{I_r}{G_r} p^{G_r} (1-p)^{I_r-G_r} \quad \text{en} \quad \tilde{\lambda}_r(p) = \frac{\lambda_r(p)}{\int_{p=0}^1 \lambda_r(p) dp}. \quad (1)$$

Hier is λ de kanswaarschijnlijkheid, λ_r is het individuele gedragsprofiel voor respondent r , p is de kans tussen 0 en 1, en $\tilde{\lambda}_r$ is het genormaliseerde individuele gedragsprofiel

voor respondent r opdat elk individueel profiel een vergelijkbare oppervlakte krijgt van 1. Vervolgens kan voor een bepaalde groep respondenten het gedragsprofiel worden berekend door het gemiddelde van alle genormaliseerde individuele profielen te nemen:

$$\bar{\lambda}(p) = \frac{1}{R} \sum_{r=1}^R \tilde{\lambda}_r(p) \quad \text{en} \quad \bar{E} = \int_{p=0}^1 p \bar{\lambda}(p) dp. \quad (2)$$

Hier is $\bar{\lambda}$ het gedragsprofiel, R is het aantal respondenten in de groep, en \bar{E} is de verwachte waarde of kans voor de groep op het gedrag. Zie Bais (2021) voor een uitvoerige verhandeling over gedragsprofielen en hun eigenschappen.

Om twee gedragsprofielen (voor groepen van bijvoorbeeld hoogopgeleide en laagopgeleide respondenten) met elkaar te kunnen vergelijken (op bijvoorbeeld het antwoorden van ‘weet niet’), hebben we een statistische maat nodig. In dit artikel introduceren we daarvoor een aanpassing op de bestaande effectmaat Cliffs Delta.

Cliffs Delta voor gedragsprofielen

Cliffs Delta δ is ontwikkeld door Norman Cliff voor het gebruik met ordinale data (Cliff, 1993, 1996ab). De effectmaat berekent de kans dat een willekeurige data-observatie x_a van een groep A groter is dan een willekeurige data-observatie x_b van een groep B , min de omgekeerde kans (Hess & Kromrey, 2004; Rousselet, Foxe, & Bolam, 2016; Rousselet, Pernet, & Wilcox, 2017):

$$\delta = p(x_a > x_b) - p(x_a < x_b). \quad (3)$$

De steekproefschatting $\hat{\delta}_O$ van Cliffs Delta wordt verkregen door elke data-observatie in groep A te vergelijken met elke data-observatie in groep B :

$$\hat{\delta}_O = \frac{\sum_{a=1}^{R_A} \sum_{b=1}^{R_B} \text{sgn}(x_a - x_b)}{R_A R_B}. \quad (4)$$

De functie $\text{sgn}(x_a - x_b)$ resulteert in 1, 0, of -1 wanneer respectievelijk $x_a > x_b$, $x_a = x_b$, of $x_a < x_b$. Het totaal aantal vergelijkingen is het product van de groepsgroottes R_a en R_b van respectievelijk groep A en B. Hoe kleiner de overlap tussen de verdelingen van twee groepen, hoe meer de groepen verschillen. Een $\hat{\delta}_o$ van -1 of 1 betekent afwezigheid van overlap tussen twee groepen en een $\hat{\delta}_o$ van 0 betekent gelijkheid van groepen (Hess & Kromrey, 2004).

Om $\hat{\delta}_o$ te transformeren in een aangepaste Cliffs Delta die gedragsprofielen kan vergelijken, realiseren we ons dat elke observatie van een groep A precies één keer met elke observatie van een groep B wordt vergeleken voor de berekening van $\hat{\delta}_o$. Voor elke afzonderlijke vergelijking hebben beide observaties aldus een 'frequentie' of 'gewicht' van 1. Als we deze frequenties toevoegen aan formule (4) verkrijgen we

$$\hat{\delta}_o = \frac{\sum_{a=1}^{R_A} \sum_{b=1}^{R_B} \text{sgn}(x_a - x_b)(w_a w_b)}{\sum_{a=1}^{R_A} \sum_{b=1}^{R_B} (w_a w_b)}, \quad (5)$$

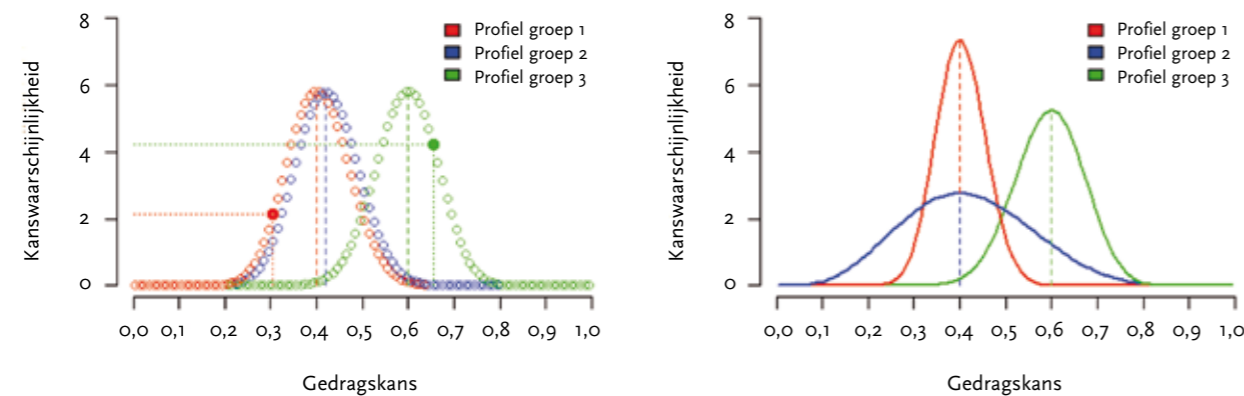
waar w_a en w_b de frequenties van respectievelijk de data-observaties x_a en x_b van de groepen A en B zijn. Wanneer we dit idee toepassen op een aangepaste Cliffs Delta $\hat{\delta}_p$ voor gedragsprofielen, dan zouden we de gedragskansen van 0 tot 1, met een bepaalde stapgrootte, als onze 'observaties' kunnen beschouwen en de waarschijnlijkheden voor elke kans als hun bijbehorende 'frequentie' of 'gewicht':

$$\hat{\delta}_p = \frac{\sum_{a=1}^A \sum_{b=1}^B \text{sgn}(p_a - p_b) \bar{\lambda}_A(p_a) \bar{\lambda}_B(p_b)}{\sum_{a=1}^A \sum_{b=1}^B \bar{\lambda}_A(p_a) \bar{\lambda}_B(p_b)}, \quad (6)$$

waar p_a en p_b de kansen van 0 tot 1, en $\bar{\lambda}_A$ en $\bar{\lambda}_B$ de gemiddelde kanswaarschijnlijkheden voor groep A en B respectievelijk zijn. Voor de gedragskansen kiezen we een stapgrootte van 0,01, zodat we voor elk profiel 100 intervallen verkrijgen. De middelste waarden van elk interval, $\{0,005, 0,015, 0,025, \dots, 0,995\}$, kunnen we als onze 'observaties' beschouwen die alle hun eigen gewicht hebben in de vorm van een kanswaarschijnlijkheid.

Zie grafiek 1 in figuur 1 voor een voorbeeldpaar van observaties in het vergelijken van groep 1 en groep 3. Beschouw de gedragskansen van 0,305 en 0,655 en hun respectievelijke waarschijnlijkheden van ongeveer 2,1 en 4,2 (zie de gestippelde lijnen). Het plaatsen van deze waarden in de teller van formule (6) geeft $\text{sgn}(0,305-0,655)$ ($2,1 * 4,2$) en levert een negatieve contributie aan de totale som van de teller. Alle positieve en negatieve contributies van alle mogelijke paren gedragskansen worden opgeteld. Het resultaat wordt gedeeld door de som van alle mogelijke paren waarschijnlijkheden. Evenals de originele $\hat{\delta}_o$ zal de aangepaste $\hat{\delta}_p$ altijd tussen de -1 en 1 liggen.

In het vergelijken van gedragsprofielen houdt $\hat{\delta}_p$ rekening met zowel de locatie van de profielen ten opzichte van de gedragskansen als met de vorm van de profielen. Zie grafiek 1 in figuur 1. Wanneer twee groepen en daarmee hun verwachte gedragskansen dichtbij elkaar liggen, is hun overlap groot en ligt $\hat{\delta}_p$ relatief dichtbij 0 (zie groep 1 en 2). Wanneer twee groepen en hun verwachte gedragskansen verder van elkaar vandaan liggen, is hun overlap kleiner en ligt $\hat{\delta}_p$ verder bij 0 vandaan (zie groep 1 en 3).



1. Respondentprofielen voor groep 1, 2 en 3: invloed van profiellocatie 2. Respondentprofielen voor groep 1, 2 en 3: invloed van profielvorm

Figuur 1. Voorbeelden van respondentprofielen met verschillende locaties (grafiek 1) en verschillende vormen (grafiek 2)

Bij het vergelijken van groep 1 met groep 3 vinden we een grote negatieve $\hat{\delta}_p$, wat betekent dat de gedragskans hoger is voor groep 3 dan voor groep 1.

Zie grafiek 2 in figuur 1. Wanneer een profiel relatief 'uitgestrekt' is over de breedte van de gedragskansen (zie groep 2) betekent dit dat er onzekerheid over de verwachte gedragskans voor de groep bestaat. In het algemeen betekent dit dat respondenten weinig items hebben ingevuld. Een uitgestrekt profiel heeft over het algemeen veel overlap met een ander profiel, waardoor $\hat{\delta}_p$ relatief dichtbij 0 ligt (zie groep 2 en 3). Wanneer een profiel relatief 'samengeperst' is (zie groep 1) betekent dit dat de verwachte gedragskans relatief zeker is. In het algemeen betekent dit dat respondenten relatief veel items hebben ingevuld. Een samengeperst profiel heeft over het algemeen weinig overlap met een ander profiel, waardoor $\hat{\delta}_p$ verder bij 0 vandaan ligt (zie groep 1 en 3).

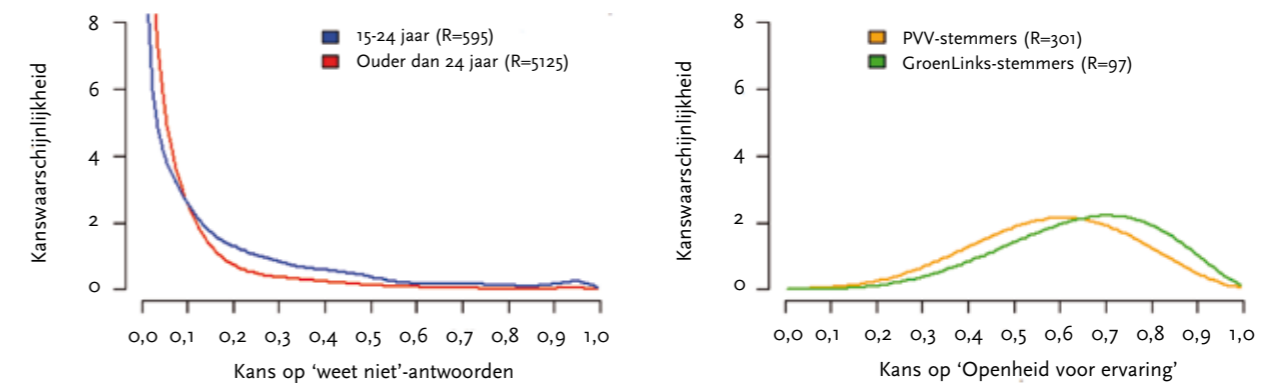
Het vergelijken van twee gedragsprofielen met $\hat{\delta}_p$ heeft veel voordelen. Zo maakt $\hat{\delta}_p$ geen assumptie over de vorm van de onderliggende verdelingen (Cliff, 1993, 1996ab; Goedhart, 2016; Vargha & Delaney, 2000) en is $\hat{\delta}_p$ robuust in het geval van outliers en scheve of andere niet-normale verdelingen (Goedhart, 2016). De effectmaat is makkelijk te berekenen en interpreteren, en gestandaardiseerd, wat betekent dat verschillende categorieën van effectgrootte kunnen worden onderscheiden (Goedhart, 2016). Voor onze $\hat{\delta}_p$ zijn relatief kleine of ongelijke groepsgroottes niet problematisch.

Twee voorbeelden

Nu laten we twee voorbeelden met surveydata zien die elk uit twee respondentprofielen bestaan waarvoor we $\hat{\delta}_p$ berekenen. We gebruiken data van het door CentERdata geadmistrateerde LISS Panel. We hanteren een bootstrapprocedure om 99% betrouwbaarheidsintervallen voor $\hat{\delta}_p$ te berekenen.

Zie grafiek 1 in figuur 2. We hebben respondenten van 15-24 jaar vergeleken met respondenten ouder dan 24 jaar op het geven van 'weet niet'-antwoorden voor items van de survey 'Politiek en Waarden' (wave 6, 2013). Vrijwel alle respondenten in beide groepen vulden 64 tot 67 items in waarbij een 'weet niet'-antwoord een optie was. De resulterende $\hat{\delta}_p$ van 0,28 kan worden gecategoriseerd als een 'medium' effect (Goedhart, 2016; Vargha & Delaney, 2000). We concluderen dat respondenten van 15-24 jaar meer 'weet niet'-antwoorden geven dan respondenten ouder dan 24 jaar op items over politieke inhoud. Zie Bais (2021) voor een eerste praktische implementatie van $\hat{\delta}_p$ voor het exploreren van de relatie tussen respondentkenmerken en antwoordgedrag.

Zie grafiek 2 in figuur 2 voor een andersoortig voorbeeld om de brede bruikbaarheid van $\hat{\delta}_p$ te illustreren. We onderzoeken de vraag of individuen die stemmen op GroenLinks mogelijk hoger scoren op de persoonlijkheidstrek 'openheid voor ervaring' dan individuen die stemmen op de PVV. Op basis van antwoorden op de vraag uit de survey 'Politiek en Waarden' op welke



1. Leeftijd: 'weet-niet'-antwoorden; Survey 'Politiek en Waarden'; Geschatte Cliffs Delta = 0,28; 99% Betrouwbaarheidsinterval = (0,22, 0,34)

2. Openheid voor ervaring voor PVV-stemmers vs. GroenLinks-stemmers; Geschatte Cliffs Delta = 0,21; 99% Betrouwbaarheidsinterval = (0,10, 0,32)

Figuur 2. Respondentprofielen en hun statistische eigenschappen voor leeftijdsgroepen voor het antwoorden van 'weet niet' voor de survey 'Politiek en Waarden' (grafiek 1) en voor PVV-stemmers versus GroenLinks-stemmers en hun score op 'Openheid voor Ervaring' (grafiek 2)

politieke partij men op 12 september 2012 heeft gestemd, hebben we een groep GroenLinks-stemmers en een groep PVV-stemmers geconstrueerd. Uit de survey 'Persoonlijkheid' (wave 6, 2013) beschouwen we de 10 items die 'openheid voor ervaring' meten op een 5-punts Likertschaal. We transformeren de respondentscores van 1 tot en met 5 op elk item in respectievelijk 0, 0,25, 0,50, 0,75 en 1. Daarna tellen we deze getransformeerde scores op en ronden het totaal af op een geheel getal voor elke respondent. Deze totalen worden gebruikt als G_i in het linkerdeel van formule (1) met $I_r=10$. De respondentprofielen voor beide politieke partijen worden vervolgens berekend met het rechterdeel van formule (1) en linkerdeel van formule (2). Vergelijking van de twee profielen geeft een $\hat{\delta}_p$ van 0,21, wat een 'klein' maar duidelijk aanwezig effect betekent. We concluderen dat individuen die op GroenLinks stemmen tot op zekere hoogte meer 'openheid voor ervaring' vertonen dan individuen die op de PVV stemmen.

Conclusie

We concluderen dat $\hat{\delta}_p$ een solide en conservatieve statistiek is die zowel nuttig als gunstig is om twee gedragsprofielen mee te vergelijken. In essentie is $\hat{\delta}_p$ een onderschatting van de 'ware' δ en convergeert $\hat{\delta}_p$ naar deze δ naarmate het aantal items toeneemt waarop profielen zijn gebaseerd. De motivering voor dit onderzoek was het gebruiken van een effectmaat voor het vergelijken van twee gedragsprofielen. Met zulke vergelijkingen kunnen relaties tussen respondent-/itemkenmerken en ongewenst antwoordgedrag worden ontdekt. Deze relaties kunnen richting geven aan het ontwerpen van surveys die ongewenst antwoordgedrag minimaliseren en datacollectie optimaliseren.

DANKWOORD

Graag bedanken we CentERdata voor de beschikbaarheid van LISS Panel data en Vanessa Torres van Grinsven voor haar hulp bij publicatie van dit artikel.

LITERATUUR

Bais, F. (2021). *Constructing behaviour profiles for answer behaviour across surveys* (Dissertation). Utrecht University. <https://doi.org/10.33540/538>

- Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114(3), 494-509. <https://doi.org/10.1037/0033-2909.114.3.494>
- Cliff, N. (1996a). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research*, 31(3), 331-350. https://doi.org/10.1207/s15327906mbr3103_4
- Cliff, N. (1996b). *Ordinal methods for behavioral data analysis*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315806730>
- Goedhart, J. (2016). *Calculation of a distribution free estimate of effect size and confidence intervals using VBA/Excel*. BioRxiv. <https://doi.org/10.1101/073999>
- Hess, M. R., & Kromrey, J. D. (2004). *Robust confidence intervals for effect sizes: A comparative study of Cohen's d and Cliff's Delta under non-normality and heterogeneous variances* (Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California). American Educational Research Association.
- Rousselet, G. A., Foxe, J. J., & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience*, 44(9), 2647-2651. <https://doi.org/10.1111/ejn.13400>
- Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, 46(2), 1738-1748. <https://doi.org/10.1111/ejn.13610>
- Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL Common Language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2), 101-132. <https://doi.org/10.2307/1165329>

FRANK BAIS heeft in 2013 een research master Psychology aan de Universiteit van Amsterdam afgerond. De afgelopen jaren deed hij promotie-onderzoek in survey-methodologie aan de Universiteit Utrecht en bij het Centraal Bureau voor de Statistiek. Hij gebruikte gedragsprofielen en de aanpassing van Cliffs Delta om de relatie tussen respondentkenmerken en antwoordgedrag te onderzoeken. Op 31 maart jongstleden is hij gepromoveerd.
E-mail: frank_bais@hotmail.com.

JOOST VAN DER NEUT heeft een BSc en MSc in Applied Geophysics van de Technische Universiteit Delft. Hij promoveerde in 2012 cum laude op seismische beeldvormingstechnieken bij TU Delft en ontving vervolgens een Veni beurs van NWO. Tevens heeft hij de laatste jaren onderzoek gedaan naar medische beeldvormingstechnieken met ultrageluid.
E-mail: jrvanderneut@yahoo.com.

Samenwerken voor duurzaamheid

Bijna tien jaar geleden schreef ik een blog (Poppelaars, 2012) over mijn bezoek aan het Wereldvoedselprogramma (WFP), de grootste humanitaire hulporganisatie gericht op het bestrijden van hongersnood in de wereld. In samenwerking met het programma Moving the World van TNT Express had mijn toenmalige team al verscheidene projecten² (in WFP-termen missies) om de voedseldistributie in onder andere Liberia, Ethiopië en Mali te verbeteren met goed resultaat afgerond (Plat, van Dijk & Poppelaars, 2009). Ik was bij het WFP in Rome om te verkennen hoe we onze samenwerking verder konden uitbreiden.

WFP wint Edelman Award

Mijn bezoek viel samen met de viering van het 50-jarig bestaan van het WFP. Tijdens de algemene vergadering hoorde ik Josette Sheeran, directeur van het WFP, over het werk en de uitdagingen van het WFP spreken. De kille

statistieken die ze in haar speech aanhaalde, gaven aan hoe ernstig de problemen op dat moment waren. Een op de zeven mensen op aarde lijdt chronisch honger, elke 10 seconden sterft er een kind aan de gevolgen van ondervoeding. Gruwelijke feiten, zeker als je beseft dat het onnodig leed is aangezien er voldoende grondstoffen en technologie beschikbaar zijn om iedereen op de wereld te voeden (FAO, z.d.).

Tien jaar later zijn de statistieken wat beter, één op de elf mensen op aarde lijdt chronisch honger. Kindersterfte is echter nog steeds een groot probleem. De gevolgen van klimaatverandering en sinds vorig jaar COVID maken het werk van het WFP niet eenvoudiger. Er is nog een lange weg te gaan om honger uit te bannen maar de bereikte verbetering biedt hoop. Het WFP heeft deze verbetering mede weten te bereiken door de innovatieve inzet van data, statistiek en operations research. Voor deze impactvolle inzet van operations research heeft het WFP in april van dit jaar de INFORMS Franz Edelman Award mogen ontvangen (Informs, 2021). Deze Award heeft ook

