



Universiteit
Leiden
The Netherlands

The construction of health state utilities

Osch, S.M.C. van

Citation

Osch, S. M. C. van. (2007, September 6). *The construction of health state utilities*. Retrieved from <https://hdl.handle.net/1887/12363>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12363>

Note: To cite this publication please use the final published version (if applicable).

Samenvatting

Weersvoorspellingen betreffen de toestand van de atmosfeer op een bepaald tijdstip en op een bepaalde locatie. Voorspellingen volgen een model. Maar geen enkel model kan alle variabelen insluiten die relevant zijn. *Zo kan een weersvoorspelling niet de wind meenemen die wordt veroorzaakt door de vleugels van een vlinder, ook als dat de oorzaak is van een latere orkaan.*

Het is gebruikelijk om gezondheidswinst uit te drukken in termen van kwantiteit en kwaliteit van leven. Een utiliteit is een maat voor de waardering van de kwaliteit van leven. Dit proefschrift behandelt de samenstelling van gezondheidsutiliteiten volgens twee besliskundige modellen, namelijk verwachte nutstheorie (expected utility (EU)) en prospect theorie (PT). EU is een normatieve theorie die voorschrijft hoe rationele beslissingen genomen zouden moeten worden volgens iemands eigen preferenties. Echter er zijn dusdanig sterke en systematische empirische afwijkingen van EU gevonden, dat men voor descriptieve doeleinden is gaan zoeken naar verbeteringen. PT is zo een verbetering. Het is een puur descriptieve theorie en het heeft geen normatieve claims. PT breidt EU uit en omschrijft de fouten en inconsistenties ervan. Dit proefschrift analyseert beslissingen zowel met behulp van het normatieve model, EU, als met behulp van het descriptieve model, PT. Doel is meten van utiliteiten binnen de gezondheidszorg te verbeteren met behulp van descriptieve bevindingen van PT.

Hoofdstuk 1 beschreef de opzet van dit proefschrift. Vooral het onderscheid tussen normatieve, prescriptieve en descriptieve modellen werd behandeld. De drie meest gebruikte methodes om utiliteiten te schatten werden besproken. Dit zijn de standard gamble (SG), de time trade-off (TTO) en de Visueel analoge schaal (VAS).

De SG heeft als enige methode een goede theoretische basis en vindt zijn oorsprong in EU. Voor de SG wordt de respondent gevraagd te beoordelen welke kans op overlijden hij of zij bereid is te accepteren om in optimale gezondheid te verkeren in plaats van een te waarderen gezondheidstoestand (de “zekere uitkomst”). Het probleem is dat mensen niet rationeel beslissen, alhoewel de onderliggende EU theorie dat wel aanneemt. De consequentie is dat de SG utiliteit meetfouten bevat. De fouten (zoals veroorzaakt door kansweging als door verliesangst) en de mogelijke correctiemethodes, worden onder andere omschreven in PT. Het referentiepunt speelt in PT een belangrijke rol. Het is het uitgangspunt waarnaar men het handelen of denken kan richten. Echter PT biedt voor medische toepassingen geen concrete aanwijzingen voor wat respondenten als referentiepunt ervaren.

Voor de TTO wordt de respondent gevraagd te beoordelen hoeveel jaren leven in optimale gezondheid men maximaal wil inleveren om niet in de te waarderen gezondheidstoestand (van minder goede kwaliteit) voor de resterende levensverwachting te leven. De TTO wordt ook verstoord door een aantal meetfouten, onder andere als gevolg van de incorrecte aanname dat we ieder toekomstig jaar evenveel waarderen. In de literatuur wordt gesuggereerd dat de TTO utiliteit de ware utiliteit het dichtst benadert omdat deze het minst last heeft van meetfouten.

De VAS is een lijn van tien centimeter met als uiteinden dood en optimale gezondheid. De respondent wordt gevraagd aan te geven ter hoogte van welk punt t.o.v. de eindpunten de te waarderen gezondheidstoestand zich bevindt wat betreft zijn waarde. De VAS geeft over het algemeen lagere waarderings dan de SG en TTO. De grootste kritiek op de VAS is dat de methode geen theoretische basis heeft.

In hoofdstuk 2 presenteerden we nieuwe inzichten in de correctiemethodes voor de meetfouten. De correctiemethodes worden besproken in de economische literatuur, maar veelal genegeerd in de medische literatuur. Dat is betreurenswaardig, omdat het corrigeren van meetfouten leidt tot betere schattingen van utiliteiten. SG utiliteiten worden beïnvloed door de meetfouten verliesangst, kansweging en schaalcompatibiliteit. TTO utiliteiten worden beïnvloed door de meetfouten verliesangst, utiliteitskromming en schaalcompatibiliteit. Verliesangst wil zeggen dat we gevoeliger zijn voor een bepaald verlies dan voor evenzoveel winst. Kansweging wil zeggen dat mensen kansen wegen, en daarbij overmatig gevoelig zijn voor kansen dichtbij onmogelijkheid (0%) en zekerheid (100%). Utiliteitskromming betreft het afnemende marginale nut van levensjaren. Schaalcompatibiliteit wil zeggen dat wanneer een attribuut van een mogelijke uitkomst van een beslissing (bijvoorbeeld levensduur) veel overeenkomsten vertoont met de respons methode (bijv. zoals het geval in de TTO, waar een antwoord gevraagd wordt in termen van levensjaren), dit attribuut meer gewicht krijgt in een beslissing.

In PT worden uitkomsten beschreven als winst of verlies ten opzichte van een (neutraal) referentiepunt. Iets wordt als verlies of winst gepercipieerd vanuit een neutraal referentiepunt. Als je € 50 hebt gewonnen, dan lijkt dat positief. Als je echter hoort dat je anders € 60 gewonnen had, dan voelt het opeens als negatief. Dit is het gevolg van een verschillend referentiepunt. In het eerste geval ervaart men een uitkomst als winst, en in het tweede geval ervaart men dezelfde uitkomst als een verlies. Het referentiepunt bepaalt hoe SG utiliteiten gecorrigeerd moeten worden voor kansweging en verliesangst. Na correctie zijn dan mogelijk: (a) de voor winst-gecorrigeerde SG utiliteit; (b) de voor verlies-gecorrigeerde SG utiliteit; of (c) de gemengde gecorrigeerde SG utiliteit. De laatste is een half winst, half verlies perceptie van de SG methode en wordt gezien als de theoretische favoriet. Dit wil zeggen dat de

zekere uitkomst, oftewel de te waarderen uitkomst, het referentiepunt was. Veelal leidden de meetfouten tot een te hoge schatting van de SG utiliteit. Het effect van schaalcompatibiliteit op de SG utiliteit was bij de start van dit onderzoek nog onbekend (zie hoofdstuk 4). Er is in de literatuur gesuggereerd dat de meetfouten in de TTO elkaar zouden neutraliseren.

De data lieten zien dat de voor winst-gecorrigeerde SG utiliteit de beste overeenkomst liet zien met de TTO utiliteit. Echter, een onverwachte uitkomst van ons onderzoek was dat de TTO voornamelijk last had van opwaartse meetfouten en dus toch een te hoge schatting gaf van utiliteiten. De meetfout utiliteitskromming had geen invloed op TTO utiliteit. Dit had volgens de literatuur de tegenhanger moeten zijn van de andere twee meetfouten die beide een opwaarts effect hebben op TTO utiliteiten. Daarnaast was de TTO hoger dan de theoretisch favoriete correctie van de SG, de gemengde correctie methode. We toonden verder aan dat zoals al vaker werd gevonden, de ongecorrigeerde SG inderdaad te hoge utiliteiten geeft. Dit kan belangrijke consequenties hebben voor kosteneffectiviteitanalyses. Hoofdstukken 3, 4 en 5 voorzien in een verder inzicht van deze bevindingen met behulp van kwalitatieve gegevens.

In hoofdstuk 3 toonden we met behulp van kwalitatieve data dat respondenten meestal de zekere uitkomst als referentiepunt percipieerden in de SG certainty equivalent (CE) methode. Dat wil zeggen dat de CE wordt gepercipieerd als zijnde deels winst en deels verlies. Dit heeft consequenties voor het gebruik van deze methode omdat de gecorrigeerde CE utiliteit verschilt van de ongecorrigeerde. Verder bleek dat het realiseren van gevormde doelen (zoals voor kinderen zorgen, promotie maken en genieten van het pensioen) voor mensen erg belangrijk is en dat dit eveneens de perceptie van het referentiepunt beïnvloedde. Tevens beïnvloedden

doelen de hoeveelheid aandacht die werd geschonken aan een uitkomst en daarmee de utiliteit. Op basis van deze gegevens argumenteerden we dat motivaties zoals persoonlijke doelstellingen van een patiënt een grotere rol zouden moeten spelen in de behandelkeuze van patiënten.

De kwalitatieve data uit hoofdstuk 4 lieten zien dat de te waarderen gezondheidstoestand in de SG meestal fungeert als referentiepunt. Dat wil zeggen dat de SG wordt gepercipieerd als zijnde gemengd. Persoonlijke doelen worden ook meestal gerelateerd aan deze uitkomst. Voorts vonden we indirect bewijs dat de meetfout schaalcompatibiliteit een opwaarts effect heeft op de SG. Van de andere twee meetfouten was reeds bekend dat ze de SG utiliteit omhoog drukken. Wederom vinden we dus bewijs dat de ongecorrigeerde SG utiliteit substantieel te hoog is.

In het vijfde hoofdstuk verzamelden we kwalitatieve data met betrekking tot de TTO. De data lieten zien dat, zoals verwacht, de gezondheidstoestand die wordt gewaardeerd als referentiepunt fungeerde. Dit leidt tot een hogere utiliteit als gevolg van de meetfout verliesangst. Ook werd meer aandacht geschonken aan lengte van leven dan aan kwaliteit van leven, hetgeen leidt tot een hogere utiliteit als gevolg van de meetfout schaalcompatibiliteit. De kwalitatieve data toonde niet de aanwezigheid van de meetfout utiliteitskromming. We beschreven verder onder andere het mogelijke effect van het geven van een label aan de te waarderen gezondheidstoestand. Het lijkt erop dat het gebruiken van een label, zoals reumatoïde artritis, ertoe leidt dat er een minder groot deel van het utiliteitscontinuüm wordt gebruikt. Hierbij lijkt anticiperende adaptatie een rol te spelen.

In hoofdstuk 6 inventariseerden we mogelijke strategieën die men hanteert bij het invullen van de VAS, om zo de onderliggende cognitieve processen te onderzoeken.

Vervolgens werd systematisch nagegaan in welke mate de strategieën werden gehanteerd. Onze experimenten lieten zien dat er vier mogelijke strategieën zijn, in de volgende voorkeursvolgorde: “Zo ongeveer”, “Halveren van de lijn”, “Numerieke uitdrukking” en “Verdeling in kleinere delen”. Indien men vooraf een extra instructie kreeg waarin de doelstelling van het onderzoek werd beschreven, werd men zich niet bewuster van de gebruikte strategie. Vervolgens bleek dat er waarschijnlijk duale informatieverwerkingsprocessen voorkomen tijdens het waarderen op de VAS. Duale informatieverwerkingsprocessen zijn een interactie tussen automatische en gecontroleerde informatieverwerking. Een bekend voorbeeld is het fietsen naar het werk. Dit gaat normaal bijna vanzelf (automatische informatieverwerking). Er is geen of maar weinig aandacht nodig. Echter, als de weg naar het werk is opgebroken, dan moet er een nieuwe route worden bedacht (gecontroleerde informatieverwerking). Hiervoor is wel aandacht nodig. Gecontroleerde processen zoals het zich bewust zijn van de strategie, verbetert mogelijk de betrouwbaarheid van de VAS, omdat het gebruik van eenzelfde strategie een volgende keer eerder hetzelfde resultaat zal geven. We vonden dat bewustzijn van strategie gerelateerd is aan een meer expliciete strategie (zoals eerst de lijn halveren), maar ook aan vertrouwen in het antwoord. Op basis van deze bevindingen beargumenteerden we dat het aanbeveling zou verdienen om respondenten vooraf te instrueren hun strategie te bepalen, bijvoorbeeld tijdens een oefenronde.

In hoofdstuk 7 introduceerden we een risicovragenlijst. Deze lijst meet gezondheidsgelateerde risicoattitude. Mensen verschillen in hun gezondheidsgelateerde risico attitude. Dit resulteert in verschillend gezondheidsgedrag, en in verschillende keuzes om een medische behandeling te ondergaan. In de medische besliskunde gebruikt men meestal de SG (of CE) om risico attitude te bepalen. Dit is echter een veelomvattende taak, met meetfouten, zoals

hierboven uitgelegd. De gezondheidsgerelateerde risicovragenlijst (HRAS) probeert te bepalen hoe een persoon haar/zijn gezondheid waardeert en omgaat met gezondheidsgerelateerde risico's. Het uiteindelijke doel is om te voorspellen hoe een persoon met riskante gezondheidssituaties zal omgaan. Items voor de schaal werden ontwikkeld met behulp van literatuurstudie en door interviews met patiënten over hun voorkeur voor een riskante medische behandeling. We testten de psychometrische eigenschappen in twee studies. Om de construct validiteit te bepalen gebruikten we algemene risico schalen, een domein specifieke risicoschaal, SG's, een beheersingsoriëntatieschaal, en een persoonlijkheidsvragenlijst. In studie 1 wordt de constructie van de HRAS omschreven en worden de validiteit en betrouwbaarheid bepaald. Op basis van een factor analyse en een betrouwbaarheidsanalyse werd de HRAS gereduceerd van 18 items tot 13 items. In studie 2 beschreven we de validiteit en betrouwbaarheid van de uiteindelijke versie van de HRAS. Relaties met andere risicoschalen waren positief. De HRAS liet een goede betrouwbaarheid zien (zowel intern als test-hertest) en een goede convergente validiteit. Het lijkt erop dat de HRAS een korte en simpele vragenlijst is die gezondheidsgerelateerde risico attitude meet.

De SG moet worden gecorrigeerd voor de opwaartse effecten van de meetfouten voor zover daar een correctiemethode voor bestaat. De zekere uitkomst fungeerde veelal als referentiepunt, dus de SG moet gecorrigeerd worden met de gemengde correctiemethode zoals voorgesteld door Bleichrodt, Pinto en Wakker (4). Er is nog geen correctiemethode beschikbaar voor schaalcompatibiliteit.

De enige neerwaartse fout in de TTO, utiliteitskromming, wordt zowel in de kwantitatieve data als kwalitatieve data niet gevonden. Er werd wel kwalitatief bewijs gevonden voor de andere meetfouten. Daarnaast bleek de gemengde gecorrigeerde SG utiliteit nog altijd lager dan de TTO. De TTO lijkt dus te lijden aan voornamelijk

opwaartse meetfouten. Omdat er nog geen correctiemethode bestaat voor verliesangst en schaalcompatibiliteit, is het enige dat we nu kunnen doen ons bewust zijn van de meetfouten die op de TTO werken.

De twee informatieverwerkingsmechanismen die gebruikt worden bij het waarderen met de VAS lijken ook aanwezig te zijn bij het waarderen met de TTO en SG. Er werd zowel op intuïtieve wijze als op meer beredeneerde wijze ingegaan op de TTO en SG. Het onderscheid tussen intuïtie (meer berustend op emoties) en redeneren is een onderwerp dat veelvuldig in de belangstelling staat. Het blijkt dat emoties en overtuigingen een rol spelen in het nemen van beslissingen en dus ook in het waarderen van kwaliteit van leven. Echter zowel emoties als overtuigingen maken geen onderdeel uit van EU. EU heeft, net als andere traditionele theorieën van economische besliskunde, als aanname dat mensen rationele wezens zijn. Echter mensen zijn op soms karakteristieke wijze irrationele wezens.

Een utiliteit kan niet worden gescheiden van emotie. Het framing effect duidt erop dat mensen zich laten beïnvloeden door hoe informatie wordt aangeboden. Het framing effect is speciaal geassocieerd met neurale activiteit in het emotionele centrum van het brein. Dit benadrukt de bevindingen dat emotionele processen een plaats moeten hebben in besliskundige modellen. In besliskundige analyses wordt meestal aangenomen dat het juiste normatieve model voor beslissingen in onzekerheid de EU is. Soms neemt men dit impliciet aan door niet te corrigeren voor meetfouten. We kunnen concluderen dat dat niet correct is. Een verder vraag is: Kunnen we zeggen dat EU correct is als er zoveel fouten en afwijkingen optreden? Bleichrodt, Pinto en Wakker stellen dat de inconsistenties niet het gevolg zijn van irrationaliteit van de persoon maar van gebrekkigheden van de meetmethode.

Maar wat is er zo erg aan irrationaliteit? In de individuele besliskunde is er mogelijk niets mis mee. Emotie wordt vaak gezien als de antithese van rede. Daarentegen kan emotie vaak worden gezien als een systematische reactie op geobserveerde feiten. Emoties spelen een rol in hoe beslissingen worden genomen. Het is misschien mogelijk deze emoties te formaliseren en deze te verwerken in een beschrijvend besliskundig model. Het lastige is het beantwoorden van de vraag of emoties een plaats hebben in een prescriptief model. Op basis van de literatuur en bevindingen uit dit proefschrift kunnen we stellen dat emoties en doelstellingen een rol spelen in de besliskunde, maar we kunnen niet de bovenstaande vraag uit de morele besliskunde beantwoorden. Een theorie die gevoelens zoals geanticipeerde spijt en verliesangst negeert is geen goede beschrijvende theorie. Daarnaast resulteert het in voorschriften die niet de utiliteit van uitkomsten maximaliseren zoals ze eigenlijk wel worden ervaren. Als emoties een standaard onderdeel zijn van het cognitieve proces, betreffende beslissingen, hoe kunnen we dan voorschrijven om te beslissen zonder hen erbij te betrekken?

Het is niet zo dat wanneer men corrigeert voor meetfouten, dat deze fouten niet mogen bestaan. Maar dat punten die aan de meetfouten ten grondslag liggen niet zijn opgenomen in het model. Aan de andere kant, niet alle elementen die relevant zijn voor besliskunde kunnen in een model worden verwerkt, en er zit een limiet aan het aantal parameters dat we kunnen gebruiken in een model om het hanteerbaar te houden. Voorlopig moeten we werken met wat we hebben omdat beslissingen nu genomen moeten worden. De eerste stap die we maken is een groter bewustzijn van het effect van meetfouten op de utiliteiten binnen de medische besliskunde. De volgende stap is om ervoor te corrigeren. Een derde stap is de rol van emoties verder te formaliseren in besliskundige modellen. Maar wees ervan bewust dat het model *“niet de wind kan meenemen die wordt veroorzaakt door de vleugels van een vlinder.”*

