

BIJLAGE 2

QALY en kwaliteit van leven metingen

Jan van Busschbach, Erasmus MC (projectleider/penvoerder)
Ben van Hout, University of Sheffield
Ardine de Wit, RIVM/UMCU

Inhoud

	Inleiding	5
1	Quality Adjusted Life Year: QALY	6
1.1	Quality Adjusted Life Years QALYs	6
1.2	Utiliteit	7
2	Correctiefactoren van kwaliteit van leven	8
2.1	Noodzakelijke meeteigenschappen	8
2.2	Het maatschappelijk perspectief	8
2.3	Generieke vragenlijsten	9
2.4	De noodzaak van Nederlandse waardeningen	10
3	Standaardanalyse; de EQ-5D-5L en alternatieven náást deze standaardanalyse	12
3.1	De standaardanalyse: de EQ-5D-5L	12
3.2	Alternatieven náást de standaardanalyse	12
3.2.1	Ziektespecifieke vragenlijsten	12
3.2.2	Domeinspecifieke vragenlijsten	13
3.2.3	Kwalitatief onderzoek	14
3.2.4	Preferentieonderzoek	14
3.2.5	Recente initiatieven: PROMs en Positive health	14
3.2.6	Andere QALY-vragenlijsten dan de standaardanalyse	15
3.2.7	Mapping	17
	Rapportage	19
	Referenties	20

Inleiding

In deze verdiepingmodule worden de methodologische keuzes bij het meten van kwaliteit van leven binnen de 'Richtlijn voor Uitvoeren van Economische Evaluaties in de Gezondheidszorg' toegelicht. Daarbij zal vooral de keuze voor de EQ-5D-5L worden gemotiveerd. Daarnaast zullen alternatieve meetmethoden besproken worden die náást dit meetinstrument gebruikt mogen worden. De opbouw van de toelichting volgt de opbouw van de richtlijn:

- QALYs; waarom wordt gekozen voor QALY en wat zijn de kenmerken van QALYs?
- Correctiefactoren van kwaliteit van leven; aan welke eisen moeten de meetinstrumenten voor kwaliteit van leven voldoen?
- Standaardanalyse; waarom de EQ-5D-5L en welke alternatieven kunnen náást deze standaardanalyse zinvol gebruikt worden?
- Rapportage; aan welke eisen moet de rapportage voldoen?

1 Quality Adjusted Life Year: QALY

Het deel van de richtlijn waarin het meten van kwaliteit van leven wordt besproken, begint met een beschrijving van QALYs. Die tekst staat hieronder afgebeeld.

1.1 Quality Adjusted Life Years (QALYs)

Interventies kunnen een effect hebben op de levensduur en/of de kwaliteit van leven. In gezondheids-economische evaluaties zijn beide belangrijk en kunnen beide gebruikt worden om de kosten te rechtvaardigen. De relevantie van de effecten op levensduur en kwaliteit van leven ten opzichte van de kosten dient te worden aangetoond door middel van een 'kostenutiliteitsanalyse' (KUA). De 'utiliteiten' van de KUA moeten worden geoperationaliseerd door 'kwaliteit van leven gecorrigeerde levensjaren', ofwel 'Quality Adjusted Life Years': QALYs.

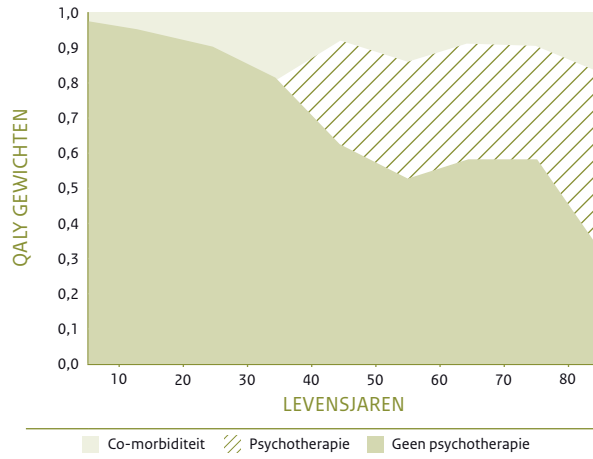
De richtlijn gaat er dus vanuit dat interventies een effect hebben op de levensduur, op de kwaliteit van leven of op allebei. In gezondheidseconomische evaluaties worden zowel levensduur als kwaliteit van leven als belangrijk ervaren. Verbeteringen in zowel levensduur als kwaliteit van leven kunnen daarom gebruikt worden om de kosten te rechtvaardigen. De effecten op overleving en kwaliteit van leven kunnen worden gezien als het 'nut' of de 'utiliteit', waarvan de waarden moeten opwegen tegen de kosten. Er is voor gekozen om deze afweging tussen de kosten en waarden van de effecten standaard te laten plaats vinden middels een kostenutiliteitsanalyse (KUA). Met 'utiliteit' of 'nut' wordt hier bedoeld de waarden van de effecten op overleving en kwaliteit van leven. Om 'utiliteit' of 'nut' te operationaliseren is gekozen voor de uitkomstmaat 'voor kwaliteit van leven gecorrigeerde levensjaren' ofwel 'Quality Adjusted Life Years', afgekort 'QALYs'.

Schematisch worden QALY meestal gepresenteerd met de levensjaren op de x-as en de kwaliteit van leven op de y-as, zoals in figuur 1. In figuur 1 wordt een denkbeeldige ontwikkeling van een persoonlijkheidsstoornis uitgebeeld door het onderste donkere vlak (geen psychotherapie). Als baby heeft de persoon een kwaliteit van leven vlak bij het optimum van 1,00. Maar tijdens de puberteit ontwikkelt zich een persoonlijkheidsstoornis en verliest die persoon kwaliteit van leven. Dit afglijden van de kwaliteit van leven bereikt een voorlopig dieptepunt rond zijn vijftigste. Na z'n 70e wordt het nog erger, want dan ontwikkelt de persoon ook allerhande lichamelijke klachten en op z'n 80e sterft de persoon. Stel dat zo iemand een bewezen effectieve psychotherapie had gekregen rond zijn 25e levensjaar. Als de therapie inderdaad werkzaam zou zijn, dan verbetert de kwaliteit van leven. Het gestreepte oppervlakte vertegenwoordigt dan de winst in termen van QALYs door de therapie. Het grijze vlak wat overblijft, vertegenwoordigt de restziekten.

De grafiek is geïllustreerd door het ziekteverloop van een patiënt met persoonlijkheidsstoornissen, maar eenzelfde grafiek had getekend kunnen worden voor patiënten met andere progressieve ziekten die de kwaliteit van het leven aantasten. Bijvoorbeeld; de patiënt gaat achteruit in gezondheid en komt er op 25-jarige leeftijd achter dat hij diabetes is. Doet de patiënt daar weinig aan, dan wordt zijn gezondheid slechter: het donkere vlak. Veranderd de patiënt na het ontdekken van de ziekte zijn levensstijl grondig, dan 'wint' de patiënt het gearceerde vlak.

Met QALYs kan je dus het verloop en de winst van behandeling van heel verschillende ziekten en behandelingen met elkaar vergelijken. Door het gebruik van deze universele uitkomstmaat kunnen kosten en effecten van verschillende typen interventies gericht op verschillende ziekten toch met elkaar vergeleken worden.¹

Figuur 1. Een denkbeeldige ontwikkeling van leven van een patiënt met een persoonlijkheidsstoornis met en zonder succesvolle behandeling



1.2 Utiliteit

In het bovenstaande is betoogd dat de woorden ‘nut’ en ‘utiliteit’ verwijzen naar de waarde (value) van de effecten. QALYs zijn samengesteld uit levensjaren en kwaliteit van leven en is de meetbare eenheid van nut of utiliteit. Verwarrend is dat een ‘kwaliteit van leven score’ soms ook een ‘utiliteit’ wordt genoemd. Dat is eigenlijk niet correct, omdat iets alleen maar ‘nut’ kan hebben als het een bepaalde tijd duurt; het is immers kwaliteit van leven. Daarom wordt in deze verdiepingsmodule de term ‘utiliteit’ gereserveerd als synoniem voor ‘QALY’. Een kostenutiliteitsanalyse is, binnen de context van de richtlijn, dus een ‘kosten per QALY analyse’. Een betere term voor de kwaliteit van leven score is ‘kwaliteit van leven gewicht’, of ‘correctiefactor voor kwaliteit van leven’. Dat laatste past ook bij de Nederlandse vertaling van QALY: kwaliteit gecorrigeerde levensjaren.

In de economie worden de woorden ‘nut’ of ‘utiliteit’ breder gebruikt dan alleen bij het meten van QALYs. In die wetenschap worden ‘nut’ of ‘utiliteit’ ook gebruikt om in het algemeen de waarde aan te geven voor goederen of diensten (commodities). Bij een perfecte marktwerking kan de waarde van die goederen en diensten bepaald worden door de prijs die er voor betaald wordt:

Utility is a metaphysical concept of impregnable circularity; utility is the quality in commodities that makes individuals want to buy them and the fact that individuals want to buy them shows that they have utility.²

In de gezondheidszorg is er, met opzet, geen sprake van marktwerking, waardoor de prijs van een behandeling niet gezien kan worden als de waarde (het nut, de utiliteit) van die behandeling. We moeten daarom op zoek naar een alternatieve manier om waarde uit te drukken. We doen dat door de kwaliteit van leven van gezondheidstoestanden te relateren aan levensduur; de QALY. Daarmee hebben we een relatieve maat waarbij overleving en kwaliteit van leven op één noemer staan, en operationaliseert de QALY ‘nut’ of ‘utiliteit’.

2 Correctiefactoren van kwaliteit van leven

Het deel van de richtlijn waarin het meten van kwaliteit van leven wordt besproken begint met een beschrijving van QALYs. Die beschrijving is hoofdstuk 1 besproken. Daarna gaat de richtlijn verder over de 'correctiefactoren van kwaliteit van leven', ofwel de kwaliteit van leven gewichten. De noodzakelijke eigenschappen van deze correctiefactoren worden in de richtlijn als volgt beschreven:

Correctiefactoren van kwaliteit van leven

De kwaliteit van leven in het kader van het meten van QALYs, moet worden vastgesteld middels gevalideerde generieke kwaliteit van leven vragenlijsten die door patiënten worden ingevuld. De uitkomstmaten van die vragenlijsten zijn 'gezondheidstoestanden' die gedefinieerd zijn in termen van gezondheidsgerelateerde kwaliteit van leven. De gevalideerde vragenlijsten koppelen een 'kwaliteit van leven correctie' voor een QALY-analyse aan de gezondheidstoestanden. Deze correctiefactor wordt een 'weging', 'waardering' of 'utiliteit' genoemd. Deze 'waarderingen voor de kwaliteit van leven' moet bestaan uit een weging van de 'kwaliteit van leven' ten opzichte van 'tijd van leven'. De waarderingen voor de gezondheidstoestanden van de patiënten moeten gebaseerd zijn op de preferenties van de algemene bevolking, zodat de analyses uitgevoerd kunnen worden vanuit het maatschappelijk perspectief.

2.1 Noodzakelijke meeteigenschappen

Een bijzondere eigenschap van de gewichten van kwaliteit van leven is dat zij op een schaal moeten passen, waarbij 0,00 de waarde van dood vertegenwoordigt en 1,00 de waarde van volledige gezondheid. Daarnaast moet deze schaal, om in QALY berekeningen te kunnen gebruiken, ratio-eigenschappen hebben: 0,25 moet de helft zijn van de waarde 0,50. Er zijn maar een paar manieren bekend waarop een vragenlijst 'gevalideerd' kan worden zodat het aannemelijk is dat de schaal die 'ratio-eigenschappen ook daadwerkelijk heeft. Op dit moment is 'time trade-off' de meest gebruikte methodologie, naast 'standard gamble'. Er zijn initiatieven gaande om te kijken of ook andere methoden geschikt zouden zijn, waarbij 'discrete choice modelling' (DCE) en aanverwante conjoint analysis technieken de beste kaarten lijken te hebben. Bij deze alternatieve nieuwe methoden blijkt het vooral lastig te zijn om de waarde van 'tijd' (levensduur) mee te wegen, zoals dat gebeurt in time trade-off en standard gamble. Als tijd wel in DCE wordt meegewogen, zoals onder andere is voorgesteld door Bansback et al. 2014³, is het nog onduidelijk hoe de wegingsfactoren geschaald kunnen worden op een schaal waarbij 0,00 de waarde van dood is en 1,00 de waarde van perfect gezond.³ Zie voor een discussie hierover, Rowen et al, 2015.⁴ Er zijn erg veel vragenlijsten beschikbaar over kwaliteit van leven. Veel kwaliteit van leven vragenlijsten kennen wel de eerste stap van het beschrijven van de gezondheidstoestand, maar niet de tweede stap, die van de weging van de gezondheidstoestand. Soms is er wel een somscore, of wordt de score van de patiënt gerelateerd aan een leeftijdsafhankelijke populatienorm. Dit zijn echter geen wegingen ten opzichte van tijd die aan de bovengenoemde eigenschappen voldoen. Daarom kunnen veel kwaliteit van leven vragenlijsten geen kwaliteit van leven gewichten genereren die gebruikt kunnen worden in QALYs. Soms worden dergelijke vragenlijsten toch toegevoegd aan het onderzoek om in kaart te brengen in welk domein van kwaliteit van leven de veranderingen zich voordoen. De informatie uit deze vragenlijst kan waardevol zijn, maar kan dus niet als vervanging dienen van vragenlijsten die wel gewichten van kwaliteit van leven kunnen genereren.

2.2 Het maatschappelijk perspectief

Het wegen van de uitkomst van de vragenlijst middels time trade-off of standard gamble kan geschieden door de patiënt zelf, door iemand uit de omgeving van de patiënt of door het algemeen publiek. De richtlijn schrijft voor dat:

De waarderingen voor de gezondheidstoestanden van de patiënten moeten gebaseerd zijn op de preferenties van de algemene bevolking, zodat de analyses uitgevoerd kunnen worden vanuit het maatschappelijk perspectief.⁵

Dat betekent dat de patiënten zelf de gezondheidstoestanden bepalen door het invullen van vragenlijsten (de eerste stap – zie hierboven), maar dat de waarderingen voor uitkomsten van die vragenlijsten worden vastgesteld door het algemeen publiek (de tweede stap).

Deze laatste stap wordt vaak raar gevonden: zouden het niet de patiënten moeten zijn die de waarderingen vaststellen? Patiënten zijn toch immers de consumenten van de zorg? Dit is evenwel niet zo vanzelfsprekend als het lijkt.

In de eerste plaats is het niet vanzelfsprekend dat patiënten de ‘consumenten’ zijn van de gezondheidszorg. Consumenten hebben een belangrijke positie bij het bepalen van de waarde van een product of dienst, mede omdat ze voor die dienst betalen. Maar doorgaans betalen patiënten niet alleen voor hun zorg. Het is het algemeen publiek dat betaalt via de bijdrage voor de ziektekostenverzekering. Daarnaast wordt zorg betaald uit bijdragen van werknemers en werkgevers en uit de rijksbegroting. De positie van de patiënt in het zorgstelsel is dan ook veel complexer dan de positie van een consument op de markt voor consumentengoederen. De relatie tussen de patiënt en het zorgstelsel moet dan ook gezien worden vanuit het verzekeringsperspectief. Bij een verzekering is het niet degene die de schade oploopt die op dat moment bepaalt hoeveel er uitgekeerd moet worden: dat gebeurt door de verzekeringsnemer vóórdat de schade is opgelopen. De verzekeringsnemer doet dat door vóór een polis met een specifieke dekking te kiezen, zonder dat hij/zij weet of hij deze verzekering ooit nodig zal hebben. Voordat de schade is opgelopen, is de patiënt een onderdeel van het algemeen publiek. Het is dus aan het algemeen publiek om vast te stellen wat zij belangrijk vinden.⁶ Voor het gebruik van QALY in economische evaluaties betekent dit dat de waarde van de kwaliteit van leven moet worden bepaald door het algemeen publiek, ook weer zonder dat zij weten of ze zelf ooit in deze gezondheidstoestand terecht zullen komen. Er is nog een tweede reden om gebruik te maken van waardering van het algemeen publiek. Het is in de praktijk namelijk moeilijker dan gedacht om invoelbare, lees valide, waarderingen te krijgen van de eigen kwaliteit van leven van patiënten. Vaak zijn de waarderingen voor duidelijk slechte gezondheidstoestand onwaarschijnlijk hoog. Dat komt omdat de patiënt zich aanpast aan zijn gezondheidstoestand. Dat wordt ook wel ‘coping’ genoemd. Bijvoorbeeld, wanneer iemand in een rolstoel belandt, zal deze patiënt in eerste instantie een lage kwaliteit van leven rapporteren. Maar na verloop van tijd zal de vraag “hoe gaat het met je?” kunnen worden beantwoord met “goed”. Dat patiënten zich aan hun gezondheidstoestand aanpassen, is een gezonde, en zelfs noodzakelijke voorwaarde om zo goed mogelijk te kunnen functioneren. Maar in geval van het vaststellen van QALYs doen patiënten zichzelf tekort: het verbeterpotentieel voor kwaliteit van leven wordt door de aanpassing verkleind. Een klein verbeterpotentieel heeft een ongunstig effect op de kosteneffectiviteit. Met andere woorden: het is dus in het belang van de patiënt dat anderen dan de patiënt zelf de kwaliteit van leven waarderen.

Het algemeen publiek moet uiteraard wel weten wat er speelt bij de patiënt om een goede waardering van de kwaliteit van leven te geven. Daarom vullen de patiënten zelf de vragenlijsten in die de kwaliteit van leven van de patiënten beschrijft in generieke termen. Die beschrijvingen worden gewaardeerd door het algemeen publiek. Daarmee wordt aan de volgende zaken voldaan:

- De patiënt geeft zelf een beschrijving van zijn of haar kwaliteit van leven;
- Door het gebruik van generieke vragenlijsten kan het algemeen publiek de beschrijving begrijpen;
- Door het gebruik van generieke vragenlijsten kan het algemeen publiek de beschrijving vergelijken met andere gezondheidstoestanden en deze waarderen ten opzichte van elkaar.

2.3 Generieke vragenlijsten

De richtlijn schrijft voor dat ‘generieke vragenlijsten’ van kwaliteit van leven gebruikt moeten worden. Een generieke vragenlijst meet gezondheidsgerelateerde kwaliteit van leven over een breed spectrum. Een ziektespecifieke vragenlijst focust op een beperkt aantal voor de ziekte relevant geachte domeinen van kwaliteit van leven. Een generieke vragenlijst meet kwaliteit van leven dus breed, terwijl een ziektespecifieke vragenlijst de diepte op zoekt. Dat de richtlijn vraagt om de meting van kwaliteit van leven over een breed spectrum te benaderen, komt onder andere voort uit het ‘maatschappelijk perspectief’ van waaruit de richtlijn is opgesteld. Dat maatschappelijk perspectief bepaalt dat alle effecten moeten worden meegewogen in de analyses. Dat betekent dat niet alleen naar de gewenste effecten van een interventie moet worden gekeken, maar ook naar de bijwerkingen. Het voorgeschreven maatschappelijk

perspectief pleit er dan ook voor om gebruik te maken van een ‘generiek meetinstrument’. Generiek wil in dit geval dus zeggen dat een breed scala van kwaliteit van leven effecten moeten worden gemeten. Dat betekent dat de vragenlijst die gebruikt wordt in QALY analyses zich niet alleen moet richten op het beoogde hoofdeffect van de interventie, maar ook de (onverwachte) bijwerkingen en de eventueel aanwezige comorbiditeit.

Een tweede reden om gebruik te maken van generieke vragenlijsten is dat het met generieke vragenlijsten gemakkelijker is om tot een standaard lijst te komen. Een ziektespecifieke vragenlijst kan immers niet gemakkelijk worden ingezet om interventies bij verschillende ziektebeelden met elkaar te vergelijken. Een voordeel van generieke vragenlijsten is dat ze breed kunnen worden ingezet: ze meten een breed scala aan kwaliteit van leven dimensies. Wanneer een generieke vragenlijst eenmaal is gevalideerd is deze in principe geschikt voor alle interventies in alle ziektegebieden. Het valideren betekent hier vooral de mogelijkheid om waarderungen aan de gezondheidstoestanden te koppelen. Gevalideerde generieke vragenlijsten vereenvoudigen het werk van de onderzoeker dan ook aanzienlijk. Toch worden vaak naast generieke- ook ziektespecifieke- en domeinspecifieke vragenlijsten, gericht op specifieke gezondheidsproblemen zoals angst of pijn, ingezet bij gezondheidseconomische evaluaties. Ook worden kwalitatieve onderzoeksmethoden gebruikt. Het is namelijk goed mogelijk dat de generieke vragenlijst niet alle voor de patiënt relevante zaken goed oppakt. Door het inzetten van ziektespecifieke vragenlijsten naast de generieke vragenlijsten kan worden onderzocht of geen ‘waarde’ van de interventie wordt gemist door de generieke vragenlijst. In hoofdstuk 3 van deze verdiepingsmodule, die gaat over de standaardanalyse wordt ingegaan op de mogelijkheden om deze extra ‘waarde’ van de interventie in kaart te brengen.

2.4 De noodzaak van Nederlandse waarderungen

Economische evaluaties in de Nederlandse context moeten gebruik maken van Nederlandse populatiestudies om waarden aan gezondheidstoestanden toe te kennen. Het is denkbaar dat de uitkomsten van de vragenlijsten van kwaliteit van leven worden gewogen door het algemeen publiek uit een ander land. Het is een oplossing die vaak gekozen wordt wanneer Nederlandse waarderungen voor de betreffende vragenlijst ontbreken. Daardoor worden de uitkomsten veel onzekerder, en wel om tenminste twee redenen. De meest voor de hand liggende reden voor extra onzekerheid is dat het niet duidelijk is of Nederlanders dezelfde gewichten toekennen aan de gezondheidstoestanden als niet-Nederlanders. Tot nu toe zijn er steeds verschillen gevonden tussen landen, met name bij het gebruik van time trade-off, en het is nog niet duidelijk waar die verschillen door veroorzaakt worden. Het is dus niet goed te voorspellen of waarderungen in twee landen gelijk zullen zijn of niet.

Verder neemt de onzekerheid toe wanneer de antwoorden van patiënten uit land A worden gecombineerd met de waardering van publiek uit land B. Het is denkbaar dat de items van de vragenlijsten in land A misschien niet dezelfde connotatie hebben als de items zoals die beschreven zijn in dezelfde vragenlijst in land B. Stel bijvoorbeeld dat een van de items van de vragenlijst zou zijn “enige problemen met lopen”. Dit is in het Engels verwoord als “some problems with walking about”. Wanneer we de Nederlandse waarderungen willen koppelen aan de Engelse items, moeten we eerst zeker weten of “lopen” precies dezelfde betekenis heeft als “walking about”. Dat laatst kan wellicht worden geïnterpreteerd als een breder begrip dan “lopen” en dus kunnen “some problems with walking about” problematischer zijn dan “enige problemen met lopen”. De waarderungen tussen Engeland en Nederland zullen dan ook verschillen, niet omdat de preferenties anders zijn, maar omdat de vragenlijsten linguïstisch niet precies het zelfde zijn. Met name bijvoeglijke naamwoorden zoals “een beetje, veel en extreem” kunnen in verschillende talen verschillende betekenissen hebben. Het is dus goed denkbaar dat deze verschillen tussen vragenlijsten er zijn. Dat betekent dat waarderungen uit het ene taalgebied niet zomaar gebruikt kunnen worden in een ander taalgebied zonder een toename van onzekerheid.

Gegeven de toegenomen onzekerheden verdient het aanbeveling om bij Nederlandse economische evaluaties gebruik te maken van Nederlandse patiënten en Nederlandse waarderungen. Wanneer alleen buitenlandse patiënten zijn geïnccludeerd, bijvoorbeeld in internationale geneesmiddelen trials, zijn er verschillende mogelijkheden. Wanneer het onderzoek is gemodelleerd in bijvoorbeeld een Markov-mo-

del, dan kunnen de Nederlandse waarderingen van de gezondheidstoestanden worden gevonden door vragenlijsten af te nemen bij Nederlandse patiënten die zich in vergelijkbare toestanden bevinden. Dit is vergelijkbaar met het kostenonderzoek, waarbij ook de kosten van een internationaal vastgestelde gezondheidstoestand voor de Nederlandse situatie wordt bepaald. De patiënten hoeven zich niet noodzakelijker wijze in een vergelijkbare RCT te bevinden: net als bij het kostenonderzoek moet wel worden aangetoond dat de gezondheidstoestanden vergelijkbaar zijn met de gezondheidstoestanden in het model. Dit biedt de mogelijkheid om dergelijke kosten- en kwaliteit van leven onderzoek uit te voeren in de dagelijkse praktijk. Een tweede, minder goede oplossing is om Nederlandse waarderingen te gebruiken voor in het buitenland gemeten gezondheidstoestanden. Daarbij wordt dus geaccepteerd dat er een foute koppeling kan plaatsvinden tussen de beschrijving van de gezondheidstoestand en de daaraan gekoppelde waardering van die toestand. Een derde oplossing is het accepteren van buitenlandse waarderingen. Wat in ieder geval geen goede oplossing is, is om de antwoorden van Nederlandse patiënten op Nederlandse vragenlijsten te wegen met buitenlandse wegingen.

3 Standaardanalyse; de EQ-5D-5L en alternatieven naast deze standaardanalyse

In de richtlijn wordt een standaard voorgeschreven voor het meten van de kwaliteit van leven: de EQ-5D-5L. De ‘standaard’ heet binnen de richtlijnen ‘de standaardanalyse’.

Standaardanalyse

Vanwege de mogelijke verschillen in de waardering van kwaliteit van leven die ontstaan bij het gebruik van verschillende vragenlijsten, dient de kwaliteit van leven altijd gemeten te worden met de EQ-5D-5L. Dit kan worden gezien als een vereenvoudiging van de werkelijkheid en is daardoor niet in alle gevallen afdoende. Alternatieve vragenlijsten en andere methoden om de kwaliteit van leven in kaart te brengen mogen daarom naast de standaardanalyse worden aangeboden. Deze alternatieven vervangen de standaardanalyse niet, maar kunnen noodzakelijk zijn om het volledige effect in kaart te brengen. Dit houdt dus in dat ook wanneer verwacht wordt dat de EQ-5D-5L niet sensitief is, de EQ-5D-5L toch moet worden afgenomen. De alternatieven worden in de verdiepingmodule besproken.

3.1 De standaardanalyse: de EQ-5D-5L

Hierboven is al beschreven dat de voorkeur uitgaat naar een gevalideerde generieke vragenlijst van kwaliteit van leven die door middel van time trade-off of standard gamble waarderingen koppelt aan gezondheidstoestanden. Die waarderingen moeten worden gegeven door een representatieve steekproef van de algemene bevolking van Nederland. Er zijn op dit moment twee vragenlijsten in Nederland die aan die eisen voldoen: EQ-5D-3L en de EQ-5D-5L. De laatste is een doorontwikkelde versie van de eerdere EQ-5D-3L, en kent 5 antwoordniveaus per vraag in plaats van 3 zoals in de ‘oude’ EQ-5D-3L. Deze doorontwikkeling heeft plaats gevonden om de sensitiviteit van het instrument te verbeteren. Daarnaast is de waardering door middel van time trade-off verbeterd. De vragenlijst wordt ontwikkeld door een internationale, niet commerciële groep van onderzoekers, met daarin een sterk Nederlands aandeel. EQ-5D-3L is wereldwijd de meest gebruikte vragenlijst bij economische evaluaties en is methodologisch gezien de meest onderzochte vragenlijst. Het gebruik van de vragenlijst is voor non-commercieel gezondheidseconomisch onderzoek gratis na registratie op www.euroqol.org.⁷ Van ander internationaal gebruikte generieke QALY vragenlijsten zoals de SF-6D, HUI 3, 15D, QWB en de AQoL-8D bestaan op dit moment geen gepubliceerde waarderingen van het Nederlandse publiek. Dit alles maakt dat de richtlijn de EQ-5D-5L noemt als standaardanalyse voor het meten van de gezondheidgerelateerde kwaliteit van leven.

3.2 Alternatieven naast de standaardanalyse

Gegeven de mogelijke verschillen in de waarderingen van kwaliteit van leven die ontstaan bij het gebruik van verschillende vragenlijsten stelt de standaardanalyse dat de kwaliteit van leven meting moet worden uitgevoerd met de EQ-5D-5L. Zoals hierboven ook al besproken is kan het zinnig zijn om náást deze standaardanalyse ook resultaten van ander kwaliteit van leven onderzoek te presenteren. Het is bijvoorbeeld denkbaar dat de EQ-5D niet alle relevante kwaliteit van leven aspecten oppakt voor een bepaalde ziekte of een bepaalde interventie. In dat geval kan men er bijvoorbeeld voor kiezen om ook ziektespecifieke of domeinspecifieke meetinstrumenten in te zetten, die wellicht meer sensitief zijn om kleine veranderingen op te pikken. De richtlijn laat nadrukkelijk deze optie open, zolang de standaardanalyse ook wordt gepresenteerd.

3.2.1 Ziektespecifieke vragenlijsten

Een ziektespecifieke vragenlijst beperkt zich tot de kwaliteit van leven effecten die voorkomen bij één bepaalde ziekte, bijvoorbeeld psoriasis of prostaatklachten. Daarnaast zijn er vragenlijsten die zich richten op een groep van ziekten, bijvoorbeeld oncologische ziekten. Ziektespecifieke vragenlijsten kunnen gedetailleerd de effecten van een interventie in kaart brengen en worden vaak samengesteld in overleg

met de betrokken patiëntengroep. Een nadeel is dat de specifieke doelgroep ook het onderzoek beperkt: er is meestal minder onderzoek mee gedaan dan met generieke lijsten. Daardoor zijn ziektespecifieke vragenlijsten vaak minder goed gevalideerd dan generieke vragenlijsten. Het is dus van belang om goed in de gaten te houden of de validiteit en betrouwbaarheid van de ziektespecifieke vragenlijst in voldoende mate is onderzocht. Een niet gevalideerde vragenlijst voegt weinig toe aan overtuigingskracht. Het is dan ook niet aan te bevelen om ad hoc dergelijk vragenlijsten samen te stellen.

De meeste ziektespecifieke vragenlijsten kunnen niet gebruikt worden in een QALY analyse, vanwege het ontbreken van waarderingen voor gezondheidstoestanden afkomstig uit de algemene populatie. Een beperkt aantal ziektespecifieke vragenlijsten vormt hierop een uitzondering: de waarderingen (de gewichten) van kwaliteit van leven zijn hier wel bekend. Voor deze vragenlijsten is dan op vergelijkbare wijze als voor de EQ-5D gebeurd is een waardenset afgeleid uit waarderingsonderzoek in de algemene bevolking. Deze vragenlijsten zouden dus in principe ook gebruikt kunnen worden in QALY-analyses. Deze ziektespecifieke vragenlijsten worden ook vaak gezien als een sensitieve vervanger van de generieke vragenlijsten zoals de EQ-5D. Toch kunnen deze ziektespecifieke vragenlijsten de generieke vragenlijsten niet vervangen. Het belangrijkste probleem is dat, uit de aard van de zaak, de ziektespecifieke vragenlijsten niet alle facetten van gezondheidsgerelateerde kwaliteit van leven meten. Daardoor wordt bij patiënten met comorbiditeit welke buiten het meetdomein ligt van de ziektespecifieke vragenlijsten, de waarde van de kwaliteit van leven overschat. Daarnaast kunnen de bijeffecten en complicaties van de behandeling worden gemist. Dat betekent dat niet alleen de absolute waarde van kwaliteit van leven, maar ook de effectgrootte wordt beïnvloed door de beperkte reikwijdte van de ziektespecifieke vragenlijst. Het valideren van vragenlijsten voor het gebruik in een QALY analyse is complex en duur. Het verkrijgen van goede time trade-off en standard gamble waarderingen is arbeidsintensief, omdat ze moeten worden vastgesteld in interviews door getrainde interviewers waarbij gebruikt wordt gemaakt van gedetailleerde protocollen. Omdat generieke vragenlijsten meer gebruikt worden, wordt er ook meer onderzoek gedaan naar de gewichten van generieke vragenlijsten dan bij ziektespecifieke vragenlijsten. Daardoor is de kwaliteit van de waarderingen van generieke vragenlijsten doorgaans beter dan die van ziektespecifieke vragenlijsten.

Een bijkomend probleem is dat het nog onduidelijk is onder welke condities de uitkomsten van ziektespecifieke vragenlijsten dezelfde waarderingen opleveren als die van generieke vragenlijsten. Een probleem van ziektespecifieke vragenlijsten is bijvoorbeeld de sterke focus op slechts één of enkele aspecten van gezondheid. Door die focus kunnen de waarderingen voor die aspecten worden overschat (framing). Tot op heden is het nog onbekend onder welke omstandigheden de waarderingen van generieke en ziektespecifieke meetinstrumenten vergelijkbaar zijn.

Samengevat kan gesteld worden dat wanneer vermoed wordt dat er meer effecten zijn dan kunnen worden waargenomen met de EQ-5D-5L het de moeite waard is om te kijken of die 'extra waarde' zichtbaar gemaakt kan worden middels ziektespecifieke vragenlijsten. In alle gevallen zullen de uitkomsten van de ziektespecifieke vragenlijsten gebruikt worden náást de QALY analyse die is uitgevoerd om te voldoen aan de standaardanalyse, dus met behulp van de EQ-5D-5L. Het gebruik van ziektespecifieke vragenlijsten is dus geen vervanging voor de EQ-5D-5L, maar een verdieping van het kwaliteit van leven onderzoek zoals voorgeschreven door de richtlijn.

3.2.2 Domeinspecifieke vragenlijsten

Een domeinspecifieke vragenlijst beperkt zich tot één domein van kwaliteit van leven, bijvoorbeeld jeuk of depressie. Voorbeelden zijn de Skindex-29 en de State-Trait Anxiety Inventory (STAI) en de Hospital Anxiety and Depression Scale (HADS). Met name in de geestelijke gezondheidszorg zijn er veel goed gevalideerde domeinspecifieke vragenlijsten voorhanden, zoals vragenlijsten naar depressie en mentale gezondheidsklachten. Net als bij ziektespecifieke vragenlijsten geldt ook voor domeinspecifieke vragenlijsten dat er maar weinig vragenlijsten zijn waarbij er gewichten voor de kwaliteit van leven bekend zijn die gebruikt kunnen worden in QALY-analyses. Daarnaast geldt, net als bij ziektespecifieke vragenlijsten, ook hier de beperkingen dat het specifieke karakter van deze vragenlijsten de lijsten minder geschikt maken voor economische evaluaties. Het gebruik van domeinspecifieke vragenlijsten is dus geen vervan-

ging voor de EQ-5D-5L, maar een verdieping van het kwaliteit van leven onderzoek zoals voorgeschreven door de richtlijn.

3.2.3 *Kwalitatief onderzoek*

Onderzoek naar kwaliteit van leven hoeft zich niet te beperken tot vragenlijstonderzoek. Het is bijvoorbeeld ook mogelijk om de effecten van een interventie te beschrijven via kwalitatief onderzoek. Kwalitatief onderzoek heeft de naam subjectief te zijn en daarmee ongeschikt om besluitvorming te ondersteunen. Inderdaad zijn er meer subjectieve elementen bij kwalitatief onderzoek, maar wanneer kwalitatief onderzoek state-of-the-art wordt toegepast kan waardevolle informatie boven tafel komen. Te denken valt aan onderzoek naar de manier waarop patiënten de ziekte en de interventie beleven en naar de belasting van de verzorgers. Kwalitatief onderzoek geeft de mogelijkheid om subtiele effecten uit te vragen en wordt daarom wel gezien als een bijzonder sensitive methode om veranderingen op te pikken. Het lijkt daarmee een instrument waarmee de gunstige effecten van een interventie gemakkelijk naar voren kunnen worden gebracht. Goed kwalitatief onderzoek richt zich evenwel niet uitsluitend op de voordelen van een interventie, maar zal ook proberen de nadelen in kaart te brengen. Daarnaast mag van goed kwalitatief onderzoek verwacht worden dat het de effecten van een interventie duidt in een groter kader. Daarmee is dus niet gezegd dat met kwalitatief onderzoek er altijd een positief effect gevonden zal worden. Immers met goed kwalitatief onderzoek zal het ook mogelijk zijn dat de bijwerkingen van de interventie naar voren komen, dat door coping de beleving van de ernst van de ziekte beperkt is, en zelfs dat de patiënten het ziektebeeld niet eens als ziekte herkennen. Een voorbeeld is een idiopathisch kleine gestalte. Kwalitatief onderzoek bracht naar voren dat het leeuwendeel van de kinderen en volwassenen met een idiopathisch kleine gestalte dit niet als een probleem ervaren, sterker nog: ze benoemen juist de voordelen.⁸

3.2.4 *Preferentieonderzoek*

In het onderzoek naar kwaliteit van leven wordt de laatste tijd veel geëxperimenteerd met zogenaamde discrete choice experiments (DCEs). Dit is een methode waarmee het relatieve belang van verschillende aspecten van een product ten opzichte van elkaar in kaart wordt gebracht. Een voorbeeld kan zijn een printer, waarbij de kleur, de grootte, de afdrukkwaliteit en de prijs bepalend kan zijn voor de aantrekkelijkheid van het product. Als product kan ook een medische interventie of een gezondheidstoestand worden ingebracht. DCEs worden nu vooral ingezet om een weging te geven aan de proceskenmerken van een interventie, zoals bij kankerscreening.⁹ DCEs bieden de mogelijkheid om de verschillende dimensies van kwaliteit van leven onderling te wegen. Die onderlinge verhouding is uit te zetten op een schaal van 0 tot 1, en zo kan een QALY-achtige schaal gemaakt worden. Het is evenwel nog niet duidelijk hoe deze schaal zich verhoudt tot de eerder genoemde time trade-off en standard gamble wegingen, omdat de afweging met tijd ontbreekt. Daarmee is het voorlopig nog geen valide vervanger van de standaardanalyse welke wel op basis van time trade off is samengesteld. Het is te verwachten dat het vele onderzoek dat nu gaande is naar DCEs in de nabije toekomst tot beter inzicht zal leiden over de vergelijkbaarheid van waarderingsmethoden die verkregen zijn met verschillende waarderingsmethoden.

3.2.5 *Recente initiatieven: PROMs en Positive health*

Er is internationaal en nationaal veel belangstelling voor het routinematige meten van 'Patients Related Outcomes', of wel PROMs. Voorbeelden zijn:

- Het Patient Reported Outcomes Measurement Information System (PROMIS) gefinancierd door het Amerikaanse 'National Institutes of Health'
- Het 'International Consortium for Health Outcome Measurement' (ICHOM) initiatief, opgericht door de Harvard Business School, het Zweedse Karolinska Instituut en de Boston Consultancy Group
- In Engeland onderneemt de NHS een dergelijk initiatief onder de naam: Patient Reported Outcome Measures (PROMs).

Er wordt vaak over gespeculeerd om deze gegevens te gebruiken bij voor het berekenen QALY. Helaas worden er vaak instrument meegenomen die het berekenen van QALYs niet toelaten. Wanneer dit soort initiatieven ontplooid worden, is het dus zaak om er goed op te letten dat ook aandacht wordt gegeven aan vragenlijsten die gebruikt kunnen worden bij gezondheidseconomische berekeningen.

Naast de PROMs initiatieven is er een over de jaren heen constante stroom van nieuwe ideeën over het definiëren en meten van het begrip gezondheid gerelateerde kwaliteit van leven. Uit het begin van de jaren negentig komt het idee om het individu zelf de mogelijkheid te geven om zijn of haar kwaliteit van leven te definiëren: The Schedule for the Evaluation of Individual Quality of Life (SEIQoL). In de jaren na de eeuwwisseling is veel onderzoek gedaan om de capability approach van Amartya Sen te integreren in gezondheid gerelateerde kwaliteit van leven. Een recente alternatieve benadering die daar op lijkt wordt beschreven in het proefschrift van Machteld Huber: “Towards a new, dynamic concept of health. Its operationalisation and use in public health and healthcare, and in evaluating health effects of food”.¹⁰ Huber stelt voor om niet zo zeer te focussen op de ziekte, maar op persoonlijke groei en ontwikkeling en het vervullen van persoonlijke doelen in het leven. Huber rekt daarmee het begrip gezondheid op tot ‘welzijn’ in plaats van ‘gezondheidsgerelateerde kwaliteit van leven’. Het benadrukken van ‘wat wel kan’ in plaats van ‘wat niet kan’, is terug te vinden in de term ‘positieve gezondheid’.

In alle gevallen proberen de methoden vast te stellen wat voor de patiënt de meeste waarde heeft. De aanpak wordt daarom ook wel ‘waarde gedreven zorg’ genoemd. Een probleem met deze aanpak is dat het vergelijken van effecten of waarden van de interventies alleen kan binnen ziektebeelden, en niet tussen ziektebeelden. Dat komt omdat de meetinstrumenten per ziektebeeld kunnen verschillen. Men past als het ware steeds de meetlat aan. Daardoor kan men wel heel verfijnd meten binnen een patiëntengroep, maar een vergelijking tussen patiëntengroepen is niet mogelijk.

Een tweede probleem daaraan gerelateerd is dat de relatie met de monetaire waarden onduidelijk is: er kan niet gezegd worden of een verbetering op een schaal een doelmatige besteding van geldelijke middelen is. Doordat deze stroming de terminologie kiest van ‘waarde gedreven zorg’ wordt de indruk gewekt dat ook de ‘monetaire waarden’ wordt vastgesteld. Dat is dus een misverstand. De enige uitzondering is het Engelse PROMs initiatief. Daar heeft men vanaf het begin gekozen voor een aanpak waarbij het vergelijken van de waarden binnen en tussen patiëntengroepen allebei mogelijk is, en waarbij een relatie gelegd kan worden met de bestedingen in de gezondheidszorg. Dat doet men door in alle gevallen de EQ-5D-3L op te nemen bij de vragenlijsten voor de patiënt. Dus ook hier heeft men er voor gekozen om de EQ-5D te gebruiken als standaardisatie en link naar monetaire waarden, náást ziektespecifieke instrumenten. Dit is in lijn met de voorgestelde richtlijn zoals hier besproken.

3.2.6 *Andere QALY-vragenlijsten dan de standaardanalyse*

Binnen deze richtlijn is de EQ-5D-5L gekozen als standaardanalyse: het is de vragenlijst die in elk geval meegenomen moet worden bij het meten van de effecten van een interventie. Dat de keuze op de EQ-5D-5L is gevallen komt voort uit het feit dat deze vragenlijst het meest onderzocht en doorontwikkeld is. Voor wat betreft het gebruik van de besproken alternatieve vragenlijsten en methoden naast de EQ-5D-5L kan een ‘voorkeursvolgorde’ worden beargumenteerd bij gebruik in het kader van economische evaluatie, en wel als volgt:

- De EQ-5D-3L
- Andere QALY vragenlijsten zoals de SF-6D en de HUI
- Domeinspecifieke QALY vragenlijsten, zoals de FACT-L en EORTC-C30
- Ziektespecifieke QALY vragenlijsten
- Mapping
- Het rechtstreeks waarderen van gezondheidstoestanden in het model
- Gebruikmaken van kwaliteit van leven gewichten uit de literatuur
- Hieronder worden deze voorkeursvolgorde uitgewerkt.

3.2.6.1 *EQ-5D-3L*

De EQ-5D-5L bouwt voort op de kennis van de EQ-5D-3L; de meest gebruikte kwaliteit van leven vragenlijst bij economische evaluaties. De EQ-5D-3L werd ook in de vorige versie van de richtlijn aanbevolen en kent ook een Nederlandse waardering. De EQ-5D-3L is dan ook in veel lopend onderzoek opgenomen. Het zou onredelijk zijn om dit lopend onderzoek nu te diskwalificeren. Dat geldt ook voor internationaal onderzoek uit landen waar de EQ-5D-5L nog niet operationeel is. Voor onderzoek dat nu start, is het evenwel niet wenselijk dat er gekozen kan worden tussen twee verschillende versies van de EQ-5D. Daarom geldt voor nieuw onderzoek dat de EQ-5D-5L gebruikt moet worden. Het is te verwachten dat

veel andere landen ook een voorkeur zullen uitspreken voor de 5 level versie van de EQ-5D. Daarmee zal de EQ-5D-3L waarschijnlijk langzaam van het toneel verdwijnen.

3.2.6.2 *Andere QALY vragenlijsten zoals de SF-6D en de HUI*

Naast de EQ-5D zijn er ook andere generieke vragenlijsten die waarderingen geven die gebruikt kunnen worden in QALY analyses. De meest bekende zijn de SF-6D en de HUI. Zoals hierboven beschreven ontbreekt het bij deze vragenlijsten aan gepubliceerde Nederlandse waarderingen. Er zijn evenwel een aantal speciale gevallen waarin het gebruik van de vragenlijsten een zinvolle toevoeging zou zijn náást het standaardanalyse. Als voorbeeld kan het gebruik van de SF-6D genoemd worden. De SF-6D is afgeleid van SF-36, een vragenlijst die speciaal ontworpen is om te gebruiken in de 'algemene populatie'. In de algemene populatie komen relatief veel milde gezondheidsproblemen voor en daarmee wordt bij de SF-36 dan ook nadrukkelijk rekening gehouden. De EQ-5D richt zich juist op de zieke populatie. De SF-36, en daarmee de SF-6D, lijken dan ook meer sensitief voor kleine verschillen in relatief milde gezondheids-toestanden dan de EQ-5D. Wanneer de interventie betrekking heeft op patiënten met een relatief goede gezondheid, kan de SF-6D dus een zinvolle toevoeging zijn.

Een ander voorbeeld is de HUI mark III: de Health Utility Index. De HUI is ontwikkeld vanuit de kindergeneeskunde, en heeft daardoor sterk ontwikkelde proxy-versies die ingezet kunnen worden bij kinderen en bij patiënten die zelf geen vragenlijsten kunnen invullen, zoals dementerenden. Wanneer men dus onderzoek doet bij deze patiëntengroepen kan de HUI een goed alternatief zijn náást de standaardanalyse. Naast de SF-6D en HUI zijn er maar weinig andere alternatieve generieke vragenlijsten waarvoor uitgebreid methodologisch onderzoek is verricht. De Finse 15D en QWB ontwikkeld aan de Amerikaanse oostkust waren ooit voorloper in het veld, maar worden tegenwoordig weinig meer gebruikt. Een opkomende vragenlijst is de Australische AQoL-8D, die een bredere definitie heeft van gezondheid: ook relaties, coping en onafhankelijkheid worden er in betrokken. Het is denkbaar dat de AQoL-8D vragenlijst zich ontwikkeld als alternatief voor de EQ-5D, de SF-6D en de HUI. Vooralsnog is er geen Nederlandse validatie en is de relatie met QALYs nog onduidelijk door het ontbreken van time trade-off waarderingen.

3.2.6.3 *Domeinspecifieke QALY vragenlijsten, zoals de FACT-L en EORTC-C30*

Naast de generieke vragenlijsten zoals de EQ-5D, SF-6D en de HUI, zijn er een aantal domeinspecifieke vragenlijsten die ongeveer op dezelfde manier gewaardeerd zijn. De bekendste voorbeelden zijn de FACT-L en EORTC-C30. Deze twee vragenlijsten zijn speciaal ontwikkeld voor oncologische patiënten. Een nadeel van deze vragenlijsten is dat ze niet generiek zijn, waardoor belangrijke domeinen van de kwaliteit van leven buiten beschouwing worden gelaten. In het geval van de FACT-L en EORTC-C30 valt dat nadeel wel mee, omdat deze vragenlijsten zich richten op een groep van ziekten en daarom kwaliteit van leven relatief breed meten. Een belangrijk nadeel van deze lijsten is dat de waarderingen door het algemeen publiek meestal via een aantal omwegen tot stand komen. Deze omwegen zijn noodzakelijk omdat de lijsten in eerste instantie niet ontworpen waren voor dit soort exercities, en omdat er vaak maar beperkte middelen voor handen waren om het waarderingsonderzoek uit te voeren. Daarom zijn eenvoudige waarderingsmethoden gebruikt zoals de visual analoge scale (VAS) bij het waarderen van de FACT-L. Deze waarderingsmethoden hebben een lagere validiteit binnen QALY analyses dan time trade-off, omdat de component 'tijd' niet betrokken wordt bij het uitdrukken van de waardering. Daardoor is de relatie met het QALY concept onduidelijker dan bijvoorbeeld bij time trade-off.

3.2.6.4 *Ziektespecifieke QALY vragenlijsten*

Hierboven is al beschreven dat gevalideerde ziektespecifieke vragenlijsten gepresenteerd kunnen worden náást het standaardanalyse. Een belangrijk nadeel bij de interpretatie van de scores van ziektespecifieke vragenlijsten is dat er meestal geen link is met QALYs. Bij een aantal ziektespecifieke vragenlijsten is die link er wel, doordat de vragenlijst op een vergelijkbare manier gevalideerd is als bijvoorbeeld de EQ-5D. Deze ziektespecifieke vragenlijsten worden ook vaak gezien als een sensitieve vervanger van de generieke vragenlijsten zoals de EQ-5D. Toch kunnen deze ziektespecifieke vragenlijsten de generieke vragenlijsten niet vervangen. Het belangrijkste probleem is dat, uit de aard van de zaak, de ziektespecifieke vragenlijsten niet alle facetten van gezondheidsgerelateerde kwaliteit van leven meten. Daardoor wordt bij patiënten met comorbiditeit welke buiten het meetdomein ligt van de ziektespecifieke vragenlijst-

ten, de waarde van de kwaliteit van leven overschat. Daarnaast kunnen bijeffecten worden gemist. Dat betekent dat niet alleen de absolute waarde van kwaliteit van leven, maar ook de effectgrootte worden beïnvloed door de beperkte reikwijdte van de ziektespecifieke vragenlijst. Een bijkomend probleem is dat het nog onduidelijk is onder welke condities de uitkomsten van ziektespecifieke vragenlijsten dezelfde waarderingsopleveren als die van generieke vragenlijsten. Een probleem van ziektespecifieke vragenlijsten is bijvoorbeeld de sterke focus op slechts één of enkele aspecten van gezondheid. Door die focus kan de waarderingsopleveren voor die aspecten worden overschat (framing). Tot op heden is het nog onbekend onder welke omstandigheden de waarderingsopleveren van generieke en ziektespecifieke meetinstrumenten vergelijkbaar zijn.

3.2.7 Mapping

Zoals boven reeds besproken, kunnen ziektespecifieke vragenlijsten op een vergelijkbare manier worden 'gewaardeerd' als bij de EQ-5D-5L. Soms wordt een andere techniek toegepast: 'mapping'. Bij mapping worden een ziektespecifieke vragenlijst en een reeds gewaardeerde generieke vragenlijst, meestal de EQ-5D of de SF-6D, parallel afgenomen bij een representatieve groep patiënten. Via een regressieanalyse wordt de waardering van de EQ-5D of de SF-6D gezondheidstoestand gekoppeld aan de score van de ziektespecifieke vragenlijst. De waardering van de EQ-5D of de SF-6D is dan voorspelbaar uit de ziektespecifieke vragenlijst. Er zijn echter verschillende problemen met mapping.

Ten eerste moet de 'mappingfunctie' steeds worden vastgesteld in precies dezelfde populatie als de populatie waarop de economische evaluatie betrekking heeft. Daarbij moet het aannemelijk zijn dat de variatie in de ziektespecifieke component van kwaliteit van leven direct gerelateerd is aan de variatie in de generieke component, en dat deze relatie niet loopt via een andere route. Bijvoorbeeld: in een trial met oudere mannen wordt een middel getest tegen plasklachten en de effecten worden getest met een ziektespecifieke vragenlijst die specifiek vraagt naar de plasklachten. Stel nu dat de relatie tussen deze ziektespecifieke plasklachtenlijst en de generieke vragenlijst wordt gemeten, bijvoorbeeld de EQ-5D in een populatie van mannen wordt afgenomen, waarin zowel oude mannen zijn vertegenwoordigd als jonge mannen. In dat geval wordt de relatie tussen kwaliteit van leven op de generieke lijst en ziektespecifieke lijst overtrokken. Dat komt omdat jonge mannen weinig tot geen plasklachten kennen én gemiddeld genomen een prima kwaliteit van leven hebben. Wanneer deze mappingfunctie wordt gebruikt in een trial met de oudere mannen, dan lijkt het dat wanneer de plasklachten verdwijnen, de kwaliteit van leven van de oude mannen op het niveau komt van de jonge mannen. Het effect van de interventie wordt daardoor overschat. Omdat de route langs covariaten nooit helemaal is uit te sluiten, blijft mapping een onzekere methode om tot kwaliteit van leven gewichten te komen.

De validiteit van de mapping neemt toe wanneer de vragenlijsten meer op elkaar lijken. Omdat bij mapping altijd een generieke vragenlijst betrokken is, bijvoorbeeld de EQ-5D of de SF-6D, geldt dat 'hoe meer generiek' de ziektespecifieke vragenlijst is, hoe beter het resultaat. Een voorbeeld daarvan is de mapping van de EORTC-C30, een vragenlijst met generieke eigenschappen die gebruikt wordt in de oncologie. In alle gevallen gaat ook informatie verloren: de mappingfunctie zal nooit perfect zijn, en daarmee gaat ook de sensitiviteit achteruit.

Mapping is een controversiële techniek omdat het ziektespecifieke karakter maakt dat er informatie gemist kan worden, en omdat de er bijna nooit garanties zijn dat regressiecoëfficiënten onafhankelijk zijn van de populatie waarin ze gevonden worden. Daarom beland mapping laag in de rangorde van alternatieven.

3.2.7.1 *Het rechtstreeks waarderen van gezondheidstoestanden in het model*

Soms bestaat de mogelijkheid om de gezondheidstoestanden die gebruikt worden bij modelmatige economische evaluaties rechtstreeks te valideren. Zo'n beslismodel kan bijvoorbeeld bestaan uit de toestand 'gebroken heup', 'vervangen heup' en 'ontstoken vervangen heup'. Deze drie gezondheidstoestanden kunnen dan rechtstreeks worden voorgelegd aan het algemeen publiek en dan worden gewaardeerd middels time trade-off.

Er zijn een aantal problemen met deze aanpak die aandacht verdienen. Ten eerste moet het algemeen publiek een juist beeld krijgen van de gezondheidstoestand. Dat is afhankelijk van de klinische beschrijving en het is lastig vast te stellen hoe adequaat die beschrijving is en hoe goed de respondenten die

beschrijving begrepen hebben. Dus het kan zijn dat de klinische beschrijving de ernst van de klachten overdrijft of dat deze juist onderschat wordt in de beschrijving. Omdat de beschrijving van de gezondheidstoestand in de handen ligt van de onderzoekers, komt er ook een subjectief element in de analyses, welke met gevalideerde vragenlijst juist wordt vermeden. Daarnaast hebben we geen idee van de variatie in de kwaliteit van leven tussen de patiënten in de gezondheidstoestanden: immers er komt maar één waarde per gezondheidstoestand. Deze variatie kan dan ook niet worden meegenomen in het gezondheidseconomische beslismodel. Daarnaast heeft deze methode ook een aantal nadelen, die ook al genoemd zijn bij de ziektespecifieke vragenlijsten, zoals het probleem dat niet alle domeinen van kwaliteit van leven gedekt worden en er sprake kan zijn van een focuseffect.

3.2.7.2 *Gebruik maken van kwaliteit van leven gewichten uit de literatuur*

Bij modelstudies waarin ook gezondheidstoestanden voorkomen met een lage frequentie (zoals bijvoorbeeld een zeldzame comorbiditeit), kan het voorkomen dat er te weinig observaties zijn om de waarde van de kwaliteit van leven betrouwbaar te schatten. In dat geval kan er voor gekozen worden om een bij de gezondheidstoestand passende kwaliteit van leven waardering op te zoeken in de literatuur. Omdat verschillende vragenlijsten, waarderingsmethoden en perspectieven verschillende waarderingsmethoden, is het zaak een onderzoek te vinden in de literatuur waarin deze zaken gelijk zijn aan de vragenlijsten, waarderingsmethode en perspectieven van het onderzoek zelf. Als bijvoorbeeld de studie is gebaseerd op de EQ-5D-5L, dan moeten ook de waarden uit de literatuur gebaseerd zijn op die vragenlijst. Daarnaast moet het onderzoek in de literatuur in een gelijke populatie hebben plaatsgevonden als de populatie in de studie. Anders ontstaan er problemen zoals die beschreven zijn bij 'mapping'.

Wanneer er helemaal geen empirische kwaliteit van leven gegevens zijn die gebruikt kunnen worden voor het berekenen van QALYs, is het terugvallen op waarderingsmethoden uit de literatuur een laatste redmiddel om het gezondheidseconomische model te vullen. Een veelgemaakte fout hierbij is dat dan waarderingsmethoden vastgesteld middels verschillende vragenlijsten, waarderingsmethoden en perspectieven worden gecombineerd. Zoals hierboven betoogd is geeft dat verschillen in zowel de absolute als in de relatieve waarde van kwaliteit van leven. Wanneer voor dit laatste redmiddel gekozen wordt, dienen waarderingsmethoden dus zoveel mogelijk gebaseerd te zijn op dezelfde vragenlijsten afgenomen bij vergelijkbare patiënten.

4 Rapportage

Hoe meer er naast de standaardanalyse wordt aangeboden, hoe uitgebreider de rapportage moet zijn om de validiteit van de analyse op waarde te kunnen schatten. Er zijn een aantal zaken die altijd vermeld dienen te worden. Voor de standaardanalyse moet altijd worden aangegeven welke vragenlijst is gebruikt: de EQ-5D-5L of de EQ-5D-3L, en welke nationale waarderingen er gebruikt zijn. Wanneer naast de standaardanalyse nog alternatieven worden gepresenteerd moet ook steeds weer worden aangegeven hoe deze alternatieven zijn gevalideerd en welke nationale waarderingen voor de gezondheidstoestanden zijn gebruikt. Een peer-reviewed gepubliceerd artikel helpt de validiteit te onderbouwen. Bij een modelstudie met gezondheidstoestanden, zoals een Markov model, moeten de waarderingen per gezondheidstoestand worden weergegeven. Wanneer de QALYs worden geschat middels een 'area under the curve' analyse, moet het verloop van de gemiddelde kwaliteit van leven in de tijd worden weergegeven. Wanneer waarderingen van kwaliteit van leven uit de literatuur worden gehaald, moet vermeld worden welke vragenlijst of waarderingsstechniek is gebruikt, wat de nationaliteit van de respondenten was en hoe representatief de respondenten zijn voor het maatschappelijk perspectief: zijn het waarderingen van het algemeen publiek of zijn het patiëntwaarderingen.

Referenties

- 1 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Fourth Edition, Oxford press 2015.
- 2 Robinson, J. *Economic Philosophy*, Chicago: Adeline Pub. Co., 1962; p. 47
- 3 Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. *Social Science and Medicine*. 2014; 114:38-48.
- 4 Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Medical Decision Making*. 2015; 35(3): 328-40.
- 5 Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. 2015 Beschikbaar via: www.zorginstituutnederland.nl
- 6 de Wit GA, van Busschbach JJ, de Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Economics*. 2000; 9: 109-26.
- 7 EuroQoL. Beschikbaar via: <http://www.euroqol.org/>
- 8 van Busschbach JJ, Hinten M, Rikken B, Grobbee DE, de Charro FT, Wit, JM. (Editors: Eiholzer U, Haverkamp F, Voss L). Some patients with idiopathic short stature see their short stature as a problem but others do not: Why this difference? In: *Growth, Stature and Psychosocial Well-Being*. Hogrefe & Huber Publishers, Seattle, Toronto, Bern, Gottingen, 1999.
- 9 de Bekker-Grob EW, Rose JM, Donkers B, Essink-Bot ML, Bangma CH, Steyerberg EW. Men's preferences for prostate cancer screening: a discrete choice experiment. *British Journal of Cancer*. 2013; 108(3): 533-41.
- 10 Huber M. Towards a new, dynamic concept of Health. Its operationalisation and use in public health and healthcare, and in evaluating health effects of food. Proefschrift Universiteit Maastricht, 17 December 2014.

Colofon

Contactpersoon

S. Knies
sknies@zinl.nl

Adres

Zorginstituut Nederland
Postbus 320
1110 AH DIEMEN