
Kostenutiliteitsanalyse

dr JND de Neeling



Kostenutiliteitsanalyse

dr JND de Neeling

Nr A03/01, Den Haag, 3 februari 2003

De Gezondheidsraad, ingesteld in 1902, is een adviesorgaan met als taak de regering en het parlement “voor te lichten over de stand der wetenschap ten aanzien van vraagstukken op het gebied van de volksgezondheid” (art. 21 Gezondheidswet).

De Gezondheidsraad ontvangt de meeste adviesvragen van de bewindslieden van Volksgezondheid, Welzijn & Sport, Volkshuisvesting, Ruimtelijke Ordening & Milieubeheer, Sociale Zaken & Werkgelegenheid, en Landbouw, Natuurbeheer & Visserij. De Raad kan ook eigener beweging adviezen uitbrengen. Het gaat dan als regel om het signaleren van ontwikkelingen of trends die van belang kunnen zijn voor het overheidsbeleid.

De adviezen van de Gezondheidsraad zijn openbaar en worden in bijna alle gevallen opgesteld door multidisciplinair samengestelde commissies van—op persoonlijke titel benoemde—Nederlandse en soms buitenlandse deskundigen.



De Gezondheidsraad is lid van het International Network of Agencies for Health Technology Assessment (INAHTA). INAHTA bevordert de uitwisseling en samenwerking tussen de leden van het netwerk.

U kunt het rapport downloaden van www.gr.nl.

Deze publicatie kan als volgt worden aangehaald:
JND de Neeling. Kostenutiliteitsanalyse. Den Haag: Gezondheidsraad, 2003; publicatie nr A03/01.

auteursrecht voorbehouden

ISBN: 90-5549-463-1

Inhoud

Ten geleide 7

Samenvatting 9

-
- 1 Inleiding 11
 - 1.1 Achtergrond 11
 - 1.2 Het rapport *Volksgezondheidszorg* 12
 - 1.3 Het advies *Grenzen van de gezondheidszorg* 13
 - 1.4 Dit rapport 14

2 Doelmatigheid 17

-
- 3 Utiliteit 21
 - 3.1 Het begrip ‘gezondheidswinst’ 21
 - 3.2 Beschrijving van gezondheidstoestanden 28
 - 3.3 Waarderen van de kwaliteit van leven 34
 - 3.3.1 Aard van de waardeoordelen 34
 - 3.3.2 Keuze van de beoordelaars 37
 - 3.3.3 Keuze van een meetprocedure 41

4 Kosten 45

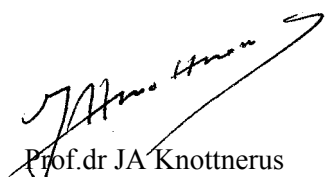
- 4.1 Productiviteitskosten 46
-

4.2	Kosten in gewonnen levensjaren	48
4.3	Marginale versus gemiddelde kosten	49
<hr/>		
5	Disconteren	53
<hr/>		
6	Verdelingseffecten	57
<hr/>		
7	Tot besluit	61
<hr/>		
	Literatuur	71
<hr/>		
	Bijlage	85
A	Verantwoording	87

Ten geleide

Voor u ligt een achtergrondstudie bij het advies *Contouren van het basispakket*. De studie is op persoonlijke titel geschreven door dr JND de Neeling, secretaris bij de Gezondheidsraad, ten behoeve van de Commissie Contouren van het Basispakket. De conclusies die de commissie eraan verbindt zijn opgenomen in het advies zelf.

Het is een doorwrocht en boeiend rapport geworden. Het schetst de mogelijkheden en beperkingen van de kostenutiliteitsanalyse bij het in kaart brengen van de doelmatigheid van de gezondheidszorg. Het rapport maakt duidelijk dat de methodologie van de kostenutiliteitsanalyse nog veel problemen in zich draagt, waardoor haar bruikbaarheid voor de vergelijking van de kostenutiliteit van verschillende voorzieningen over de volle breedte van de gezondheidszorg bescheiden is. De gepresenteerde analyse bevat daarmee enerzijds uitdagingen voor het wetenschappelijk veld op het gebied van het doelmatigheidsonderzoek. Anderzijds biedt zij een nuchtere plaatsbepaling van de kostenutiliteitsanalyse in het kader van het beleid.



Prof. dr JA Knottnerus
voorzitter Gezondheidsraad

Samenvatting

Kan bij het afbakenen van een collectief te financieren basispakket de kostenutiliteitsanalyse (KUA) dienen als instrument om de doelmatigheid van alle preventieve en curatieve gezondheidszorgvoorzieningen op één uniforme wijze in maat en getal uit te drukken? Deze vraag, in 1997 door de WRR in het rapport *Volksgezondheidszorg* bevestigend beantwoord, staat centraal in deze achtergrondstudie, die is geschreven ter ondersteuning van de discussie in de Gezondheidsraadcommissie Contouren van het Basispakket. De studie inventariseert de aannamen en keuzen in de methodologie van de KUA en geeft een overzicht van wat daarover geschreven is in de gezondheidseconomische, medische en medisch-ethische literatuur.

Allereerst wordt nagegaan hoe in de KUA de algemene richtinggevende notie van de doelmatigheid getransformeerd wordt tot een op de gehele gezondheidszorg toepasbare verhouding tussen input ('kosten') en output ('gezondheidswinst') van gezondheidszorgvoorzieningen: de kostenutiliteitsverhouding.

Het daarop volgende, meest uitvoerige hoofdstuk is gewijd aan de algemeen toepasbare maat voor gezondheidswinst die de KUA onderscheidt van andere vormen van doelmatigheidsanalyse: het Quality Adjusted Life Year (QALY). De weg die leidt naar de QALY blijkt geplaveid met een reeks voetangels en klemmen waarover nog volop discussie is: problemen bij het integreren van 'kwaliteit' en 'kwantiteit' van leven, bij het optellen van QALY's van verschillende personen, bij het op een gestandaardiseerde wijze beschrijven van gezondheidstoestanden, bij het formuleren van een theorie over voorkeuren voor gezondheidstoestanden, bij het ontwerpen van procedures voor het meten van die voorkeuren en bij de keuze van de beoordelaars.

De volgende twee hoofdstukken gaan in op onderdelen van de KUA die inherent zijn aan elke vorm van kwantitatieve doelmatigheidsanalyse: het kwantificeren van de kosten en het terugrekenen van toekomstige gezondheidswinst naar het heden ('disconteren'). In de wetenschappelijke literatuur blijkt sprake van verschillende benaderingswijzen van diverse kostenposten – bijvoorbeeld van de directe gezondheidszorgkosten, de productiviteitskosten en de kosten in gewonnen levensjaren – die tot belangrijke verschillen in de schatting van de kosten kunnen leiden. Over de vraag of en, zo ja, hoe toekomstige gezondheidswinst gedisconteerd moet worden bestaat evenmin consensus.

Vervolgens wordt vastgesteld dat de KUA, toegepast als hulpmiddel om te komen tot een maximaal aantal gewonnen QALY's bij een gegeven gezondheidszorgbudget, tal van implicaties heeft voor de verdeling van de behaalde gezondheidswinst.

In het laatste hoofdstuk wordt geconcludeerd dat de QALY en de KUA, ruim 25 jaar na hun introductie, nog steeds in allerlei opzichten ter discussie staan. Bij het beoordelen van de doelmatigheid van zorgvoorzieningen biedt de KUA slechts beperkt houvast. Alternatieven voor de QALY, zoals het *saved young life equivalent* en de *willingness to pay*-benadering, verdienen nadere exploratie. Maar om een collectief gefinancierd basispakket af te bakenen, zal het vooral zaak zijn een transparante besluitvormingsprocedure te ontwerpen, waarin – binnen deelgebieden van de gezondheidszorg – de partijen die de gezondheidszorg van binnenuit kennen, zorgverleners en patiënten voorop, actief participeren. De KUA kan in dat besluitvormingsproces een hulpmiddel zijn om overzicht en overeenstemming te bereiken.

Inleiding

1.1 Achtergrond

In de gehele westerse wereld worstelen regeringen met het probleem hoe tegemoet te komen aan de groeiende vraag naar gezondheidszorg, zonder de collectieve uitgaven aan gezondheidszorg te sterk te laten stijgen (Kle93). Zo ook in Nederland.

In februari 2001 vroeg de minister van VWS, in het kader van de voorbereiding van de politieke discussie over de hervorming van het zorgverzekeringsstelsel, de Gezondheidsraad zijn visie te geven op de “op wetenschappelijke inzichten gebaseerde, operationaliseerbare criteria voor de in een basispakket op te nemen zorgvoorzieningen”. De minister vestigde de aandacht speciaal op het criterium van de doelmatigheid en gaf te kennen de mening van de Raad te willen vernemen over “de mogelijkheid het doelmatigheidscriterium te gebruiken voor het bepalen van de samenstelling van een basispakket”.

De voorzitter van de Gezondheidsraad heeft de beantwoording van de adviesaanvraag van de minister opgedragen aan de door hem ingestelde Commissie Contouren van het Basispakket. Dit rapport is een achtergrondstudie ter ondersteuning van de discussie in die commissie over de bruikbaarheid van het doelmatigheidscriterium.

1.2 Het rapport *Volksgezondheidszorg*

In 1997 publiceerde de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) het rapport *Volksgezondheidszorg* (WRR97). De minister van VWS vroeg de Gezondheidsraad expliciet dit rapport in zijn beschouwing te betrekken.

De centrale probleemstelling van het WRR-rapport is hoe de algemene toegankelijkheid en een voor iedereen gelijke kwaliteit van de gezondheidszorg in Nederland op de langere termijn gewaarborgd kunnen blijven. De WRR stelde vast dat het noodzakelijk is een collectief te financieren basispakket van zorg te definiëren. In de besluitvorming over de afbakening van het preventieve en curatieve zorgpakket kende de raad een centrale rol toe aan een specifieke operationalisatie van het doelmatigheids criterium: de kostenutiliteitsverhouding (K/U-verhouding).^{*} De K/U-verhouding of, kortweg, de kostenutiliteit, is gedefinieerd als de *meerkosten* gedeeld door de *meeropbrengst*, gemeten met een op de gehele gezondheidszorg toepasbare maat voor gezondheidswinst, van de op zijn kostenutiliteit te beoordelen voorziening, in vergelijking met een relevant alternatief.^{**} Het alternatief kan de gebruikelijke zorg zijn, een concurrerende voorziening of géén voorziening. Het instrument om de kostenutiliteit te bepalen is de kostenutiliteitsanalyse (KUA).

Als het gaat om de afbakening van het pakket *preventieve* voorzieningen, moet volgens de WRR het maximaliseren van gezondheidswinst, gegeven een bepaald budget, de primaire leidraad zijn. Dit vereist een ordening van alle preventieve programma's naar mate van kostenutiliteit. De WRR stelt voor daarbij als eenheid van gezondheidswinst de DALY te hanteren, het *Disability Adjusted Life Year*, ontleend aan de *Global Burden of Diseases Study* van de WHO (Mur96, Mur97). De DALY is een maat die in veel opzichten lijkt op het in de medische en gezondheidseconomische literatuur bekendere *Quality Adjusted Life Year*, de QALY. Maximalisatie van

* De WRR gebruikte in zijn rapport de algemenere term kosten-effectiviteitsverhouding. Volgens het gangbare gezondheidseconomisch spraakgebruik kan bij het berekenen van een *costeffectiveness* verhouding de opbrengst van de zorgvoorziening in uiteenlopende effectiviteitsmaten worden uitgedrukt: gewonnen levensjaren, klachtenvrije dagen, millimeters Hg bloeddrukverlaging etc. Bij het berekenen van een *costutility* verhouding worden specifieke eisen gesteld aan de uitkomstmaat: hij moet toepasbaar zijn op de gehele gezondheidszorg en de waarde representeren die aan het totaal van alle gezondheidseffecten van een voorziening wordt toegekend (Dru97).

** Een doelmatigheidsanalyse kan ook leiden tot de bevinding dat de kosten en/of de gezondheidsopbrengst van de te onderzoeken voorziening *V* *lager* is dan die van het alternatief *A*. Wanneer de kosten van *V* *lager* zijn en de opbrengst *hoger* is dan die van *A*, dan heet het in de gezondheidseconomie dat *A* door *V* *gedomineerd* wordt. Is het omgekeerde het geval, dan wordt *V* gedomineerd door *A*. In beide gevallen is berekening van een kostenutiliteitsverhouding niet zinvol. Zijn zowel kosten als opbrengst van *V* *lager* dan die van *A*, dan ligt het voor de hand de vergelijking om te draaien en te spreken van de kostenutiliteitsverhouding van *A* ten opzichte van *V*. In de meeste gevallen zal echter van sprake zijn van *meerkosten* en *meeropbrengst* van *V* in vergelijking met *A*.

gezondheidswinst is volgens de WRR te realiseren door de verschillende programma's in volgorde van afnemende kostenutiliteit te financieren, tot het beschikbare budget is uitgeput (p. 148-50).

In de *curatieve* gezondheidszorg zou volgens de WRR gestreefd moeten worden naar een "egalisering van K/U-verhoudingen", over alle categorieën curatieve voorzieningen heen (p. 150-2). De gedachte is dat de besluitvorming over het al of niet collectief financieren van vooral nieuwe curatieve voorzieningen zich moet laten leiden door het principe dat in alle categorieën voorzieningen voor het realiseren van één DALY eenzelfde maximumbedrag moet gelden. Toepassen van de "methodiek" van het egaliseren van de kostenutiliteitsverhouding kan volgens de WRR leiden tot een "objectivering" van de besluitvorming over het wel of niet collectief beschikbaar stellen van curatieve voorzieningen (p. 155). Dat neemt niet weg, aldus de Raad,

"[...] dat ook in de voorgestelde methodiek lastige keuzes overblijven. Zo moet bij gebruik van de DALY-maat gekozen worden wat wel en wat niet als kosten wordt geteld (indirecte kosten, directe kosten enz.) en op welke wijze de baten moeten worden gewaardeerd. Deze moeilijke punten zijn echter geen reden een dergelijke methodiek niet te gebruiken. In zeker opzicht gaat het om technische vragen van uitwerking. Niet alleen kunnen deze keuzes expliciet bediscussieerd worden, bovendien blijft het voordeel behouden dat alle voorzieningen op dezelfde wijze worden gewogen" (p. 155).

De WRR laat in het midden in welk opzicht de "lastige" vragen technisch zijn. De suggestie is dat ze niet alleen technisch zijn in zoverre dat specifieke deskundigheid nodig is om ze te doorgronden, maar ook in die zin dat ze verschillend beantwoord kunnen worden – zodat de 'machinerie' van de KUA op onderdelen een verschillende aanblik kan vertonen – zónder dat dit van wezenlijke invloed is op het uiteindelijke resultaat, immers: "[...] het voordeel [blijft] behouden dat alle voorzieningen op dezelfde wijze worden gewogen". De WRR beargumenteert niet waarom een bepaalde verzameling van methodologische opties, toegepast op verschillende voorzieningen, met uiteenlopende gezondheidseffecten bij divers samengestelde patiëntengroepen, zou leiden tot dezelfde rangordening naar mate van kostenutiliteit als een andere, op "technische" punten anders ingerichte KUA-methodologie.

1.3 Het advies *Grenzen van de gezondheidszorg*

Ruim tien jaar vóór het rapport *Volksgezondheidszorg* verscheen het advies *Grenzen van de gezondheidszorg* (GR86). De Gezondheidsraad kende daarin aan de KUA een veel bescheidener rol toe dan later de WRR. De resultaten van KUA's zouden volgens de Gezondheidsraad niet doorslaggevend moeten zijn bij het rangordenen van aspirant-verstrekkingen, en derhalve bij de besluitvorming. Ze zouden "*behulpzaam* kunnen zijn

bij het verhogen van de kwaliteit van de besluitvorming, maar meer niet” (p. 89). Een rangordening naar mate van kostenutiliteit zou kunnen leiden tot willekeur. Zo’n ordening zou immers afhankelijk zijn van het aanbod van zorg dat zich op enig moment voordoet en van de beperkte mogelijkheden om KUA’s uit te voeren. Bovendien zou over het hoofd gezien kunnen worden dat in de loop der tijd, bij toenemende ervaring en ruimere toepassing, de kostenutiliteit van elementen van zorg kan verbeteren. Tenslotte zijn naast kostenutiliteit nog andere maatstaven van belang voor het al of niet opnemen van verstrekkingen in het pakket. Is de (aspirant-) verstrekking van levensbelang of anderszins essentieel? Is de zorg voor een ieder die ervoor in aanmerking komt, bereikbaar? Draagt de zorg bij aan de gewenste verdeling van de gezondheidszorg, binnen de bevolking en tussen groepen patiënten (p. 90)? Aldus de Gezondheidsraad in 1986.

1.4 Dit rapport

Sinds het verschijnen van het Gezondheidsraadadvies *Grenzen van de gezondheidszorg* is vijftien jaar verstreken. Talloze artikelen zijn gepubliceerd waarin uitgangspunten en methodologie van kostenutiliteitsanalyses zijn bediscussieerd en nieuwe voorstellen zijn gedaan om knelpunten weg te nemen. Het aantal beschikbare KUA’s is verveelvoudigd. In het licht van deze wetenschappelijke ontwikkelingen ligt het voor de hand het oude Gezondheidsraadstandpunt ten aanzien van de bruikbaarheid van kostenutiliteitsanalyses opnieuw te overwegen, temeer daar de minister in haar adviesaanvraag expliciet verwijst naar het WRR-rapport, dat met het oude standpunt van de Gezondheidsraad op gespannen voet staat.

Dit rapport is bedoeld als aanzet tot zo’n nieuwe standpuntbepaling van de Gezondheidsraad. De primaire doelstelling ervan is beperkt: het inventariseren van aannamen en methodologische problemen en daaruit voortvloeiende beperkingen van de bruikbaarheid van de KUA als instrument ter ondersteuning van de besluitvorming over de afbakening van het basispakket. Het rapport is daarmee grotendeels gewijd aan de “lastige keuzes” in de methodologie van de KUA, waarvan de WRR al gewag maakte. Het geeft een overzicht van feiten, meningen en overwegingen die dienaangaande in de gezondheidseconomische, medische en medisch-ethische literatuur naar voren zijn gebracht. In het verlengde hiervan wordt in het slothoofdstuk verkend welke lering hieruit te trekken is ten aanzien van de algemenere vraag naar de toepasbaarheid van het doelmatigheids criterium bij de afbakening van het collectief te financieren basispakket.

De KUA wordt in dit rapport, overeenkomstig het gangbaar gezondheidseconomisch spraakgebruik, primair opgevat als een doelmatigheidsanalyse met de QALY als maat voor gezondheidswinst (CVZ99, Dru97, Hur00, Joh96, Ric94, Tor86, Uyl00). Daarbij wordt ervan uitgegaan dat de KUA wordt uitgevoerd vanuit een

algemeen maatschappelijk perspectief, zodat alle effecten en kosten relevant zijn, ongeacht wie ze ten deel vallen of wanneer ze zich voordoen.

De QALY wordt in de wetenschappelijke literatuur vrij algemeen beschouwd als de theoretisch best onderbouwde en in praktische, methodologische zin verst ontwikkelde uitkomstmaat voor KUA's. Mogelijke alternatieven, zoals de DALY, het *Healthy Year Equivalent* (HYE) en het *saved young life equivalent* (SAVE), komen in het rapport slechts summier ter sprake. Geheel niet aan de orde zullen komen de talrijke methodologische keuzen die niet alleen in KUA's, maar *in het algemeen* van belang zijn voor de kwaliteit van onderzoeksgegevens over de effectiviteit van gezondheidszorgvoorzieningen, hoewel duidelijk is dat ze de kwaliteit en bruikbaarheid van KUA's in belangrijke mate kunnen beïnvloeden. Te denken valt aan keuzen bij het ontwerpen van effectiviteitsonderzoek, de uitvoering ervan en de wijze waarop de resultaten van verschillende onderzoeken in meta-analyses zijn samengevat. Het rapport gaat ook voorbij aan de modellering van gezondheidsuitkomsten op de lange termijn: een terrein van onderzoek vol uitdagingen, voetangels en klemmen, dat niet alleen, maar zeker óók voor KUA's van essentieel belang is (Bar98, Bon91, Maa95, Man96a). *Randomized clinical trials* en andere vormen van effectiviteitsonderzoek hebben meestal een looptijd van ten hoogste enkele jaren; vanuit een algemeen maatschappelijk perspectief uitgevoerde KUA's beogen de integrale effecten van gezondheidszorgvoorzieningen in beeld te krijgen, wat vaak uitspraken over de invloed op de totale levensloop noodzakelijk maakt. Modellering van levenslange behandelingseffecten – op basis van empirisch effectiviteitsonderzoek met een beperkte tijdshorizon, aangevuld met andere gegevens – introduceert in de KUA aanzienlijke onzekerheden. Hetzelfde geldt voor de daaraan gerelateerde voorspelling van de kosten op langere termijn.

De opzet van het rapport is als volgt. Om te beginnen wordt in hoofdstuk 2 nagegaan hoe in de KUA de algemene richtinggevende notie van de doelmatigheid getransformeerd wordt tot een over de gehele gezondheidszorg toepasbaar, kwantificeerbaar begrip dat verwijst naar de verhouding tussen input ('kosten') en output ('gezondheidswinst') van gezondheidszorgvoorzieningen. Hoofdstuk 3, het meest uitvoerige, is gewijd aan de uitkomstmaat die de KUA onderscheidt van andere vormen van doelmatigheidsanalyse: de QALY. De weg die moet leiden naar een algemeen toepasbare maat voor gezondheidswinst blijkt geplaveid met een reeks van aannamen en keuzen waarover het laatste woord niet is gezegd. De hoofdstukken 4 en 5 gaan vervolgens kort in op twee onderdelen van de KUA die niet specifiek zijn voor deze, maar inherent aan elke vorm van kwantitatieve doelmatigheidsanalyse: het kwantificeren van de kosten en het terugrekenen van toekomstige gezondheidswinst naar het heden ('disconteren'). Hoofdstuk 6 stelt aan de orde, eveneens in kort bestek, wat de effecten zijn van allocatiebeslissingen op basis van KUA's op de verdeling van de behaalde gezondheidswinst. In hoofdstuk 7, ten slotte, worden de implicaties van de

voorafgaande hoofdstukken verkend voor de toepasbaarheid van het doelmatigheidscriterium.

Doelmatigheid

Een kostenutiliteitsanalyse geeft op een specifieke manier invulling aan het begrip doelmatigheid. De econoom Hurley constateert in een recent gezondheidseconomisch handboek dat de term *efficiency* in de economie een bepaalde, technische betekenis heeft die niet overeenkomt met het algemene gebruik ervan. In het dagelijkse – dat wil in zijn geval zeggen, Canadese – spraakgebruik heeft het woord, zo stelt hij vast, een sterk normatieve lading: het verwijst naar “not wasting resources” – iets waar men bezwaarlijk tegen kan zijn (Hur00).

Die wervende betekenis lijkt vaak ook aan de orde in het dagelijks gebruik van de Nederlandse woorden ‘doelmatigheid’ en ‘efficiëntie’. Als het gaat om de gezondheidszorg, komt de notie dat middelen niet verspild mogen worden niets tekort aan aantrekkingskracht en relevantie, maar het is duidelijk dat deze algemene, richtinggevende gedachte ver afstaat van een strak gedefinieerd, laat staan één kwantificeerbaar, over de volle breedte van de gezondheidszorg toepasbaar doelmatigheidsbegrip. Ook de betekenis die de Gezondheidsraad in *Medisch handelen op een tweesprong* (GR91) aan het begrip doelmatigheid toekent – “de baten van een verrichting afgezet tegen de kosten in geld, mankracht, middelen en tijd” – is van een dergelijk algemeen en kwantitatief begrip nog ver verwijderd. Hetzelfde geldt voor de definitie die de minister van VWS hanteert in de beleidsbrief *MTA en doelmatigheid van zorg* (Bor95): “doelmatige zorg is doeltreffende zorg, tegen zo laag mogelijke kosten”. Uit geen van deze beide omschrijvingen spreekt de ambitie om vormen van zorg met verschillende doeleinden onderling op doelmatigheid te kunnen vergelijken.

Van de gebruikelijke betekenis van het woord doelmatigheid naar de betekenis in KUA's is, kortom, een lange weg te gaan. De kostenutiliteitsverhouding van de KUA is een speciale transformatie van het informele doelmatigheidsbegrip. Hij komt voort uit een kwantitatief model van de relatie tussen input en output bij gezondheidszorgvoorzieningen en vertegenwoordigt daarmee een vereenvoudiging van de werkelijkheid.

In dat model wordt de output afgebakend als: gezondheidswinst, meer bepaald: als de waarde die die gezondheidswinst vertegenwoordigt voor patiënten of potentiële patiënten. Verschillende auteurs hebben gewezen op aspecten van de opbrengst van gezondheidszorg die op die manier uit zicht blijven (Ger93, Goo96, Moo94a, Moo94b, Moo00, Nee00, RGO01): de waarde van informatie *per se*, in de vorm van de overdracht van diagnostische, etiologische, pathofysiologische en prognostische inzichten; de waarde van geruststelling; de waarde van respect, zorg en aandacht die patiënten ondervinden en van wat meer in het algemeen wel de *process utility* van gezondheidszorg is genoemd (Don97); het herstel of de bevestiging van hun autonomie; de waarde van de continuïteit in de zorg (Shi97); de positieve effecten die van gezondheidszorg uitgaan op de naasten van patiënten (Bar99, Bro01, Har01, Kor99, Loo89, Wei97a); de waarde die voor mensen persoonlijk en voor de maatschappij als geheel schuilt in de wetenschap dat werkzame zorgvoorzieningen voor iedereen beschikbaar zijn (Hur00, Moo00); meer specifiek: de waarde van het voorkómen van ongerustheid bij ouders, en van maatschappelijke onrust, als het gaat om bijvoorbeeld het al of niet invoeren van een vaccinatie tegen meningokokken bij kinderen; de waarde van herstel of verbetering van iemands uiterlijk (Bon01a); de waarde van nieuw menselijk leven (Bro93). Voor een deel zijn deze aspecten inherent aan gezondheidszorg in het algemeen en aan een goede bejegening van patiënten en hun naasten. Maar voor een niet onaanzienlijk deel zijn het ook ingrediënten en voortbrengselen van gezondheidszorg die bewezen effectief is, die voldoet aan de eisen van de tijd, zorg waar patiënten en zorgverleners vertrouwen in kunnen hebben, kortom: het zijn voor een belangrijk deel aspecten waarin specifieke voorzieningen wel degelijk van elkaar kunnen verschillen.

Ook de input van gezondheidszorg wordt in KUA's op een specifieke wijze gedefinieerd. Uitgangspunt is dat middelen die aan de ene vorm van zorg onttrokken worden zonder meer naar een andere, meer doelmatige vorm zijn over te hevelen (Dru97). Veronachtzaamd wordt dat de kosten op zichzelf ook opbrengsten impliceren, in de vorm van arbeidssatisfactie en inkomen van zorgverleners. De afbakening van kosten die wel en die niet in de analyse betrokken worden – zoals productiviteitsverliezen, kosten van gezondheidszorg in door medische behandeling gewonnen levensjaren – is aan discussie onderhevig (Bro01, Hur00, Koo98, Luc96). Er wordt veelal aan voorbijgegaan dat de reële kosten afhankelijk kunnen zijn van de

schaal waarop een voorziening wordt toegepast, van de ervaring die ermee is opgedaan en, meer in het algemeen, van allerlei tijd- en plaatsgebonden factoren (Ash00, Ger93). Voor zowel input als output van voorzieningen geldt dat ze in de regel door middel van discontering naar één bepaald moment worden toegerekend, zonder rekening te houden met de ongewisse, maar niettemin tot op zekere hoogte – onder meer door politieke beslissingen – beïnvloedbare dynamiek van maatschappelijke en wetenschappelijke ontwikkelingen.

De K/U-verhouding, zo stelden we vast, vloeit voort uit een model en is dus een vereenvoudiging van de werkelijkheid. Dat is, zoals bekend, tegelijk de kracht en de zwakte van een model. Een model biedt overzicht en inzicht dat anders door de complexiteit van de werkelijkheid moeilijk te verwerven zou zijn. Het biedt de mogelijkheid om de overwegingen die bij een bepaald besluit relevant zijn, in kaart te brengen. Het kan helpen om de discussie met diverse partijen die bij dat besluit betrokken zijn, ordelijk te voeren, punten van consensus en meningsverschil te identificeren en zo stapsgewijs te komen tot een overeenstemming die anders moeilijk te bereiken zou zijn, gegeven de complexiteit van het probleem en de vele tegenstrijdige opinies en belangen (Vle90). Maar het gevaar bestaat dat de ontwerpers en de gebruikers van het model, al of niet bewust, vergeten dat zij met een model te maken hebben en overwegingen die niet of onvoldoende in het model tot uitdrukking komen, buiten de orde verklaren of niet goed op waarde weten te schatten. Bovendien is het van belang om in te zien dat bij de modellering tal van normatief geladen keuzen in het spel komen (Ber98).

Hurley, eerder genoemd, wijst op het feit dat technische economische begrippen als ‘efficiëntie’ en een ‘optimale’ verdeling van middelen in de publieke arena onvermijdelijk een normatieve betekenis krijgen. De zogenaamde welvaartseconomie, de economische traditie waarin veel economisch georiënteerde KU-onderzoekers hun werk plaatsen, is onder meer daardoor, zo stelt hij, hoezeer veel economen er ook een ‘objectieve wetenschap’ in willen zien, onvermijdelijk een vorm van sociale ethiek en dient als zodanig behandeld te worden (Hur00). Deze vaststelling is van belang voor de interpretatie van op KUA’s gebaseerde uitspraken over de doelmatigheid van gezondheidszorgvoorzieningen. Zij impliceert dat een kostenutiliteitsverhouding niet op te vatten is als de weergave van een objectieve stand van zaken in de werkelijkheid, maar als een normatief geladen stellingname in het debat over de verdeling van schaarse middelen in de gezondheidszorg.

Utiliteit

Het woord ‘utiliteit’ in de kostenutiliteitsverhouding kán, maar hoeft niet te verwijzen naar een van de betekenissen die in de loop der tijd door filosofen en economen aan die term zijn toegekend (Sen91). De term ‘kostenutiliteitsanalyse’ staat in het gangbare gezondheidseconomisch spraakgebruik voor die doelmatigheidsanalyse die voor het kwantificeren van de opbrengst van een gezondheidszorgvoorziening een maat hanteert die, ten eerste, toepasbaar is op de gehele gezondheidszorg en, ten tweede, de waarde representeert die aan het totaal van alle gezondheidseffecten van die voorziening wordt toegekend (Dru97). In dit hoofdstuk passeren achtereenvolgens de verschillende methodische stappen die bij het meten van de waarde van gezondheidswinst in KUA’s in het geding zijn, de revue.

3.1 Het begrip ‘gezondheidswinst’

Hoe is de opbrengst van een medische behandeling te kwantificeren? De meest gangbare benadering bij het schatten van de gezondheidswinst in kostenutiliteitsanalyses gaat uit van de gedachte dat de opbrengst van een gezondheidszorgvoorziening te ontleden is in twee dimensies: levensduur en gezondheidsgelateerde levenskwaliteit. Een medische behandeling kan het leven verlengen, de kwaliteit van het leven bevorderen, of allebei. In deze gedachtegang is het mogelijk te spreken over de ‘kwaliteitsgecorrigeerde levensduur’ van een persoon, die te berekenen is door voor elke levensperiode die in een bepaalde gezondheidstoestand wordt doorgebracht, de duur te vermenigvuldigen met de bij die gezondheidstoestand

behorende kwaliteitsfactor, variërend van 0 (dood) tot 1 (volmaakt gezond), en vervolgens deze rekenkundige producten te sommeren.* De opbrengst van een behandeling, in vergelijking met geen behandeling, bij één persoon is dan te kwantificeren als het verschil tussen de kwaliteitsgecorrigeerde levensduur mét en zonder behandeling.

Dit is de gedachtegang die ten grondslag ligt aan de QALY als eenheid van gezondheidswinst. Een QALY, één kwaliteitsgecorrigeerd levensjaar, staat voor één levensjaar doorgebracht in een toestand van volmaakte gezondheid, of voor twee jaren doorgebracht in een gezondheidstoestand die ‘half zo wenselijk’ is, enzovoort. Een behandeling die het leven met tien jaar verlengt, terwijl de gezondheidstoestand ‘half zoveel waard is’ als een toestand van volmaakte gezondheid, en een behandeling die het leven niet verlengt, maar voor de duur van 10 jaar de gezondheidsgerelateerde kwaliteit van het leven van 0,5 tot 1 doet toenemen, leveren beide 5 QALY’s op.

Hoe aantrekkelijk ook in zijn eenvoud, het QALY-concept heeft van meet af aan kritische vragen opgeroepen, bij kostenutiliteitsonderzoekers en bij minder ingewijden. Is het reëel om ervan uit te gaan dat aan een bepaalde gezondheidstoestand één kwaliteitswaarde kan worden toegekend, onafhankelijk van de duur van die toestand, van de toestanden die eraan vooraf gaan en die er mogelijk op volgen en van de eventuele alternatieve toestanden, die bij (een andere) behandeling bereikt kunnen worden? Is het reëel om elk gezond levensjaar even zwaar te wegen, onafhankelijk van de vraag of het het laatste levensjaar is of het eerste in een lange reeks? Zijn QALY’s van verschillende personen zodanig vergelijkbaar, dat ze bij elkaar kunnen worden opgeteld als het erom gaat de opbrengst van een medische voorziening te bepalen bij een groep personen? En als deze veronderstellingen niet, niet in alle omstandigheden, of niet volledig opgaan, wat zijn daarvan dan de consequenties voor de interpretatie van gezondheidswinst gemeten in QALY’s?

Met deze vragen is in de gezondheidseconomische en daarmee samenhangende besliskundige en ethische literatuur, vanaf de introductie van de term QALY in de jaren zeventig (Ber73, p 196, Wei77, Zec76) tot op de dag van vandaag, onafgebroken geworsteld. Onder de gezondheidseconomische auteurs die zich ermee ingelaten hebben, zijn globaal twee stromingen te onderscheiden, gekenmerkt door twee verschillende benaderingen van het QALY-concept: de welvaartseconomische (*welfarist*) en de extra-welvaartseconomische of besluitvormersbenadering (*extra-welfarist* of *decision maker’s approach*) (Ble97a, Bro00a, Bus00, Cul98, Hur98, Joh96, Wei97b). De welvaartseconomische benadering streeft naar een zo goed mogelijke inbedding van het QALY-concept in de economische welvaartstheorie en probeert op die

* Het feit dat althans sommige mensen sommige gezondheidstoestanden ervaren als erger dan de dood heeft in de QALY-methodologie tot nu toe weinig aandacht gekregen (Hou01, Mac01).

manier duidelijkheid te scheppen over de interpretatie van de QALY onder verschillende omstandigheden. In deze visie dienen QALY's een valide afspiegeling te zijn van de waardeoordelen van de individuele personen waar het in de gezondheidszorg uiteindelijk om gaat. De besluitvormersbenadering daarentegen gaat ervan uit dat degenen die moeten beslissen over de verdeling van een beperkt gezondheidszorgbudget zelf moeten bepalen welke 'doelstellingsfunctie' zij met deze verdeling trachten te maximaliseren. In de regel wordt aangenomen dat de doelstellingsfunctie van de besluitvormers een operationalisatie van het criterium 'gezondheid' zal zijn en dat 'de som van de geproduceerde QALY's' daar de op dit moment best denkbare invulling van is.

Beide benaderingen hebben hun problemen. De welvaartseconomische benadering stelt zichzelf tot taak om aan te tonen dat de QALY, onder bepaalde voorwaarden, op te vatten is als een maat voor gezondheidsgerelateerd 'nut' of 'utiliteit'. De gedachte is dat mensen erop uit zijn hun individuele 'utiliteit' te maximaliseren, dat de maatschappelijke utiliteitsfunctie een aggregaat dient te zijn van de individuele utiliteiten en dat het doel van de verdeling van schaarse middelen zou moeten zijn die maatschappelijke utiliteitsfunctie te maximaliseren (Gar96). Het begrip 'utiliteit', waaraan in de filosofische en economische literatuur uiteenlopende betekenissen worden toegekend (Sen91), wordt in de welvaartseconomisch georiënteerde gezondheidseconomische literatuur in het algemeen gebruikt in de zin van de *expected utility theory*, waarvoor de fundamenteën in 1944 gelegd werden door Von Neumann en Morgenstern (Neu44). De aan deze theorie ontleende *standard gamble*-methode* gold lange tijd onder gezondheidseconomen als de theoretisch meest zuivere manier om individuele gezondheidsgerelateerde utiliteit te meten. De laatste jaren wordt echter het belang van de *expected utility theory* als theoretische basis van de QALY door sommige gezondheidseconomische onderzoekers bestreden. Zij erkennen de waarde ervan als *normatieve* theorie, maar betogen dat de uit het werk van onder meer Kahneman en Tversky bekende *descriptieve* tekorten (Kah79), de *expected utility theory* diskwalificeren als theoretisch fundament voor de descriptieve activiteit die het afleiden van voorkeuren voor gezondheidstoestanden uit het keuzegedrag van mensen hun inziens is. Een *nonexpected utility theory* als de *rank-dependent utility theory* is volgens hen meer in overeenstemming met de voorkeuren die mensen ten aanzien van gezondheidstoestanden tentoonspreiden en biedt een beter uitgangspunt voor de ontwikkeling van instrumenten om die voorkeuren te meten (Ble97b, Ble02).

* Zie de eerste voetnoot bij paragraaf 3.3.3.

In 1980 formuleerden Pliskin en collega's drie voorwaarden waaraan voldaan moet zijn om de QALY als eenheid van utiliteit te kunnen interpreteren (Pli80, Wei80): *

- de waardering van de kwaliteit en die van de duur van het leven moeten wederzijds onafhankelijk zijn (*utility independence*)
- de proportie van de resterende levensduur die men bereid is op te geven ten behoeve van een bepaalde verhoging van de kwaliteit moet onafhankelijk zijn van de lengte van die resterende levensduur (*constant proportional trade-off*)
- er moet sprake zijn van een risiconutraliteit ten aanzien van levensduur (*risk neutrality on life years*)**: verondersteld wordt dat bijvoorbeeld de keuze tussen, enerzijds, de zekerheid van vijf levensjaren met een utiliteit van 0,5 en, anderzijds, deelname aan een loterij met 50% kans op onmiddellijk overlijden en 50% kans op vijf geheel gezonde levensjaren, iemand om het even is (Boe98).***

Nadien wordt vastgesteld dat bovendien, in het in de praktijk veel voorkomende geval van een gezondheidstoestand die varieert in de tijd, de waarde die aan de gezondheidstoestand in een bepaalde periode wordt toegekend, onafhankelijk moet zijn van de gezondheidstoestanden in alle andere perioden (*additive utility independence*, ook wel *additive separability* genoemd) (Ble95, Joh96, Dol00). Dolan onderscheidt daarnaast nog de voorwaarden dat waarderingen van gezondheidstoestanden stabiel dienen te zijn gedurende iemands hele leven (*stability of lifetime preferences*) en onafhankelijk van het moment, dichtbij of verder weg in de toekomst, waarop ze zich voordoen (*zero rate of time preference*) (Dol00).

Dit zijn geen geringe voorwaarden. Uiteenlopende auteurs beargumenteren dat er in de praktijk in het algemeen niet aan voldaan zal worden (Ble95, Boe98, Bro93, Car89, Dru97, Gaf95, Gar97, Goo96, Gui99, Hey90, Loo89, Nor94, Tor89). Uit empirisch onderzoek komen ook tal van aanwijzingen voort die daarop wijzen. De waardering van

* Eveneens in 1980 onderscheidde Alan Williams op het World Congress on Health Economics een zevental, tot op zekere hoogte vergelijkbare, aannamen: "very stringent conditions, which make empirical work to elicit such valuations rather difficult. They are needed in order to avoid the logistically impossible alternative, which is to present subjects with all possible future time profiles of health states, and then to get them to value each profile as a whole relatively to each other profile as a whole." Williams' 'logistiek onmogelijke alternatief' wordt later, ietwat gemodificeerd, gepropageerd door Mehrez en Gafni, in de vorm van het verderop in deze paragraaf beschreven *Healthy Year Equivalent*.

** Bleichrodt en collega's toonden later aan dat de derde voorwaarde, die van de *risk neutrality*, de andere twee voorwaarden *impliceert* onder de - alleszins aannemelijke - voorwaarde dat voor een duur van nul levensjaren alle gezondheidstoestanden equivalent zijn. Zij benadrukten dat deze conclusie niet als *verdediging* van het QALY-concept moet worden opgevat. En inderdaad: of *utility independence* en *constant proportional trade-off* nu als zelfstandige voorwaarden worden gezien, of als inbegrepen in de aanname van *risk neutrality*, voor de plausibiliteit van de drie condities tezamen lijkt dat weinig uit te maken.

*** In theorie is het mogelijk om de berekening van QALY's aan te passen aan niet-neutrale risico-attitudes, maar dat vergt een – lastige – schatting van de relevante risico-attitudeparameter (Kra98b, Miy85). In de praktijk wordt die weg nauwelijks bewandeld (Joh96).

een gezondheidstoestand blijkt afhankelijk van de tijdsduur die iemand in die toestand doorbrengt (Kir00, Sac78), van iemands vooruitzicht op verbetering of juist verslechtering van zijn of haar toestand (Kra98a, Ric96), van de prognose op de lange termijn en de doodsoorzaak (Hal92). Dolan geeft een overzicht van een groot aantal onderzoeken die één of meer van de zes genoemde voorwaarden op hun houdbaarheid hebben onderzocht en stelt vast dat voor geen ervan duidelijke empirische ondersteuning te vinden is (Dol00). Vooral de aanname van de *additive separability* lijkt door de beschikbare informatie op overtuigende wijze te worden weersproken, in overeenstemming met het betoog van Broome, die temidden van diverse “heroic assumptions” precies deze voorwaarde kwalificeert als “the most dubious condition” (Bro93). De voorlopige slotsom van de welvaartseconomische discussie over de QALY lijkt een adequate samenvatting te vinden in het boek van Drummond en collega’s, waar zij hun vraag “Is a QALY a utility” beantwoorden met een kernachtig “In general, it is not” (Dru97).

Om dit fundamentele probleem het hoofd te bieden, introduceerden Mehrez en Gafni in 1989 het *Healthy Year Equivalent* (HYE) (Meh89). Anders dan bij de QALY, die gebaseerd is op een waardering van gezondheidstoestanden zonder gespecificeerde duur, wordt bij de meting van gezondheidsgerelateerde utiliteit in HYE’s aan respondenten gevraagd de mate van wenselijkheid te bepalen van een totaal *lifetime health profile*, dat wil zeggen: van een specifieke volgorde van gezondheidstoestanden, elk met een bepaalde duur, tot aan het overlijden. Ook de bepaling van de utiliteit in termen van HYE’s kan niet zonder de aanname van risiconutraliteit ten aanzien van (gezonde) levensduur (Dol00, Rie98). Voor het overige zijn echter de onderliggende aannames van de HYE veel minder restrictief dan die van de QALY. De theoretische superioriteit van het waarderen van *lifetime health profiles* in plaats van geïsoleerde gezondheidstoestanden wordt in de literatuur dan ook vrij algemeen erkend (Ble95, Bon92, Bra99, Bus00, Dol00, Dru97, Gol96a, Rie98). Even algemeen echter is de vaststelling dat de praktische uitvoerbaarheid van de HYE-benadering twijfelachtig is. Het waarderen van één opeenvolging van diverse gezondheidstoestanden is al lastig, het waarderen van een uitkomstscenario dat een groot aantal verschillende *lifetime health profiles* omvat, is een immense opgave. Het noopt tot uitgebreide interviews en stelt hoge cognitieve eisen aan de respondenten, met alle problemen van dien op het vlak van de kosten, het aantrekken van beoogde respondenten en de betrouwbaarheid en validiteit van de verkregen uitkomsten (Bon92, Dru97, Joh96, Ric96). De HYE-benadering is in empirisch onderzoek dan ook weinig toegepast (Lle02).

Sommige welvaartseconomisch georiënteerde gezondheidseconomen suggereren dat de aannames die aan de opvatting van de QALY als eenheid van utiliteit ten grondslag liggen in de praktijk misschien niet helemaal opgaan, maar wel bij benadering en voldoende om het gebruik van QALY’s in KUA’s te rechtvaardigen (Gar96). De

(gedeeltelijke) empirische toetsing van deze claim die in principe mogelijk is door een vergelijking van de QALY en de HYE-benadering is echter tot op heden achterwege gebleven. Studies naar de vraag in welke mate de resultaten van de beter uitvoerbare kwantificering van gezondheidswinst in QALY's in de praktijk afwijken van die van de theoretisch gezien superieure kwantificering in HYE's, zijn nauwelijks voorhanden (Bra99).

Stuit aldus de interpretatie van de QALY als eenheid van individuele utiliteit op ernstige problemen, ook de aggregatie van individuele gezondheidgerelateerde utiliteiten tot een waardebeoordeling op maatschappelijk niveau, die het sluitstuk vormt van de welvaartseconomische benadering van de QALY, is niet onproblematisch. Of en onder welke condities het aggregeren van individuele utiliteiten in een sociale utiliteitsfunctie mogelijk is, is de inzet van een intensief debat in de welvaartseconomie, dat ook zijn weerslag heeft in de gezondheidseconomie (Fro89c, Gar96, Sas01, Tor86, Wag91). Sommige auteurs stellen dat aggregatie van QALY's onder niet al te beperkende voorwaarden mogelijk is (Ble97a), anderen menen dat de noodzakelijke interpersoonlijke vergelijkbaarheid van QALY's onvoldoende is te waarborgen (Bro00a, Ube00c). Dat een niet te verwaarlozen aantal mensen de toestand 'dood' hoger waardeert dan sommige gezondheidstoestanden is een gegeven dat de aggregatie van waarderungen van gezondheidstoestanden verder compliceert (Bon01a, Mac01). Het debat duurt voort.

Gezien de hardnekkige theoretische en empirische problemen waarmee de welvaartseconomische benadering van de QALY worstelt, is het niet verbazingwekkend dat de extra-welvaartseconomische of besluitvormersbenadering, die geen fundering in individuele utiliteiten nastreeft en, althans in de ogen van sommige auteurs, een meer pragmatische, *common sense* aanpak toestaat (Bro00a), de laatste jaren een groeiende populariteit geniet (Bus00). De extra-welvaartseconomische zienswijze veronderstelt dat besluitvormers bij de verdeling van middelen in de gezondheidszorg streven naar 'maximalisatie van gezondheid' en dat 'maximalisatie van de som van de geproduceerde QALY's' een voor hen aanvaardbare invulling van die doelstelling is (Bus00, Dru97, Wag91). Tegelijkertijd zien de aanhangers van deze benadering onder ogen dat 'maximalisatie van QALY's' heel goed slechts één van de doelstellingen van de besluitvormers kan zijn en dat die besluitvormers vaak ook andere overwegingen in hun besluitvorming betrekken, zoals verdelende rechtvaardigheid, mate van afhankelijkheid, recht op privacy (Bus00, Rus96, Wei97b). Zij benadrukken – en claimen daarmee “the prevailing view of cost-effectiveness analysis” tot uitdrukking te brengen – dat de KUA beschouwd moet worden als “an aid to decision making, not a complete prescription for social choice” (Rus96, Wei97b).

Is de extra-welvaartseconomische opvatting van de QALY een alternatief voor de welvaartseconomische? Waarom zou de QALY als maat voor gezondheid slagen waar

de QALY als maat voor utiliteit faalt? Men hoeft zich slechts af te vragen wat men zich voor moet stellen bij de vermenigvuldiging van levensduur met gezondheid, om in te zien dat een kritische bevraging van de QALY als eenheid voor het meten van de ‘hoeveelheid gezondheid’ onvermijdelijk terugvoert naar alle complexiteit die het begrip van de QALY als utiliteit omgeeft. Het kan dan ook nauwelijks verbazing wekken dat het in de praktijk van het KU-onderzoek niet blijkt uit te maken of men welvaartseconomische of extra-welvaartseconomische QALY’s kwantificeert. “[...] there is nothing observable about the methods themselves to indicate whether the analysis is intended to be welfarist or extra-welfarist”, constateert Hurley. “Nothing observable distinguishes which approach is being invoked” (Hur00). In de praktijk blijken alle onderzoekers te werk te gaan alsof hun interpretatie van de QALY een welvaartseconomische is. Zij trachten de voorkeuren die individuele patiënten en mensen uit de algemene bevolking hechten aan gezondheidstoestanden, zo goed mogelijk te meten en gebruiken de gevonden waarden als factoren voor de weging van de levensduur. Een groot deel van het methodologisch onderzoek is erop gericht technieken te ontwikkelen om die individuele voorkeuren adequaat te kwantificeren. Impliciet lijkt de meerderheid van KU-onderzoekers de welvaartseconomische interpretatie van de QALY te ondersteunen (Ble96). Diezelfde meerderheid legt echter tegelijkertijd, conform “the prevailing view of cost-effectiveness analysis”, de verantwoordelijkheid voor de *weging* van de aldus geïnterpreteerde QALY’s in het besluitvormingsproces bij de besluitvormers.

En wat moeten die daarvan vinden? Hurley merkt op dat, ironisch genoeg, de pogingen van welvaartseconomisch en extra-welvaartseconomisch georiënteerde KU-onderzoekers om helderheid te scheppen over aannamen en betekenis van QALY’s, de verwarring bij de gemiddelde potentiële gebruiker van KUA-uitkomsten alleen maar hebben doen groeien. Hij concludeert, met een parafrase op een bekend gezondheidseconomisch adagium (Wei88, Wil92): “It turns out that a QALY is *not* just a QALY”. Soms blijkt de QALY een maat voor ‘utiliteit’, dan weer een maat voor ‘gezondheid’ te zijn: “the interpretation is in the eyes of the beholder, depending on what assumptions one is willing to make” (Hur00).

De KU-onderzoekers lijken zich intussen te laten leiden door de door Williams welsprekend verwoorde overtuiging dat voor de QALY nu eenmaal geen beter alternatief voorhanden is:

“[...] there is no perfect system on offer, and we can’t wait. As with a well-conducted clinical trial, the new has to be compared systematically, according to preselected criteria, with what already exists. This is what needs to happen in the field of priority setting. If the same criteria as are used to criticise the QALY approach, were used *in an even handed way* (*cursivering Williams, NdN*) to criticise current practice, or any feasible alternative to it, how would these alternatives make out?” (Wil96, p 1801)

Interessant genoeg demonstreert dit citaat niet alleen hoe althans deze prominente gezondheidseconoom denkt over de huidige praktijk van het stellen van prioriteiten in de gezondheidszorg, maar ook hoe hij, zijn extra-welvaartseconomische betogen ten spijt (Sug78, Wil93), de *QALY approach* opvat: wel degelijk als een alternatief voor de huidige praktijk van *priority setting*, en niet als een bescheiden ondersteuning daarvan.

Het woord is aan de *decision makers*. Zij zullen primair moeten varen op hun eigen inzicht en intuïtie. Empirisch onderzoek dat licht zou kunnen werpen op het belang van schendingen van aannamen van QALY's in verschillende omstandigheden, is niet of nauwelijks voorhanden (Bra99). Eén van de vragen die zij voor zichzelf moeten beantwoorden is of het wel verstandig is om het verschil tussen leven en dood én verschillen in waarde van gezondheidstoestanden in één uitkomstmaat te willen vangen. Anand en Hanson schrijven, in een kritische analyse van de DALY, die grotendeels ook op de QALY van toepassing is:

“While having an indicator that combines states of imperfect health with death is clearly convenient, there is an obvious information loss in reducing death to simply another health state. Some will argue that the events are incommensurable, and that a lexical priority attaches to life over death. At any rate, this suggests that information about mortality and morbidity should be presented separately – even if trade-offs were conceded between the two events.” (Ana97)

3.2 Beschrijving van gezondheidstoestanden

Om de opbrengst van een medische behandeling in QALY's te kunnen meten, moet de gezondheidgerelateerde kwaliteit van leven gekwantificeerd worden. Dat gebeurt in het algemeen in twee fasen. Ten eerste wordt patiëntgebonden onderzoek uitgevoerd, dat uitmondt in een *beschrijving* van de gezondheidstoestanden van de personen die de behandeling in kwestie hebben ondergaan. Ten tweede worden de beschrijvingen van de gezondheidstoestanden die bij de behandeling in het geding zijn, *gewaardeerd* op een intervalschaal* van 0 tot 1. Het is in principe mogelijk om de kwaliteitscorrectie-factor die nodig is voor het berekenen van QALY's, in het patiëntgebonden onderzoek direct te bepalen, dus zonder de tussenstap van de beschrijving van gezondheidstoestanden. Aan de in het onderzoek betrokken patiënten wordt in dat geval gevraagd een waardeoordeel uit te spreken over hun eigen gezondheidstoestand(en). Men spreekt van een directe utiliteitsbepaling (*direct utility assessment*). Deze benadering wordt echter in de praktijk bijzonder weinig toegepast, omdat de meeste onderzoekers ervan uitgaan dat in KUA's

* Een intervalschaal is een schaal die zodanig geconstrueerd is dat aan de intervallen tussen de waarden op die schaal een kwantitatieve betekenis kan worden gehecht (bijv. "het verschil tussen A en B is twee keer zo groot als het verschil tussen A en C"). Omdat geen absoluut nulpunt is gedefinieerd, hebben de onderlinge verhoudingen van de meetwaarden zelf echter geen betekenis ("A is twee keer zo groot als B" is een betekenisloze uitspraak).

niet de patiënt zelf, maar (een steekproef uit) de algemene bevolking de meest geëigende bron van waarderingen van gezondheidstoestanden vormt (zie 2.4.2), en ook omdat klinici een directe utiliteitsbepaling vaak als een ontoelaatbare belasting van patiënten ervaren (Bra99).

De wijze waarop de effecten van medische behandelingen op de gezondheidstoestand van patiënten onderzocht moeten worden, is het onderwerp van een intensieve discussie in de medisch-wetenschappelijke literatuur, waarvoor de eerste aanzetten werden gegeven aan het eind van de jaren veertig van de vorige eeuw (Fei87, Kar48, Kar49, Nee91). Sindsdien zijn, vooral gedurende de afgelopen twintig jaar, vele honderden instrumenten ontwikkeld voor de meting van de 'gezondheidstoestand', de 'gezondheidgerelateerde kwaliteit van leven' of aspecten daarvan (Spi90, Spi96). Die meetinstrumenten hebben meestal de vorm van een lijst van vragen over aspecten van 'gezondheid' of 'gezondheidgerelateerde kwaliteit van leven'. De vragenlijst kan door een patiënt zelf of door een interviewer worden ingevuld.

Globaal zijn vier types van meetinstrumenten te onderscheiden. Allereerst zijn er de domeinspecifieke instrumenten (Ess00). Dit zijn instrumenten die één domein van de gezondheidstoestand bestrijken. Ze kunnen bij verschillende ziekten worden toegepast. Voorbeelden zijn de *index of independence in Activities of Daily Living* (ADL-index; Kat63) en de *McGill Pain Questionnaire* (Mel75).

Dan zijn er de ziektespecifieke instrumenten. In een overzicht van 154 gezondheidseconomische evaluatie-onderzoeken was dit type meetinstrument het meest gebruikte (Bra99). Deze instrumenten zijn in het algemeen ontwikkeld voor medisch onderzoek ter evaluatie van behandelingen voor één bepaalde ziekte. Ze richten zich op die dimensies van het begrip gezondheid die door die specifieke ziekte worden aangedaan. Voorbeelden zijn de *Health Assessment Questionnaire*, ontwikkeld voor onderzoek bij mensen met reumatische aandoeningen (Fri80), en de *EORTC QLQ-C30*, de *core questionnaire* van de European Organization for Research and Treatment of Cancer, die voor specifieke vormen van kanker met aparte modulen kan worden aangevuld (Aar00).

De derde categorie omvat de instrumenten die beogen het begrip gezondheid in het algemeen te operationaliseren: de algemene gezondheidstoestand-instrumenten. Ook deze op psychometrische leest geschoeide instrumenten worden in economisch evaluatie-onderzoek veelvuldig toegepast (Bra99). Ze bestaan uit een groot aantal vragen ('items'), die ieder een bepaald aspect van het complexe begrip gezondheid vertegenwoordigen. De scores op de afzonderlijke items kunnen bij elkaar worden opgeteld tot scores voor de belangrijkste 'dimensies' van gezondheid (bijvoorbeeld de fysieke, de psychische en de sociale dimensie). Die laatste kunnen op hun beurt, althans bij sommige instrumenten, weer opgeteld worden tot een totaalscore voor de gezondheid als geheel. De item- en dimensiescores kunnen bij de optelling verschillend gewogen

worden. Voor het bepalen van de weegfactoren worden verschillende methoden toegepast, bijvoorbeeld: empirisch onderzoek naar de statistische correlatie tussen de scores op bij elkaar behorende items bij de *SF-36* (War96); waardering van elk item op een 15-puntsschaal van minimaal tot maximaal ‘disfunctioneel’ door een panel van beoordelaars bij de *Sickness Impact Profile* (Ber76); panelbeoordeling van de ‘ernst’ van de items bij de *Nottingham Health Profile* (Hun80). Hoewel bij de presentatie van de meetresultaten dus onvermijdelijk een vorm van weging van afzonderlijke itemscores in het geding is, worden de algemene gezondheidstoestand-instrumenten in het algemeen opgevat als beschrijvende instrumenten (Ess95).

In dat opzicht onderscheiden ze zich van de vierde en laatste categorie instrumenten, die van de gezondheidstoestand-classificatiesystemen (*preference classification systems* (Gol96a), *multi-attribute health status classification systems* (Dru97), *multi-attribute utility scales* (Bra99)). Deze instrumenten belichamen zowel de beschrijvende als de waarderende fase van het onderzoek dat nodig is voor een schatting van de kwaliteitscorrectiefactor van de QALY. Het waarderende element komt aan de orde in de volgende paragraaf. Hier volstaat de vaststelling dat de meeste gezondheidstoestand-classificatiesystemen gebaseerd zijn op een classificatie van gezondheid in een, in vergelijking met de algemene gezondheidstoestand-instrumenten, relatief klein aantal discrete toestanden. Voorbeelden zijn de *Health Utilities Index*, die in zijn derde en meest uitgebreide versie 972 000 verschillende toestanden onderscheidt (Fee96) en de *EuroQuol* (Eur90), die in de *EQ-5D* versie 243 toestanden omvat (Kin96). Ter vergelijking: de *SF-36*, één van de meest toegepaste algemene gezondheidstoestand-instrumenten, beschrijft meer dan 10 miljoen mogelijke gezondheidstoestanden (Dol00). Hoewel gezondheidstoestand-classificatiesystemen al zo’n vijftwintig jaar beschikbaar zijn, worden ze, getuige onder meer een overzicht van tot en met 1997 gepubliceerde KUA’s, in economisch evaluatie-onderzoek relatief weinig toegepast (Bra99, Neu97, Neu00a).

In de literatuur bestaat in grote lijnen overeenstemming over de voor- en nadelen van deze verschillende types meetinstrumenten als het gaat om toepassing in KUA’s. De ziekte- en domeinspecifieke vragenlijsten zijn gevoelige instrumenten voor het meten van de verschijnselen waarop ze zijn toegesneden (Ess00, Gol96a, Guy96, Pat89). Een evident nadeel bij gebruik in economisch evaluatie-onderzoek is dat ze geen vergelijking toestaan van de effecten van behandelingen voor ziekten die verschillende dimensies van de gezondheid beïnvloeden. De algemene gezondheidstoestand-instrumenten daarentegen maken een beschrijving van gezondheidseffecten mogelijk in termen die toepasbaar zijn op veel verschillende ziekten, zij het ten koste van een geringere gevoeligheid voor sommige subtiele of ziektespecifieke effecten en veranderingen (Ber89, Edg98, Guy96, Pat89, Ver01). Voor zowel de ziekte- en domeinspecifieke als de algemene gezondheidstoestand-instrumenten geldt dat nog

onduidelijk is hoe de waardering van de resulterende gezondheidstoestand-beschrijvingen, die noodzakelijk is om tot de berekening van QALY's te kunnen komen, tot een goed einde kan worden gebracht (Bra99, Dru97). De gezondheidstoestand-classificatiesystemen zijn precies ontwikkeld om voor dat laatste probleem een oplossing te bieden, maar schieten, meer nog dan de algemene gezondheidstoestand-instrumenten, tekort waar het gaat om het detecteren van subtiele en ziektespecifieke verschijnselen en veranderingen daarin (Bra99, Bus00, Don88, Gol96a, Guy96, Hol95, McK98, Neu00a, Tor99, Ver01).

Het ideale type meetinstrument dat breed zou kunnen worden ingezet om over alle ziektecategorieën heen de *vergelijkbaarheid* van de gemeten gezondheidswinst te kunnen garanderen, is nog niet voorhanden: daarover zijn alle auteurs het eens. Er worden verschillende oplossingen aangedragen. De gezondheidstoestand-classificatiesystemen vormen daarvan een vast onderdeel, maar ook voor de ziektespecifieke en de algemene gezondheidstoestand-instrumenten lijkt nog een rol te zijn weggelegd. Het is vooral de wijze waarop de uitkomsten van deze verschillende instrumenten in één maat moeten worden samengevat, waarover in de literatuur onduidelijkheid blijft bestaan.

De auteurs van de Canadese richtlijnen voor farmaco-economisch onderzoek uit 1997 (CCO97) – door het College voor zorgverzekeringen tot uitgangspunt gekozen bij de ontwikkeling van de Nederlandse richtlijnen (CVZ99) – stellen dat het onwaarschijnlijk is dat ooit één instrument de status van ‘gouden standaard’ voor de meting van de gezondheidsgerelateerde kwaliteit van leven zal bereiken. Zij bevelen farmaco-economische onderzoekers aan om zo mogelijk uit elk van de categorieën instrumenten – de ziekte- en domeinspecifieke, de algemene gezondheidstoestand-instrumenten en de classificatiesystemen – één instrument te selecteren, op grond van de diverse onderzoek-specifieke overwegingen die bij zo'n keuze in het geding kunnen zijn. Ze laten daarbij in het midden hoe de resultaten van deze metingen kunnen worden omgezet in wegingsfactoren voor het berekenen van QALY's die voor verschillende onderzoeken vergelijkbaar zijn (CCO97).

Rotterdamse onderzoekers bevelen aan om voor het berekenen van QALY's een beroep te doen op een van de gezondheidstoestand-classificatiesystemen, maar om daarnaast in economisch evaluatie-onderzoek steeds ook een specifiek en een algemeen gezondheidstoestand-instrument op te nemen. Zij gaan, net als de Canadese richtlijnen, niet in op de vraag of en, zo ja, hoe de met die laatste instrumenten te verkrijgen meetresultaten verdisconteerd zouden kunnen worden in de berekening van over verschillende onderzoeken heen vergelijkbare QALY's (Bus00, Ess00).

Het Amerikaanse *Panel on Cost-Effectiveness in Health and Medicine*, het zogenaamde Washington Panel, constateert dat “the field of measurement and valuation of outcomes is still a developing one”. Het bepleit de ontwikkeling van één algemeen

toepasbaar gezondheidstoestand-classificatiesysteem dat beter is dan de al bestaande. Tegelijk stelt het panel vast dat ook een ideaal instrument de ziektespecifieke instrumenten waarschijnlijk niet overbodig zal maken. Het beschrijft in algemene termen hoe de resultaten van een kleine subcategorie van de ziektespecifieke instrumenten, de ziektespecifieke classificatiesystemen, geprojecteerd zouden kunnen worden op de dimensies van een algemeen classificatiesysteem, om berekening van QALY's mogelijk te maken (Gol96a).

De opstellers van een uitgebreid literatuuroverzicht dat drie jaar na het boek van het Washington Panel verschijnt, laten zien dat de methoden om de resultaten van ziektespecifieke en algemene gezondheidstoestand-instrumenten op een economisch valide wijze te waarderen, nog in de kinderschoenen staan. Ze tonen zich weinig optimistisch over de vraag of ze in de praktijk ooit werkbaar zullen blijken (Bra99). De schrijver van het hoofdstuk over de meting van de gezondheidgerelateerde kwaliteit van leven in een gezondheidseconomisch handboek is resoluut in zijn conclusie: noch de ziektespecifieke, noch de algemene gezondheidstoestand-instrumenten zijn geschikt voor toepassing in KUA's. Wanneer het erom gaat de uitkomsten voor uiteenlopende ziekten en behandelingen met elkaar te kunnen vergelijken, dan zijn alleen de gezondheidstoestand-classificatiesystemen bruikbaar (Dol00).

Het is duidelijk dat hierover het laatste woord niet is gezegd. Voorlopig zit de discussie over de in KUA's te gebruiken instrumenten voor het meten van gezondheidstoestanden echter in een lastig parket. Enerzijds bestaat brede overeenstemming over het feit dat ziektespecifieke instrumenten onmisbaar zijn om relevante behandelingseffecten niet te missen en op waarde te kunnen schatten. Anderzijds wordt juist groot belang gehecht aan de algemeen toepasbare, maar voor specifieke effecten betrekkelijk ongevoelige gezondheidstoestand-classificatiesystemen, omdat het genereren van waarderingen van gezondheidstoestanden die breed vergelijkbaar zijn, alleen daarmee uitvoerbaar lijkt.

De praktijk van de KUA's lijkt zich intussen aan de groeiende overeenstemming over het belang van de gezondheidstoestand-classificatiesystemen nog weinig gelegen te laten liggen. De meeste tussen 1975 en 1995 gepubliceerde onderzoeken maken gebruik van ziektespecifieke instrumenten en ontlenen de waarden die aan gezondheidstoestanden worden toegekend aan het oordeel van de onderzoekers (Neu97). Ook in tot en met 1997 gepubliceerde KUA's blijken de gezondheidstoestand-classificatiesystemen nog weinig te worden toegepast (Neu00a).

Hoe belangrijk is de discussie over de keuze van meetinstrumenten voor de besluitvormer die overweegt verdelingsbeslissingen – mede – op KUA's te baseren? De besluitvormer kan aan de literatuur weinig houvast ontlenen als het erom gaat de mogelijke kwantitatieve invloed op doelmatigheidsschattingen te bepalen van de keuze tussen een algemeen classificatiesysteem of een ziektespecifiek instrument. Eén

onderzoek, naar de kostenutiliteit van mammografische screening op borstkanker, wijst erop dat de keuze tussen ziektespecifieke instrumenten en een algemeen classificatiesysteem in elk geval van groot praktisch belang *kan* zijn. De kosten per gewonnen kwaliteitsgecorrigeerd levensjaar bleken in dat onderzoek, waarin gebruik gemaakt werd van ziektespecifieke beschrijvingen van gezondheidstoestanden, ruim twee keer zo hoog als die per niet-gecorrigeerd gewonnen levensjaar. Het feit dat in een ander onderzoek een dergelijk verschil niet werd gevonden, schrijven de onderzoekers toe aan het feit dat in dat onderzoek een meetinstrument werd toegepast dat ongevoelig was voor specifieke aantastingen van de gezondheid die geassocieerd zijn met de diagnose en behandeling van borstkanker, zoals bezorgdheid over de diagnose, pijnlijkheid, stijfheid en zwelling die het gevolg kunnen zijn van borstamputatie, angst voor terugkeer van de kanker en zorgen over het uiterlijk (Hal92).

Ook bij de keuze tussen de beschikbare classificatiesystemen biedt de literatuur weinig houvast. De schrijvers van een al eerder aangehaald literatuuroverzicht wijzen op de duidelijke verschillen in descriptieve inhoud tussen de classificatiesystemen onderling en stellen dat deze de keuze voor een bepaald systeem mede afhankelijk maken van de ziekte en de leeftijd van patiënten. Vergelijkend onderzoek naar de meetresultaten van de verschillende classificatiesystemen bij dezelfde patiëntenpopulaties is nog bijzonder schaars en lastig te interpreteren (Bra99). Zo vonden twee onderzoeksgroepen die de *Health Utilities Index* en het *EuroQol*-instrument vergeleken, *verschillende* overeenkomsten en verschillen (Bos00, Gli99).

De complexiteit waarvoor de besluitvormer zich gesteld ziet, wordt verder verhoogd door het probleem van de interculturele vergelijkbaarheid van de meetresultaten. Bijna alle vragenlijsten voor het meten van gezondheidstoestanden zijn ontwikkeld op basis van onderzoek waarin vrijwel uitsluitend vertegenwoordigers van de relatief goed opgeleide, blanke meerderheid van een westers land betrokken waren. Het ligt voor de hand om te veronderstellen dat die vragenlijsten – met hun specifieke selectie van gezondheidsaspecten en hun specifieke weging van de itemscores – bij toepassing in uiteenlopende andere (sub)culturen, nationaal en internationaal, niet zonder meer vergelijkbare onderzoeksresultaten zullen opleveren. Uit onderzoek blijkt dat ook (Hut96, Hun98a, Ste00). In verschillende culturen kunnen bijvoorbeeld verschillende gedachten gangbaar zijn over wat het betekent om zich goed, gezond of juist ziek te voelen en deze ideeën geven mede vorm aan de ervaring van pijn, ongemakken en beperkingen en aan het relatieve gewicht van verschillende symptomen. Belangrijk zijn verder de verschillen in vocabulaire als het gaat om persoonlijke ervaringen en de conventies die de communicatie over die gevoelens met andere mensen bepalen (Hun98a). Organisatoren van grote internationale *clinical trials* hebben het probleem van de interculturele (on)vergelijkbaarheid van gezondheidstoestandmetingen al jaren geleden gesignaleerd (Spi96). In de discussie over de methodologie van KUA's heeft het

echter nog weinig aandacht gekregen. Het Washington Panel, bijvoorbeeld, stelt wel vast dat de *voorkeuren* voor gezondheidstoestanden uitgesproken door de ene bevolkingsgroep niet zonder meer geldig zijn voor anders samengestelde groepen, maar gaat voorbij aan het vergelijkbare probleem bij de *meting* van de gezondheid (Gol96a). Ook in het literatuuroverzicht van Dolan komt het niet ter sprake (Dol00).

3.3 Waarderen van de kwaliteit van leven

3.3.1 Aard van de waardeoordelen

Koesteren mensen waardeoordelen over gezondheidstoestanden en, zo ja, wat is de aard van die oordelen? Het is een vraag die in de gezondheidseconomische literatuur slechts zelden wordt gesteld. Een leerzame uitzondering vormt het betoog van Dolan (Dol00). Dolan maakt in navolging van de psycholoog Fischhoff onderscheid tussen de *philosophy of articulated values* en de *philosophy of basic values* (Fis91). Gezondheidseconomen gaan ervan uit dat de voorkeuren voor gezondheidstoestanden die zij bij hun respondenten meten ook de werkelijke voorkeuren van deze mensen zijn. Hun vooronderstelling is dat de respondenten uitgesproken waardeoordelen hebben over de meest uiteenlopende gezondheidstoestanden. De gezondheidseconoom hoeft ze hun slechts te ontlokken door de juiste vragen te stellen.

Deze *philosophy of articulated values* staat op gespannen voet met de uitkomsten van tal van onderzoeken die erop wijzen dat subtiele veranderingen in de meetprocedure, die logisch gezien niets toe of af zouden moeten doen aan de waarde van de te beoordelen gezondheidstoestanden – bijvoorbeeld veranderingen in de aard van het keuzeprobleem dat de proefpersonen wordt voorgelegd, in de verwoording van de vragen en in de volgorde waarin goede en slechte gezondheidstoestanden de revue passeren – wel degelijk van invloed zijn op de voorkeuren van de respondenten.

Dit type bevindingen ondersteunt de *philosophy of basic values*, die stelt dat voorkeuren, in het leven van alledag, zo goed als in gezondheidseconomisch onderzoek, niet simpelweg van een inwendige moederlijst worden afgelezen, maar ad hoc geconstrueerd worden door de respondent, die daarbij optreedt als *adaptive decision maker* (Slo95). In die opvatting is de meetprocedure een belangrijke factor in de vorming van de voorkeuren die respondenten tot uitdrukking brengen. Los van de meting koesteren zij ten aanzien van gezondheidstoestanden slechts ongearticuleerde, ‘basale’ waarden.

Een tussenpositie tussen de *philosophy of articulated values* en de *philosophy of basic values* is denkbaar. Mensen zouden, in plaats van over voorgevormde, direct toegankelijke waardeoordelen over gezondheidstoestanden, kunnen beschikken over “stable values of moderate complexity” (Fis91). Deze waarden zouden nader

vormgegeven en aan het licht gebracht kunnen worden in een meetprocedure die een beroep doet op de reflexieve vermogens van de proefpersoon. Dolan spreekt zich uit vóór een dergelijke tussenpositie, die het mogelijk maakt er wel van uit te gaan dat de gemeten voorkeuren voor gezondheidstoestanden de ‘ware’ voorkeuren van de respondenten in kwestie weerspiegelen, maar die tegelijk noopt tot het ontwikkelen van complexere meetprocedures. Dat zullen in de meeste gevallen meer interactieve procedures zijn dan de nu gebruikelijke: meetprocedures die niet bestaan uit het simpelweg ‘aftappen’ van reeds aanwezige voorkeuren, maar die recht doen aan het feit dat de waardeoordelen van de respondenten tijdens de meting hun uiteindelijke vorm moeten krijgen (Shi97). Deze tussenpositie maakt een breed spectrum aan benaderingen mogelijk. Maar, benadrukt Dolan, aangezien de meeste mensen slechts weinig verschillende gezondheidstoestanden uit eigen ervaring kennen en over veel toestanden geen uitgekristalliseerde waardeoordelen hebben, zou het gezondheidseconomische onderzoek er goed aan doen een uitgangspunt te kiezen dat dichter bij de *philosophy of basic values* dan bij de *philosophy of articulated values* ligt (Dol00).

Tegen de achtergrond van Dolan’s uiteenzetting over de invloed van de meetprocedure op de gemeten voorkeuren, is het opmerkelijk hoe weinig gewicht er in de gezondheidseconomische literatuur gehecht is aan het onrealistische karakter van de keuzesituaties waarin respondenten geacht worden zich te verplaatsen bij de belangrijkste gezondheidseconomische procedures* voor het meten van de waarde van gezondheidstoestanden (Joh96). In sociaal-wetenschappelijke en medische literatuur is gewezen op het abstracte en artificiële karakter van de keuzesituatie waarin de respondenten zich moeten inleven. Om die reden zijn vraagtekens geplaatst bij de gedachte dat de kwantitatief gemeten voorkeuren er ook maar enige aanspraak op zouden kunnen maken valide afspiegelingen te zijn van de ongearticuleerde, veranderlijke en contextafhankelijke ideeën over de waarde van gezondheidstoestanden die mensen er in werkelijkheid op nahouden (Ash89, Car89, Lep97, Nee92). De gemeten waarderingen van gezondheidstoestanden, zo luidt de kritiek, zijn artefacten. Willen het *relevante* artefacten zijn, die de rol die hun bij de berekening van QALY’s is toegedacht geloofwaardig kunnen vervullen, dan zal op zijn minst uit de vragen aan de respondenten duidelijk moeten zijn voor welk doel hun antwoorden gebruikt zullen worden. “Without some such explicit link between the questions used to establish quality indices and the allocations generated by QALYs, QALYs will remain persistently suspicious” (Men90).**

* Zie voor een beschrijving van deze procedures de eerste voetnoot bij paragraaf 3.3.3.

** Over een vergelijkbaar probleem – de weging van gezondheidstoestanden door deskundigen, ten behoeve van de berekening van DALY’s – merken Anand en Hanson op (Ana97): “[...] the meaning attached to the different weighting of health states depends in an important way on the precise question that was asked of these experts. Their responses would also depend on understanding the use to which such estimates would be put.”

In de gezondheidseconomische literatuur heeft deze kritiek weinig gehoor gevonden. Iemand die zich er wel ontvankelijk voor betoont is de Noorse onderzoeker Nord. Hij lanceert in 1992 het *saved young life equivalent* (de SAVE) als alternatief voor de QALY (Nor92a). Zijn voorstel, in Nederland met waardering begroet (Leu93) en in recente gezondheidseconomische overzichtsteksten vaak genoemd als veelbelovend, maar nog niet ver genoeg uitgewerkt en onvoldoende in de praktijk beproefd (Bus00, CCO97, Dru97, Gol96a), is voor een deel ingegeven door de vaststelling dat de meetprocedures die gebruikt worden om de Q van de QALY te kwantificeren op geen enkele wijze beantwoorden aan de aard van de keuzesituaties die de bestaansredenen vormen voor KUA's (Nor92a, Nor94). Nord poneert het *saved young life equivalent* als meeteenheid voor de opbrengst van gezondheidszorg: men produceert één SAVE als men één jeugdig persoon van een wisse dood redt en diens gezondheid volledig weet te herstellen. Hij stelt voor om aan respondenten in gezondheidseconomisch onderzoek de vraag te stellen hoeveel uitkomsten x van een gegeven medische interventie equivalent zijn met één SAVE, waarbij x de opbrengst van die interventie voor één patiënt representeert, gedefinieerd in termen van de gezondheidstoestand voorafgaand aan de interventie, de verbetering van die gezondheidstoestand, de risico's van de interventie, de leeftijd van de patiënt en andere relevante variabelen. Als het antwoord tien is, dan is dus de waarde van uitkomst x gelijk aan ééntiende SAVE.

Nord's SAVE-procedure vat het gezondheidseconomisch waarderingsprobleem expliciet op als een probleem van de relatieve waardering van de gezondheidsbaten die mensen ondervinden van uiteenlopende gezondheidszorgvoorzieningen (Nor93, Nor94, Nor95, Pin97). Een hiermee samenhangend voordeel is dat de SAVE gebaseerd is op een directe waardering van specifieke *veranderingen* in de gezondheid, in plaats van, zoals de QALY, op waarderungen van gezondheidstoestanden, waaruit vervolgens de waarde van de overgang van de ene toestand naar de andere wordt afgeleid (Nor99, Ube00a). Bovendien ontsnapt de SAVE aan de in 3.1 besproken beperkende aannames die aan de QALY ten grondslag liggen én worden, althans in principe, een aantal verdelingsproblemen die de kosten/QALY-verhouding aankleven (zie hoofdstuk 6), tot een oplossing gebracht: de vraag hoe gezondheidsopbrengsten die toevallen aan jongeren en aan ouderen, aan ernstig zieken en aan bijna gezonden, aan enkelingen en aan velen, ten opzichte van elkaar gewaardeerd dienen te worden, maakt deel uit van het waarderingsprobleem dat aan de respondenten wordt voorgelegd.

Deze voordelen hebben hun prijs op het praktische vlak. De waarderingstaak waar de SAVE-procedure de respondenten voor plaatst is nog aanzienlijk complexer dan de opgave die voor het schatten van *Healthy Year Equivalents* (zie 3.1) moet worden volbracht. De voordelen in vergelijking met de QALY zijn echter van dien aard, dat het te betreuren is dat het idee van de SAVE tot op heden nauwelijks ingang heeft gevonden in het gezondheidseconomisch onderzoek – afgezien van enkele recente theoretische en

methodologische bijdragen gewijd aan de *person trade-off* benadering, die de procedurele kern vormt van het SAVE-voorstel (Gre01, Nor99, Ube99, Ube00b, Ube00c). Als het erom gaat de doelmatigheid van voorzieningen over de volle breedte van de gezondheidszorg op één dimensie te kwantificeren, lijkt het ontwikkelen van complexe, interactieve procedures – in overeenstemming met het betoog van Dolan en met het pleidooi van Edgar en collega's voor een discursief-ethische benadering van het waarderingsprobleem (Edg98) – waarin een relevant panel zich gaandeweg een oordeel vormt over de waarde van de opbrengst van gezondheidszorgvoorzieningen in termen van SAVE's, een benadering die grondige exploratie verdient.

De QALY's zoals die nu figureren in kostenutiliteitsonderzoek zijn, anders dan de SAVE's, gebaseerd op abstracte waarderingsexercities die geen relatie hebben met het beoogde gebruiksdoel van de ontlokte waardeoordelen. Dit werpt een schaduw over de toepasbaarheid van de resultaten van dit onderzoek.

3.3.2 Keuze van de beoordelaars

Zo weinig in de literatuur te doen is over de aard van de waardeoordelen over gezondheidstoestanden, zo veel wordt er gediscussieerd over de vraag aan wie ze ontlokt moeten worden. De twee meest besproken opties zijn: de patiënten die de ziekte en behandeling in kwestie zelf (hebben) ondergaan en de algemene bevolking. Andere mogelijke bronnen van gezondheidstoestand-waarderingen worden incidenteel genoemd, maar over het algemeen snel afgedaan. Dat geldt voor de onderzoekers die de KUA uitvoeren: in de ogen van gezaghebbende auteurs een nauwelijks serieus te nemen optie (Joh96), maar wel in een reeks tussen 1975 en 1995 gepubliceerde KUA's verreweg de meest frequente bron van kwaliteitscorrectiefactoren (Neu97). Het geldt ook voor de beroepsbeoefenaren in de gezondheidszorg. In een tijd waarin artsen vooral gewaardeerd worden als dragers van medisch-wetenschappelijke kennis en niet zozeer als morele autoriteiten, kan de gedachte dat zij waardeoordelen over de kwaliteit van leven van hun patiënten zouden moeten uitspreken, op weinig bijval rekenen.* Artsen, verpleegkundigen en andere hulpverleners beschikken weliswaar over inzicht in en ervaring met uiteenlopende ziekten en behandelingen, maar uit onderzoek blijkt dat hun oordeel over de (on)wenselijkheid van een toestand dikwijls niet spoort met dat van

* In strijd met de heersende opvattingen in de gezondseconomische en ethische literatuur over KUA's maakten Murray en collega's, bij het bepalen van de *disability weights* voor het berekenen van *disability-adjusted life years*, gebruik van panels van beroepsbeoefenaren in de gezondheidszorg. Murray verdedigt dit op praktische gronden: "To speed up the already demanding process, the protocol [voor het wegen van gezondheidstoestanden, NdN] has been designed to be used with health-care providers so that less time needs to be invested to describe each of the conditions. Health care providers are selected because of their knowledge, not because they have 'better' judgment" (Mur97). Ook de *Dutch Disability Weights* groep deed een beroep op medische deskundigen, vooral huisartsen (Sto00a).

patiënten (Fro89c). Zij kunnen bovendien om verschillende redenen, bewust of onbewust, geneigd zijn tot een overschatting van de ernst van ziekte-toestanden waar zij zelf professioneel bij betrokken zijn (Tor86, McK98).

Blijven over de patiënten en de algemene bevolking. Dolan heeft er terecht op gewezen dat dit geen elkaar uitsluitende categorieën zijn: onder de ‘gezonde’ bevolking zijn er velen die verschillende ziekten uit eigen ervaring kennen; onder het label ‘patiënten’ gaan mensen schuil die pas net ziek geworden zijn of al lang geleden (Dol99, Dol00). Waar het uiteindelijk om gaat, is een beslissing over hoe de voorkeuren van verschillende bevolkingsgroepen gewogen moeten worden (Loo89) – indien er al systematische verschillen in beoordeling van gezondheidstoestanden tussen die groepen aan de dag zouden treden. Zijn er zulke verschillen? Sommige onderzoeken laten zien dat mensen die zelf in een slechte gezondheidstoestand verkeren, hun eigen toestand hoger waarderen dan anderen (Boy90, Len99, Sac78); in andere onderzoeken is geen verschil gevonden. In 1989 is in een overzichtsartikel vastgesteld dat de beschikbare onderzoeksresultaten geen duidelijke conclusie toelaten, maar dat reële verschillen, door de geringe omvang van de meeste onderzoeken en de grote interindividuele variabiliteit, heel goed onopgemerkt kunnen zijn gebleven (Fro89c). Later is in een relatief groot onderzoek weer wel een duidelijk verschil gevonden: patiënten waarderen uiteenlopende gezondheidstoestanden hoger dan niet-patiënten (Dol96). Een recent overzicht van 38 onderzoeken bevestigt die observatie (Wit00). Verreweg de meeste auteurs gaan er dan ook van uit dat serieus rekening gehouden moet worden met de mogelijkheid dat patiënten hun eigen toestand hoger aanslaan dan mensen die er geen ervaring mee hebben. Vooral bij mensen met een chronische ziekte zou dit het geval kunnen zijn. Onbekendheid van gezonde personen met ziekte-toestanden, maar zeker ook succesvolle aanpassing van patiënten aan hun toestand zou dit verschil in waardering kunnen verklaren (Dol00, Gol96a, McK98, Men99).

Welke waardeoordelen moeten het zwaarst wegen, die van de patiënten of die van de algemene bevolking? Niemand bestrijdt dat alleen de patiënten zelf weten hoe het is om in een bepaalde ziekte-toestand te verkeren. Zij beschikken over een door niemand te evenaren direct inzicht in de gevoelens en beperkingen die met hun toestand gepaard gaan. Niet-patiënten hebben in vergelijking met patiënten een immense ‘informatieachterstand’, die ook met de meest subtiele beschrijvingen niet te overbruggen is. Maar hoe belangrijk dit, zeker praktisch gezien (Tor86), ook is, de cruciale principiële vraag is toch of het de *waarden* van patiënten zijn, of die van de algemene bevolking, die er toe doen bij de verdelingsbeslissingen waar het in KUA’s om draait.

Vóór de patiënten, of meer algemeen: de doelgroepen van gezondheidszorgvoorzieningen, pleit dat zij degenen zijn om wier welzijn het in de gezondheidszorg begonnen is. Het lijkt dan moeilijk verdedigbaar om juist aan hun

waarden voorbij te gaan. Dit antwoord stemt volgens sommige gezondheidseconomische auteurs het best overeen met de principes van de welvaartseconomie, die stelt dat verdelingsbeslissingen gebaseerd moeten zijn op de utiliteitsfuncties van degenen die van die beslissingen zouden kunnen profiteren (Bra99, Joh96).

Tegen deze benadering is ingebracht dat patiënten tot strategisch antwoordgedrag geneigd zouden kunnen zijn. Zij zouden in de verleiding kunnen komen de onwenselijkheid van hun toestand te overschatten, om zo de doelmatigheid van voorzieningen gericht op preventie of genezing van hun toestand hoger te doen lijken dan hij is (Loo89, Tor86). Anderen betwijfelen of respondenten in de praktijk in staat zijn om te voorzien hoe hun antwoorden de verdeling van middelen zou kunnen beïnvloeden. Bovendien suggereert het beschikbare empirisch onderzoek dat patiënten hun eigen gezondheidstoestand eerder over- dan onderwaarden (Joh96, Gol96a). Belangrijker is dan ook de tegenwerping dat patiënten, voor zover zij zich succesvol aan hun toestand hebben aangepast, ook wanneer zij geheel oprecht zijn, in hun antwoorden hun aangepaste waarden tot uiting zullen brengen. En het is de vraag of het juist deze *aangepaste* waarden zijn waarop de waardering van gezondheidstoestanden in KUA's gebaseerd zou moeten worden (Bro95, Dol00, Men02).

Degenen die pleiten voor de algemene bevolking als bron van gezondheidstoestandwaarderingen beantwoorden deze laatste vraag ontkennend. Zij stellen dat de verdeling van collectieve middelen in de gezondheidszorg een zaak is van de gehele maatschappij, waarin patiënten met een bepaalde ziekte maar een kleine minderheid vormen. In de verdelingsbeslissingen zullen vooral de waarden tot uitdrukking moeten komen van de grote meerderheid gezonde mensen die volledig bevoegd en in staat zijn om hier en nu, zonder te weten welke ziekten zich in de toekomst bij hen zullen voordoen, te bepalen in welke voorzieningen hun verzekeringspremies geïnvesteerd moeten worden (Bus00, CCO97, Gol96a, Rut98). Dat zij slechts gebrekkig op de hoogte zijn van de werkelijke aard van gezondheidstoestanden, doet in deze redenering niets af aan de geldigheid van hun waardeoordelen: onvolledige en onvolmaakte informatie is nu eenmaal een onontkoombaar *fact of life* (Loo89).

Tegen deze argumentatie kan worden ingebracht dat, al zou het volstrekt legitiem zijn dat de algemene bevolking de gezondheidstoestanden waardeert, dit nog niet betekent dat zij de waarden toekent die ethisch gezien de *juiste* zijn. Dolan stelt de vraag of het wel verdedigbaar is om aan het oordeel van iemand die aan een bepaalde ziekte-toestand nooit enige gedachte heeft besteed evenveel waarde te hechten als aan dat van iemand die de ziekte kent van nabij (Dol99). Gezonde mensen zonder ervaring met bepaalde ziekten of beperkingen kunnen zich mede laten leiden door stereotypen of anderszins misplaatste vooroordelen (Gol96a). Zo geven Froberg en Kane een paar voorbeelden van onderzoeksresultaten die erop wijzen dat de waarden die respondenten

aan gezondheidstoestanden toekennen afhangen van de vraag of die toestanden henzelf of hypothetische patiënten betreffen; in het laatste geval blijken ze afhankelijk van verschillende patiëntenkenmerken die met de gezondheidstoestand op zichzelf niets te maken hebben. Zij stellen vast dat “the results of these studies highlight the serious ethical considerations that arise when social preferences are used to make public policy decisions” (Fro89c).

Een ander tegenargument is dat het onverstandig zou zijn om géén rekening te houden met hoe mensen in het algemeen hun toestand beoordelen wanneer zich eenmaal een bepaalde ziekte of beperking bij hen heeft geopenbaard – ook wanneer men zonder voorbehoud de logica dat de verzekerde nu eenmaal het recht heeft te bepalen waarvoor hij zich verzekert, van toepassing acht op de verdeling van collectieve middelen in de gezondheidszorg (Bro95, Men02). Men wil zich immers nú verzekeren van gezondheidszorg waarmee men zich dán zo goed mogelijk geholpen weet.

De opstellers van gezaghebbende buitenlandse richtlijnen voor gezondheidseconomisch onderzoek, zoals die van het Washington Panel en de Canadese richtlijnen voor farmaco-economisch onderzoek, en ook die van de richtlijnen van het College voor zorgverzekeringen en het instituut voor Medical Technology Assessment, nemen in deze discussie een duidelijk standpunt in. Zij stellen onomwonden dat waarderings van gezondheidstoestanden afkomstig moeten zijn van steekproeven uit de algemene bevolking (Bus00, CCO97, CVZ99, Gol96a). Bij de ontwikkeling van de in 2.3 besproken gezondheidstoestand-classificatiesystemen is deze richtlijn ook in praktijk gebracht (Bra99). De schrijvers van recente literatuuroverzichten die zich niet ten doel stellen om richtlijnen te formuleren, afkomstig uit zowel ethiek (Edg98, McK98) als gezondheidseconomie (Bra99, Dol00, Joh96), beschouwen de kwestie van de keuze van de relevante beoordelaars echter geenszins als afgedaan. De recente literatuur lijkt hun gelijk te geven: het debat gaat voort.

Interessant is het voorstel van Nord en collega's om onderscheid te maken tussen de utiliteit van gezondheidstoestanden, die het best door de patiënten zelf beoordeeld zou kunnen worden, en de sociale waarde van door medische voorzieningen teweeggebrachte veranderingen in gezondheidstoestanden, die het best door een steekproef uit de algemene bevolking bepaald zou kunnen worden. De utiliteit die patiënten zelf toekennen aan hun eigen gezondheidstoestanden zou, aldus deze auteurs, door de vertegenwoordigers van de algemene bevolking meegewogen moeten worden bij hun beoordeling van de sociale waarde van gezondheidsopbrengsten (Men99, Nor99, Ube00b, Ube00c, Men02). Het is niet te verwachten dat dit voorstel – dat voor de bepaling van de sociale waarde een beroep doet op de onder gezondheidseconomen vooralsnog weinig populaire *person trade-off* benadering – op korte termijn veel gehoor zal vinden. Het werpt echter een verhelderend licht op de discussie over de keuze van beoordelaars, al biedt het géén oplossing voor de inhoudelijke ethische vraag welk

gewicht moet worden toegekend aan de door succesvolle aanpassing beïnvloede utiliteitsschattingen van patiënten (Men02).

In hoeverre is deze discussie van praktisch belang? Kan de mate waarin de waardeoordelen van algemene bevolking en patiënten worden meegewogen consequenties hebben voor de ordening van voorzieningen naar mate van kostenutiliteit? Een onderzoek naar de waardering van de kwaliteit van leven bij interventies ter voorkoming van osteoporose laat zien dat de invloed aanzienlijk kan zijn. De waarden die door vrouwen met en zonder ervaring met heupfracturen werden toegekend aan osteoporose-gerelateerde gezondheidstoestanden bleken aanmerkelijk te verschillen. De geschatte kostenutiliteit van osteoporose-preventie liep daardoor uiteen van \$25.000 tot \$100.000 per gewonnen QALY, al naar gelang de waarden van de vrouwen zonder of met fractuur gebruikt werden (Gab99).

Net als bij het meten van gezondheidstoestanden, lijkt ook bij de keuze van de beoordelaars van die toestanden het gegeven van de culturele diversiteit een factor van belang te kunnen zijn. In de literatuur is hieraan nog weinig aandacht besteed (Bad01, Gol96a).

3.3.3 Keuze van een meetprocedure

Meer nog dan de keuze van de beoordelaars is de keuze van de juiste procedure voor het meten van voorkeuren voor gezondheidstoestanden een onderwerp uit de methodologie van de kostenutiliteitsanalyse dat gezondheidseconomen intensief heeft beziggehouden. Bijna dertig jaar van theoretisch debat en empirisch onderzoek heeft geen overeenstemming opgeleverd. De vaststelling dat die overeenstemming ontbreekt, is echter een opmerkelijk eensgezinde (Dol00, Dru97, Fro89b, Fro89d, Gol96a, Joh96, Len00, Nor92b).

Vele verschillende meetprocedures zijn voorhanden. Ze vallen in twee categorieën uiteen. Ten eerste zijn er de op keuzes gebaseerde procedures: de *standard gamble**

* Bij de SG wordt respondenten de keuze voorgelegd tussen twee alternatieven: een bepaalde chronische toestand of een loterij met als mogelijke uitkomsten 'volmaakte gezondheid' en 'dood'. De vraag is bij welke kans op 'volmaakte gezondheid' de keuze tussen beide alternatieven hun om het even is. Ook bij de TTO worden respondenten voor een keus geplaatst: de rest van het leven doorbrengen in een bepaalde chronische toestand of voor een kortere periode in volmaakte gezondheid voortleven. Hier is de vraag welk deel van hun levensverwachting zij bereid zijn op te geven om de toestand van volmaakte gezondheid te bereiken. De PTO wordt toegepast bij de SAVE (zie 3.3.1). Een andere mogelijke formulering is: van hoeveel mensen in een bepaalde chronische toestand zou het leven met 1 jaar verlengd moeten worden om deze optie gelijkwaardig te maken aan de verlenging van het leven van 100 gezonde mensen met eveneens 1 jaar? Bij de *paired comparison* moeten respondenten van een reeks paren van gezondheidstoestanden bepalen welke van de twee zij prefereren. Bij de *rating scaling*, de *visual analogue scale* en de *magnitude estimation* moeten ze aangeven welke waarde ze toekennen aan gezondheidstoestanden door respectievelijk een cijfer toe te kennen, een streepje op een lijn te zetten of door te zeggen 'hoeveel keer zo goed' (of slecht) een toestand is in vergelijking met een andere (Dru97, Gol96).

(SG), de *time trade-off* (TTO) en de *person trade-off* (PTO). Ten tweede zijn er de psychometrische, zoals de *paired comparison*, de *rating scaling*, de *visual analogue scale* en de *magnitude estimation*. De verschillende procedures zijn op hun beurt families van gelijksoortige technieken, die bij nadere inspectie blijken te verschillen in aspecten als de keuze van de tijdsduur van de te waarderen gezondheidstoestanden bij de SG, de duur van de resterende levensverwachting bij de TTO, de definitie van de ankerpunten van de intervalschaal, de wijze waarop de gezondheidstoestanden beschreven worden en de precieze verwoording van de vragen: aspecten die ieder voor zich hun invloed kunnen hebben op de meetresultaten (Len00, Nor92b, Smi93).

Het Washington Panel (Gol96a), de opstellers van de Canadese richtlijnen (CCO97) en Froberg en Kane, schrijvers van een viertal wat oudere overzichtsartikelen waarnaar nog steeds veel wordt verwezen (Fro89a t/m d), beschouwen zowel de psychometrische als de op keuzes gebaseerde procedures als geschikt voor toepassing in KUA's.

Anderen spreken om theoretische redenen een duidelijke voorkeur uit voor de op keuzes gebaseerde procedures, omdat die, anders dan de psychometrische, uitgaan van situaties waarin de keuze voor het ene de opoffering van het andere impliceert, waarmee ze aansluiten bij het economische gedachtegoed (Bra99, Bus00, Dol00, Joh96). In hun ogen zijn de SG en de TTO de meest serieuze kandidaten. Maar welke van de twee is te verkiezen? Het debat hierover onder gezondheidseconomen is volop gaande (Dol00).

Beide procedures hebben hun problemen. Voor zowel SG als TTO geldt dat de hypothetische keuzes die aan respondenten worden voorgelegd, niet realistisch zijn (Joh96; vergelijk 3.3.1). Bovendien kunnen de meetresultaten van beide procedures, zij het niet op dezelfde manier, worden beïnvloed door een, van persoon tot persoon wisselende, *reluctance to give up life*, die in werkelijkheid niets te maken hoeft te hebben met de relatieve waardering van gezondheidstoestanden (Bra99, Fow95). Tegen de SG pleit voorts dat er sterke aanwijzingen zijn dat mensen in de praktijk de axioma's van de *expected utility theory* waarop de SG gebaseerd is, systematisch geweld aandoen (Ble02). Een belangrijk bezwaar van de TTO is dat de meetresultaten onderhevig zijn aan de tijdsvoorkeur van de respondenten (Bra99, Dol00, Fro89b, Joh96). Bleichrodt identificeert vier vormen van vertekening die zowel SG als TTO parten spelen; bij de SG resulteren ze in het algemeen in een overschatting, bij de TTO lijken ze elkaar op te kunnen heffen, maar een garantie is er niet (Ble02). De derde keuze-gebaseerde procedure, de PTO, is – daarover zijn vrijwel alle auteurs het eens – nog onvoldoende onderbouwd en uitgewerkt. Als methode voor het meten van de waarde van gezondheidstoestanden voor individuen heeft hij bovendien het nadeel dat voorkeuren ten aanzien van de *verdeling* van gezondheidswinst er onvermijdelijk een rol in spelen.*

* Bij het bepalen van *disability weights* voor het berekenen van DALY's wordt gebruik gemaakt van een specifiek voor de *Global Burden of Disease Study* ontwikkelde versie van de PTO (Mur97, Sto00a).

Als methode voor het bepalen van de sociale waarde van de opbrengst van gezondheidszorgvoorzieningen lijkt de PTO zich echter in een toenemende belangstelling te mogen verheugen (Gre01, Men99, Nor99, Ube00b, Ube00c).

Voor zover nagegaan in onderzoek, blijken de meetresultaten van de psychometrische procedures slechts matig te correleren met die van SG en TTO (Bra99). De met de SG en de TTO gemeten waarden vertonen onderling een hogere correlatie, maar de SG-waarden neigen ertoe wat hoger te zijn dan de TTO-waarden (Ble97d, Bra99), in overeenstemming met de verwachtingen op theoretische gronden (Dol00). Niemand waagt zich in de literatuur aan algemene uitspraken over de mate waarin de meetwaarden van de verschillende procedures van elkaar verschillen. Het Washington Panel concludeert in 1996 dat nog veel meer onderzoek noodzakelijk is naar de wijze waarop de meetresultaten van de verschillende procedures zich onder verschillende omstandigheden tot elkaar verhouden (Gol96a).

Zes later staat die conclusie nog recht overeind. Onder gezondheidseconomen lijkt de TTO, nadat lange tijd de SG als de – zeker theoretisch gezien – superieure methode werd beschouwd, steeds meer de wind in de zeilen te krijgen (Ble02). Maar van een algemene consensus is geen sprake en er blijft onzekerheid over de validiteit van de waarderings van gezondheidstoestanden, ook als die voortkomen uit een TTO-procedure (Gre01, Len00). Bleichrodt, zelf een verklaard pleitbezorger van de TTO, stelt in een recent artikel vast:

“the existence of biases in SG and TTO utilities, the direction of which cannot always be predicted, highlights that the major challenge for health utility measurement is to develop utility measures and/or utility elicitation procedures that avoid or minimize the impact of biases on health utilities.” (Ble02)

De onzekerheid over de juiste procedure voor het meten van voorkeuren voor gezondheidstoestanden treft ook de door gezondheidseconomen aanbevolen gezondheidstoestand-classificatiesystemen (zie 3.2). De waarden die in de verschillende classificatiesystemen aan gezondheidstoestanden worden toegekend, zijn verkregen met uiteenlopende meetprocedures, zowel psychometrische als op keuzen gebaseerde, en zijn alleen daarom al niet zonder meer vergelijkbaar (Bra99). Ze zijn bovendien afhankelijk van de houdbaarheid van de *multi-attribute utility models* die eraan ten grondslag liggen: de wiskundige modellen die beschrijven hoe de utiliteit van de gezondheidstoestanden die door het classificatiesysteem gedefinieerd worden, samenhangen met de utiliteit van de verschillende aspecten of dimensies van die toestanden (Dru97). Dankzij het gebruik van deze modellen is het niet nodig om een groot aantal multidimensionale gezondheidstoestanden op hun utiliteit te beoordelen, maar kan volstaan worden met het waarderen van de verschillende niveaus op de afzonderlijke dimensies van die gezondheidstoestanden, terwijl de overige dimensies

constant worden gehouden (Fro89a). Deze vereenvoudiging van de waarderingstaak heeft haar prijs in de vorm van een of meer stringente voorwaarden van *utility independence*. De minst restrictieve voorwaarde, behorend bij het meest complexe model, eist dat de utiliteit van de verschillende niveaus op elke dimensie (bijvoorbeeld de mate van mobiliteit) onafhankelijk is van het niveau waarop de overige dimensies (bijvoorbeeld pijn of emotionele gesteldheid) gefixeerd zijn (Fro89a). Er is nog discussie over de vraag of deze zogenaamde *decomposed* benadering van het waarderingsprobleem, waarop de gezondheidstoestand-classificatiesystemen in de regel gebaseerd zijn, de voorkeur verdient boven de *composite* of *holistic* benadering, die de waardering van een groot aantal complete gezondheidstoestanden vereist (Dol00). Duidelijk is wel dat zonder de vereenvoudigende aannamen van de *multi-attribute utility models* het waarden van een compleet universum van mogelijke gezondheidstoestanden een complexe en tijdrovende aangelegenheid is (Bra99, Dru97). Vergelijkend empirisch onderzoek naar de validiteit van de *decomposed* en de *composite* benaderingen is nog niet gedaan (Dol00).

Kosten

Als een KUA erop gericht is een maatschappelijke beslissing te ondersteunen over, bijvoorbeeld, het al of niet uit collectieve middelen vergoeden van een gezondheidszorgvoorziening, dan moet die KUA in principe alle kosten verdisconteren die bij die voorziening in het geding zijn. Het gaat dan niet alleen om de kosten die gemoeid zijn met de toepassing ervan (de directe gezondheidszorgkosten: kosten van gebouwen en apparatuur, materiaalkosten, personeelskosten), maar ook om de kosten die patiënten moeten maken om de behandeling te kunnen ondergaan, zoals reiskosten, kosten van kinderopvang en van speciale diëten, om de tijd die patiënten en mensen uit hun omgeving besteden aan verzorging en behandeling (de directe niet-gezondheidszorgkosten), en om de verminderde ‘productiviteit’ in werk en vrije tijd ten gevolge van ziekte en overlijden (de indirecte of productiviteitskosten).

Tot zover bestaat in de literatuur over KUA’s een grote mate van overeenstemming, zij het dat de precieze terminologie en de categorisering van de kostenposten nog onderwerp van discussie zijn (Oos00). Minder duidelijk is hoe deze kostenposten geschat moeten worden. De schatting van de directe gezondheidszorgkosten bijvoorbeeld is verre van eenvoudig. Soms zal het mogelijk zijn af te gaan op geldende prijzen en tarieven. Vaak zullen die echter geen goede afspiegeling vormen van de werkelijke kosten, omdat de gezondheidszorg een niet-competitieve, in hoge mate gereguleerde maatschappelijke sector is. In dat geval is het nodig die kosten te schatten. De beoordeling of zo’n kostenschatting inderdaad noodzakelijk is en, zo ja, welke mate van grondigheid en precisie daarbij betracht moeten worden, is van vele factoren afhankelijk, onder meer van de mate waarin in een sensitiviteitsanalyse de uitkomsten

van de KUA door deze kosten beïnvloed blijken te worden. Ondubbelzinnige beslisregels zijn hiervoor niet te formuleren (Luc96).

Vooraf de productiviteitskosten en de kosten die zich voordoen tijdens door medische behandeling gewonnen levensjaren zijn het onderwerp van al lang bestaande controverses (Gar96, Hur00, Sto00b).

4.1 Productiviteitskosten

Waar het gaat om de productiviteit van zieken is er vrijwel algemene overeenstemming over het feit dat de productiviteitswinst ten gevolge van het uitstel van het overlijden geacht moet worden verdisconteerd te zijn in de *noemer* van de K/U-verhouding, dat wil zeggen: in de gezondheidswinst gemeten in QALY's. Productiviteit maakt deel uit van de waarde die levensverlenging voor ons vertegenwoordigt. Er zou sprake zijn van dubbeltelling wanneer de productiviteitswinst door levensduurverlenging daarnaast ook, als negatieve kosten, in de *teller* van de kosten/QALY-verhouding verrekend zou worden (Gar96, Luc96, Wei97a). Weinstein wijst erop dat door de verrekening in QALY's niet alle kosten van het overlijden van een productief lid van de maatschappij in beeld komen. Afgezien van de effecten op familie en vrienden, waaraan in KUA's gewoonlijk voorbij wordt gegaan – “perhaps an important limitation of the current practice” –, onderscheidt hij twee kostenposten die in de teller, de kostencomponent van de kosten/QALY-verhouding, verrekend zouden moeten worden: ten eerste de frictiekosten, dat wil zeggen: de kosten die gemoeid zijn met de vervanging van de overleden persoon op zijn of haar arbeidsplaats, en, ten tweede, dat deel van het inkomen van de overledene waarvan niet hij- of zijzelf, maar anderen profiteren: “external consumption by others in society, financed by taxes and other transfers” (Wei97a). Een methode voor het meten van deze laatste kostenpost geeft hij niet. In gangbare richtlijnen voor het uitvoeren van KUA's wordt eraan voorbijgegaan (CCO97, CVZ99, Luc96).

De discussie in de literatuur spitst zich toe op productiviteitsverlies in relatie tot morbiditeit. Het principe dat dubbeltelling vermeden dient te worden, geldt ook hier; minder duidelijk is wat het in dit geval impliceert. Verschillende standpunten zijn in de literatuur vertegenwoordigd. Het Washington Panel kiest voor een benadering die in overeenstemming is met de zojuist ten aanzien van mortaliteit geschetste en beveelt aan om de volledige invloed die ziekte uitoefent op de kwaliteit van leven van de patiënt in kwestie, met inbegrip van de mate van productiviteit, te verdisconteren in de noemer van de kosten/QALY-verhouding (Luc96, Wei96). Daarnaast onderscheidt het panel, net als Weinstein bij het overlijden, de frictiekosten en de *consumption externalities*, die respectievelijk door werkgever en collega's en door de rest van de maatschappij

gedragen worden. Deze beide kostenposten zouden in de teller van de K/U-verhouding moeten worden meegeteld (Wei96, Wei97a).

Anderen stellen dat het weinig voor de hand ligt, en ook in strijd is met de huidige onderzoekspraktijk, om ervan uit te gaan dat mensen bij de waardering van gezondheidstoestanden mogelijke inkomensverliezen mee zouden wegen (Bro97a, Mel99a). Zouden ze dat wel doen, dan zouden de waarderingen van gezondheidstoestanden afhankelijk zijn van het nationale niveau van sociale zekerheid (Bro97a). Rotterdamse onderzoekers betogen dat vanuit een maatschappelijk perspectief de frictiekosten – breed opgevat, als de som van de productieverliezen die optreden zolang de zieke medewerker niet vervangen is en de kosten die gemoeid zijn met het realiseren van adequate vervanging (Koo94, Koo95) – het werkelijke verlies aan productiviteit vertegenwoordigen (Bro97a, Bro97b). In deze benadering worden de productiviteitskosten ten gevolge van ziekte tijd- en plaatsafhankelijk. De zogenaamde frictieperiode, de periode die nodig is om een zieke medewerker te vervangen, is immers afhankelijk van de krapte op de arbeidsmarkt. Zo nam in Nederland de gemiddelde frictieperiode tussen 1995 en 1998 toe van 96 tot 123 dagen (Oos00). Andere critici van het Washington Panel hebben weinig op met de Rotterdamse frictiekostenmethode en menen dat het de voorkeur verdient het gehele verlies aan productiviteit van de zieke medewerker in de teller van de K/U-verhouding te verdisconteren (Mel99a).

In reactie op de kritiek op hun voorstel de productieverliezen ten gevolge van ziekte te verdisconteren in de QALY's, geven enkele leden van het Washington Panel toe dat hun consensus op dit punt broos, theoretisch zwak onderbouwd en voorlopig was. Maar zij houden vast aan het panelstandpunt, "to avoid monetizing items that could be viewed as consequences of health-related quality of life" (Wei97a, Wei99), zij het met de kanttekening dat meer onderzoek noodzakelijk is om tot betere, werkbare en geloofwaardige oplossingen te komen (Gar99). Eén prominent panellid vindt echter dat het bij nader inzien beter was geweest de productiviteitsverliezen buiten de QALY's te houden, uitgaand van de stelregel dat "QALYs should and do reflect health apart from its uses" (Rus99).

Zowel panelleden als critici tonen zich overtuigd van het belang van de kwestie. Rotterdamse onderzoekers laten zien dat de invloed op de kostenschattning van het al of niet verdisconteren van productieverliezen kan variëren van nihil tot enkele tientallen procenten, afhankelijk van het type voorziening, de ziekte en de patiëntengroepen die in het geding zijn (Koo94).

4.2 Kosten in gewonnen levensjaren

Even hardnekkig als de controverse over het verrekenen van productiviteitskosten is die omtrent de kosten die zich voordoen in de door medische behandeling gewonnen levensjaren (Wei97b). Het Washington Panel maakt onderscheid tussen de medische kosten die gemaakt worden in verband met ziekten die respectievelijk wel en niet gerelateerd zijn aan de interventie in kwestie, en de niet-medische kosten. Het panel beveelt aan de wel gerelateerde kosten in de analyse te betrekken. Over het al of niet verrekenen van de ongerelateerde medische kosten bereikt het Washington Panel geen overeenstemming (Wei96, Wei97b). Het beveelt aan om zo nodig in een sensitiviteitsanalyse na te gaan in welke mate de ongerelateerde kosten de K/U-verhouding beïnvloeden (Luc96). Niet-medische kosten zouden buiten de analyse gehouden moeten worden (Wei96).

Anderen betogen dat vanuit een maatschappelijk perspectief hoe dan ook alle in toegevoegde levensjaren gemaakte kosten in de analyse betrokken moeten worden, de gerelateerde en ongerelateerde medische en de niet-medische (Mel99a, Mel99b). Met hun pleidooi voor het verrekenen van de niet-medische kosten (consumptie minus productiviteit) keren zij zich tegen de opvatting dat productiviteitswinst door uitstel van overlijden verdisconteerd is in de gezondheidswinst gemeten in QALY's. Ze wijzen op de consequenties van het niet of niet volledig verdisconteren van de zogenaamde *mortality costs*: bij groepen die per saldo middelen verbruiken, zoals ouderen, kan het leiden tot een aanzienlijke begunstiging van voorzieningen die primair levensverlenging bewerkstelligen, ten koste van voorzieningen die vooral de kwaliteit van leven bevorderen (Mel97, Joh97). Het omgekeerde geldt voor groepen die per saldo meer middelen produceren dan ze verbruiken, zoals de jongere leeftijdsgroepen (Mel00). De leden van het Washington Panel houden echter vast aan hun standpunt (Gar99, Wei99). De controverse blijft onopgelost.

Afgezien van de complexiteit van het schatten van de gezondheidszorgkosten en de geschilpunten omtrent productiviteitskosten en kosten in gewonnen levensjaren kleven er aan de schatting van de kosten in KUA's nog andere, in de literatuur minder uitgebreid bediscussieerde, problemen, zoals die rond de kosten van informele zorg, de waardering van het verlies van onbetaald werk en de kosten van verloren vrije tijd van patiënten en mensen die hen ondersteunen (Bro00a, Koo98). Tezamen met de complexiteit van het schatten van de directe gezondheidszorgkosten, maken ze de schatting van de teller van de kostenutiliteitsverhouding tot een delicate aangelegenheid, die zich weinig leent voor rechttoe rechtaan sturing door ondubbelzinnige richtlijnen. Veel moet worden overgelaten aan de deskundigheid en objectiviteit van de

onderzoekers, onder meer bij de beoordeling van verkennende sensitiviteitsanalyses (Luc96).

De praktijk van het KU-onderzoek tot op heden stemt op dit punt niet tot optimisme. Een overzicht over ruim tweehonderd tussen 1975 en 1997 gepubliceerde KUA's laat zien dat in slechts 17% van de analyses enige vorm van directe niet-gezondheidszorgkosten of tijdskosten is verrekend. De kosten van de tijd van patiënten worden in minder dan 10% in de beschouwing betrokken, productiviteitsverliezen in 8% (Sto00b). De schrijvers constateren dat er een aanzienlijke variatie is in het kostenonderzoek in KUA's, zowel in de keuze van kostenposten die erin betrokken worden als in de wijze waarop ze worden geschat, en dat dit een belangrijke invloed kan hebben op de geschatte K/U-verhoudingen. Bovendien stellen ze vast dat de bron waaraan kostenschattingen ontleend worden vaak onduidelijk of in het geheel niet wordt gerapporteerd. Ze concluderen dat "if CUAs are to play a major role in informing policy makers, more uniformity and transparency in cost estimation are needed as well as continued vigilance on the part of analysts, reviewers, and journal editors" (Sto00b).

4.3 Marginale versus gemiddelde kosten

Behalve de genoemde methodologische keuzen en problemen is er nog een andere moeilijkheid die in het algemeen de zeggingskracht van kostenschattingen in KUA's beperkt. In het voorafgaande – en in de meeste KUA's – wordt eraan voorbijgegaan dat de gezondheidszorgkosten die met de toepassing van een bepaalde voorziening zijn gemoeid, veelal afhangen van de schaal waarop deze wordt toegepast, van de ervaring die ermee is opgedaan, van de mate waarin de voorziening is ingebed in alledaagse routines en logistiek, kortom: van allerlei tijd- en plaatsgebonden factoren (Ash00, Ger93). Dit inzicht vindt zijn weerspiegeling in het gezondheidseconomische concept van de marginale kosten: de extra kosten die gemaakt moeten worden om één extra eenheid van de gezondheidsopbrengst in kwestie te genereren (Gol96b, p 401).

Het marginale-kostenbegrip ontleent zijn belang aan het economisch inzicht dat in veel situaties niet een dichotome keuze aan de orde is – moet een bepaalde voorziening worden toegepast of niet –, maar de vraag of de investeringen in een voorziening, gegeven een zeker niveau van toepassing, verhoogd dan wel verlaagd moeten worden. Over de betekenis die dit begrip moet hebben voor de inrichting van KUA's wordt verschillend gedacht. Het Washington Panel formuleert enerzijds als aanbeveling dat "costs in CEA should reflect the marginal or incremental resources consumed, rather than average costs, from a long-run perspective" (Luc96, p 209), maar constateert anderzijds, verwijzend naar "economies of scale, scope, indivisibilities, or learning-by-doing", dat "in general, unless these effects are likely to be large, analysts can assume that the marginal costs of interest are constant", om te besluiten met het weinig

zeggende advies dat onderzoekers, indien de schaal- en leereffecten te groot lijken om te kunnen negeren, “should consult the literature for appropriate adjustments” (Luc96, p 199). Rotterdamse onderzoekers kennen aan het begrip een veel beperktere betekenis toe, waar zij poneren dat marginale kosten “voor specifieke onderzoeksvragen bij uitstek relevant zijn”, maar dat “integrale kosten zijn te prefereren [...] als een generalisatie naar landelijke kosten noodzakelijk is” (Koo98, Rut00b). Zij voeren aan dat op de lange termijn die in het geding is bij beslissingen op nationaal niveau, het onderscheid tussen marginale en integrale kosten zijn betekenis verliest (Oos00, Rut00b).

Haaks op het Rotterdamse standpunt staat de opinie van Mooney, voor wie *marginal analysis* de kern vormt van de bijdrage die economische evaluatie aan het stellen van prioriteiten in de gezondheidszorg te leveren heeft. Hij stelt dat in het algemeen de verleiding groot zal zijn om de gemiddelde kosten-gegevens te gebruiken, aangezien gegevens over *average costs* vaak wel beschikbaar zijn, terwijl die over *marginal costs*, op verschillende niveaus van toepassing van de voorziening, in de regel zullen ontbreken. Maar, zo stelt Mooney, “the chances are very strong that the use of these average cost figures will lead to the costings being wrong”. De aanname dat de relevante marginale kosten van de gemiddelde kosten verschillen is zijns inziens een veel beter vertrekpunt dan de aanname dat ze gelijk zouden zijn (Moo94b, p 31). In dezelfde geest betogen Ashton en zijn collega’s dat, waar het stellen van prioriteiten moet leiden tot budgettaire verschuivingen aan de marges van gezondheidszorgprogramma’s, het relevante kostenbegrip dat van de marginale kosten is. En die marginale kosten zijn helaas moeilijk te meten, omdat ze zowel context- als volumeafhankelijk zijn (Ash00).

Uit de standpunten ten aanzien van de verhouding tussen marginale en gemiddelde kosten blijkt dat er geen consensus is over wat precies onder de in K/U-verhoudingen te verdisconteren gezondheidszorgkosten verstaan zou moeten worden. Gaat het om de gemiddelde integrale kosten van een voorziening op de lange termijn? Maar hoe lang moet die termijn zijn om te kunnen veronderstellen dat ook de kosten die op korte termijn vast zijn, variabel en dus relevant worden (Rut00b)? En hoe verhoudt die lange termijn zich tot de mogelijkheid dat de kosten veranderen door leer- en schaaffecten en door organisatorische en technologische ontwikkelingen? Of zou het moeten gaan om de gemiddelde kosten van een gezondheidszorgprogramma van een specifieke omvang, bij een specifieke vorm van organisatie, in een specifiek stadium van de technologische ontwikkeling, op een specifiek punt van de leercurve van de betrokken beroepsbeoefenaren? Of om de marginale kosten van een bepaalde uitbreiding

* Vergelijk de opmerkingen van Van Hout: “Naar mijn idee moet er nog heel veel werk worden verricht om het kostenbegrip uit economische evaluatiestudies te laten aansluiten bij het kostenbegrip op de werkvloer. [...] Economische analyses zullen in de toekomst rekening moeten houden met de schaalgrootte en de tijdshorizon van de programma’s.” (Hou01)

daarvan?* En hoe zijn dergelijke tijd- en plaatsafhankelijke kostenschattingen te extrapoleren naar andere tijden en plaatsen (Oos00)? Hoe dan ook is duidelijk dat gebruikers van kostenutiliteitsschattingen er niet zonder meer van kunnen uitgaan dat de in een KUA geschatte gezondheidszorgkosten in hun situatie ook werkelijk relevant zijn.

* Het is duidelijk dat het bij beslissingen op lokaal niveau vaak zal gaan om de marginale kosten. Maar ook op landelijk niveau zullen niet zelden de marginale kosten van groter belang zijn dan de gemiddelde. Voor de meeste gezondheidszorgvoorzieningen geldt immers dat de doelmatigheid, behalve in het geval van totale ineffectiviteit, sterk afhankelijk is van de indicaties waarbij ze worden toegepast (CVZ01, Mul97, Nee00). Een effectieve voorziening is slechts zelden doelmatig of ondoelmatig *per se*, maar meestal doelmatig in een bepaald, soms zeer beperkt, indicatiegebied en minder doelmatig daarbuiten. Een beleidsvraag die zich in de praktijk dan ook vaak zal voordoen, is of het vanuit het gezichtspunt van de doelmatigheid verantwoord is om een voorziening die al vergoed wordt voor een beperkte indicatie, te gaan vergoeden voor een ruimer indicatiegebied – en dus op grotere schaal ingang te doen vinden. Dit is een beleidsvraag waarbij het relevante kostenbegrip dat van de marginale kosten is.

Disconteren

Mensen hebben in het algemeen, wat economen noemen, een positieve tijdsvoorkeur. Over positief gewaardeerde middelen beschikken we liever nu dan later; in de tussentijd kunnen we er plezier van hebben. Onwelkome lasten daarentegen ontlopen we liefst zo lang mogelijk. In de financiële sfeer weerspiegelt dit zich in het rentepercentage.

Kostenutiliteitsonderzoekers zijn het erover eens dat dit consequenties moet hebben voor de waardering van toekomstige kosten in een KUA. Om huidige en toekomstige kosten bij elkaar op te kunnen tellen, moeten ze in één eenheid worden uitgedrukt. Daarvoor kiest men huidige euro's. Dit betekent dat toekomstige euro's naar het heden moeten worden toegerekend door er een lagere waarde aan toe te kennen dan aan huidige euro's. Dit heet disconteren.

Maar hoe zit het met in de toekomst gewonnen *QALY's*? Zijn ook die minder waard dan nu genoten *QALY's*? Het rapport van het Washington Panel meldt in het hoofdstuk over de theoretische fundamenten van KUA's dat disconteren waar het gaat over gezondheidseffecten 'controversieel' is (Gar96). De auteurs kennen een belangrijke rol toe aan empirisch onderzoek naar de tijdsvoorkeur die mensen in het algemeen aan de dag blijken te leggen, in hun antwoorden op vragenlijsten en in hun feitelijk gedrag. Ze stellen dat op basis van dit onderzoek een zekere mate van disconteren, ook van gezondheidseffecten, voor de hand ligt, maar dat er *substantial disagreement* is over de vraag of het discontopercentage voor marktgoederen en voor niet-marktgoederen hetzelfde zou moeten zijn. Als er geen direct beschikbare marktgoederen zijn die als substituuut zouden kunnen dienen voor een niet-marktgoed als gezondheid, dan hoeven

de discontopercentages niet voor alle goederen en diensten gelijk te zijn. Voor die stelling, zo schrijven ze, zijn krachtige argumenten aan te voeren (Gar96).

In een ander hoofdstuk geeft het panel vervolgens een uitgebreider overzicht van de relevante literatuur en formuleert het zijn aanbevelingen ten aanzien van het disconteren (Lip96). Hier komt onder meer de veelgehoorde suggestie aan de orde dat disconteren van effecten en kosten tegen eenzelfde percentage zou leiden tot een systematische en onterechte onderwaardering van de doelmatigheid van preventieve interventies. De auteurs wijzen deze suggestie resoluut van tafel, met een verwijzing naar een uitspraak van de economen Fuchs en Zeckhauser: “self-respecting economists should not [...] use different rates because it is health that is being valued.” Als er al een reden zou zijn om de waarde van preventieve programma’s te benadrukken, dan zou de waardering van de toekomstige gezondheidsbaten naar boven moeten worden bijgesteld, en niet de discontervoet naar beneden. Aldus Fuchs en Zeckhauser in 1987 en, in hun voetspoor, het Washington Panel in 1996.

Uit de rest van het hoofdstuk blijkt echter dat er onder economen nog volop discussie is, juist over de vraag of er niet toch redenen kunnen zijn om kosten en gezondheidseffecten verschillend te disconteren. Bovendien geeft het hoofdstuk een overzicht van het empirisch onderzoek naar *individual time preferences*, waaruit interessante bevindingen naar voren komen die onder meer relevant zijn in verband met de waardering van preventieve interventies:

- mensen hanteren vaak discontopercentages die buiten het conventionele bereik van 0 tot 10% liggen; 0% komt heel vaak voor, maar ook percentages *lager* dan 0 en (veel) hoger dan 10% doen zich voor met enige regelmaat
- de gehanteerde discontopercentages neigen ertoe lager te zijn wanneer toekomstige, maar zeer zwaarwegende uitkomsten in het geding zijn
- ze neigen ertoe lager te zijn naarmate het tijdsinterval waarover gediscoteerd wordt langer is
- discontopercentages voor verliezen zijn in het algemeen lager dan die voor baten.

Er is, zo concluderen de schrijvers van het hoofdstuk, volop *behavioral evidence* die erop wijst dat de voorkeuren van individuele personen niet in overeenstemming zijn met discontering van alle mogelijke goederen tegen één vast percentage.

Het panel besluit niettemin zijn uiteenzetting met de aanbeveling om in de *reference case analysis*, de methodologische standaard-KUA die het panel propageert om de vergelijkbaarheid van KUA’s te bevorderen, kosten en gezondheidseffecten tegen hetzelfde percentage te disconteren. Maar zij voegen daar aan toe dat, aangezien theorie en empirisch onderzoek geen duidelijkheid verschaffen omtrent de relatie tussen individuele tijdsvoorkeuren en gangbare discontopercentages, “sensitivity analyses based on models that allow health and costs consequences to be discounted at different

rates may be conducted.” De Canadese en Nederlandse richtlijnen voor farmacoeconomisch onderzoek bevatten gelijksoortige aanbevelingen (CCO97, CVZ99).

Zoals te verwachten maken deze aanbevelingen geen eind aan de discussie in de literatuur. Van Hout beargumenteert dat voor een ongelijke discontering van kosten en gezondheidseffecten wel degelijk veel te zeggen is (Hou98). Hij verwerpt de belangrijkste theoretische argumenten die het Washington Panel aanvoert ten gunste van een gelijke discontering – het *consistency argument* van Weinstein en Stason (Wei77) en de zogenaamde *Keeler-Cretin paradox* (Kee83). Bij een KUA vanuit maatschappelijk perspectief zou het discontopercentage voor de kosten gebaseerd moeten zijn op de verwachte toename van het nationale inkomen, zo betoogt Van Hout. Het discontopercentage voor de gezondheidswinst zou daarentegen een weerspiegeling moeten zijn van de toegenomen efficiëntie waarmee die gezondheidswinst gegenereerd wordt en van de waarde die de maatschappij eraan hecht. Hij concludeert dat het goed te verantwoorden is om de gezondheidsbaten van bijvoorbeeld een vaccinatieprogramma voor kinderen niet of tegen een lager percentage te disconteren dan de kosten, zolang men ervan uitgaat dat het welzijn van huidige en toekomstige generaties even zwaar gewogen moet worden en het niet zonder meer aannemelijk is dat allerlei ziekten in de toekomst efficiënter dan nu genezen kunnen worden.

Ook andere Nederlandse onderzoekers bekritisieren de gangbare aanbevelingen ten aanzien van het disconteren van gezondheidswinst (Bar99, Ble00, Bon01b, Bro00b). “De beslissing om effecten tegen hetzelfde percentage te disconteren als kosten is arbitrair”, schrijven Bleichrodt en Brouwer. “De in de literatuur gegeven argumenten om één disconteringspercentage te hanteren voor effecten en kosten zijn niet relevant voor de meeste beleidsbeslissingen” (Ble00). En de Nederlanders staan niet alleen (Ana97, Goo96, Kra93). In hun kritiek op het discontopercentage van 3% dat deel uitmaakt van de DALY-formule stellen Anand en Hanson, met een verwijzing naar enkele klassieke economische argumenten tegen de gedachte dat toekomstige utiliteit of welzijn gedisconteerd zou moeten worden, dat het enige verdedigbare argument vóór het disconteren van gezondheidswinst is “the possibility that the world may end”. Let wel, zo vervolgen ze: “a 3% discount rate implies a 50% chance that the world will end in 23.4 years” (Ana97).

De discussie over de juiste wijze van disconteren van gezondheidswinst is in de wetenschappelijke literatuur dus nog volop gaande. Bij de interpretatie van K/U-verhoudingen, wanneer het erom gaat de doelmatigheid van voorzieningen met verschillende tijdsrelaties tussen investeringen en gezondheidswinst te vergelijken – bijvoorbeeld preventieve en curatieve voorzieningen –, is dit een belangrijk gegeven: de

keuze van de discontovoet kan de doelmatigheid van vooral preventieve voorzieningen maken en breken (Bar99, Bon01b).

Verdelingseffecten

De onzekerheden over de juiste methode voor het kwantificeren van de kostenutiliteitsverhouding kunnen consequenties hebben voor de verdeling over de bevolking van de gezondheidswinst die met een gegeven collectief gezondheidszorgbudget gegenereerd kan worden. Een gezondheidstoestandmeting die ongevoelig is voor de ziektespecifieke effecten van een bepaalde behandeling, een steekproef uit de algemene bevolking die het subtiele verschil tussen de gezondheidstoestanden voor en na een behandeling onvoldoende op waarde weet te schatten, een waarderingsmethode die door de *reluctance to give up life* van de respondenten het verschil in utiliteit van twee toestanden onvoldoende tot uitdrukking brengt: al die en vele andere factoren kunnen ertoe leiden dat een KUA de gezondheidswinst die een bepaalde gezondheidszorgvoorziening teweegbrengt onder- of juist overschat. Voor zover de resultaten van die KUA er indirect, door de betekenis die de verantwoordelijke besluitvormers eraan toekennen, toe leiden dat een voorziening waar een groep patiënten baat bij zou kunnen hebben, niet collectief gefinancierd wordt, wordt hierdoor een specifieke categorie patiënten benadeeld.

Van veel onzeker- en onvolkomenheden in de KUA-methodologie is niet in algemene zin aan te geven hoe ze voor de verdeling van de gezondheidswinst zullen uitpakken: de verdelingseffecten zijn afhankelijk van de specifieke combinaties van de kenmerken van populatie, ziekte, behandelingsindicaties en vergeleken voorzieningen enerzijds en de in de KUA's gemaakte methodologische keuzen anderzijds. Maar er zijn ook kenmerken van KUA's die in de literatuur de aandacht hebben getrokken juist omdat relatief makkelijk na te gaan is (of lijkt) welke verdelingseffecten ervan uitgaan.

Zo is uitvoerig gediscussieerd over het feit dat, *ceteris paribus*, ouderen, door hun kortere levensverwachting, minder QALY's te winnen hebben dan jongeren. Verdelingsbeslissingen op basis van KUA's bevoordelen daardoor jongeren boven ouderen. Over de vraag of dat gerechtvaardigd is wordt heel verschillend gedacht (Cal87, Cal90, Dan88, Eme89, GR86, Har87, Keu91, McK98, Men90, Mus87, Raw89, Rig86, Sas01, Smi87, Tsu00, Wag00, Wil92). Een gelijksoortig effect doet zich voor op de kwaliteitsdimensie, bij mensen met chronische ziekten en permanente functiebeperkingen: zij hebben bij een bepaald aantal jaren levensverlenging minder QALY's te winnen dan mensen zonder zulke ziekten en beperkingen (Ana97, Men99, Nor99, Nor01). Er is gewezen op het feit dat de KUA-systematiek geen rekening houdt met wat wel de *rule of rescue* (Jon86) is genoemd: de overtuiging dat kosten nog moeite gespaard mogen worden wanneer iemand hier en nu van een dreigende dood te redden is (Had91, Men99). Voorts is betoogd dat KUA's, doordat zij de behandelingswinst meten in QALY's, er ten onrechte aan voorbijgaan dat vertegenwoordigers van de algemene bevolking, wederom: *ceteris paribus*, een bepaalde verandering in de gezondheidstoestand van een ernstig zieke in het algemeen hoger waarderen dan dezelfde verandering in de toestand van een vrijwel gezond persoon (Dan93, Men99, Nor99, Sas01).

Minder vaak bediscussieerd, maar niet minder juist, is de vaststelling dat, behalve leeftijd, ook andere demografische kenmerken, zoals geslacht, ras en sociaal-economische status, krachtige voorspellers zijn van het resultaat van een medische behandeling, zodat een verdeling van middelen op basis van KUA's op gespannen voet staat met het streven naar een voor iedereen, ongeacht deze kenmerken, gelijke toegankelijkheid van gezondheidszorg (Bon91, Bon93). Verdelingseffecten vloeien ook voort uit de wijze waarop in KUA's het aggregatieprobleem wordt opgelost. KUA's belichamen, door de in QALY's gemeten gezondheidsbaten van verschillende personen bij elkaar op te tellen, een specifiek – en discutabel – antwoord op vragen als hoe de geredde levens van enkelingen afgewogen dienen te worden tegen de genezen keelonsteking van velen (Dan93, Sas01).

Naast deze implicaties van de wijze waarop in KUA's de gezondheidsopbrengst gemeten wordt, zijn er nog de verdelingseffecten die ontspringen aan de kostenkant van de kostenutiliteitsverhouding. Het verdisconteren van de verloren vrije tijd van patiënten en informele zorgverleners, het waarderen van onbetaald werk, de wijze van omgaan met leeftijds- en geslachtsafhankelijke verschillen in beloning: het zijn onderdelen van het kostenonderzoek in KUA's waar verschillend mee omgesprongen wordt, met onmiskenbare verdelingseffecten als mogelijk gevolg (Rus96).

Duidelijk is dat de KUA, toegepast als hulpmiddel om te komen tot maximalisatie van gewonnen QALY's bij een gegeven gezondheidszorgbudget, allerlei meer en minder zichtbare implicaties voor de verdeling van de behaalde gezondheidswinst in zich

draagt. Een verdeling van het collectieve gezondheidszorgbudget op grond van kostenutiliteitsverhoudingen, waarbij alle QALY's even zwaar gewogen worden, ongeacht de kenmerken van degenen aan wie ze ten goede komen, leidt tot een specifieke verdeling van de behaalde gezondheidswinst, die niet noodzakelijkerwijs een rechtvaardige verdeling hoeft te zijn.

In de gezondheidseconomische literatuur is het inzicht dat 'QALY-maximalisatie' op gespannen voet kan staan met 'rechtvaardigheid' inmiddels gemeengoed. Erkend wordt dat bij de allocatie van middelen in de gezondheidszorg regelmatig *equity-efficiency trade-offs* aan de orde zijn: doelmatigheid moet tot op zekere hoogte worden opgeofferd aan rechtvaardigheid en *vice versa*. De vraag hoe deze afwegingen het best benaderd kunnen worden, krijgt de laatste jaren in de theoretische en methodologische gezondheidseconomische literatuur steeds meer aandacht, maar een bevredigend antwoord is voorlopig niet in zicht (Ble97a, Pol02, Sas01, Wag91, Wil00). Enerzijds wordt gepoogd om te komen tot gewichten waarmee QALY's die terechtkomen bij verschillende categorieën personen – jongeren en ouderen, ernstig zieke en vrijwel gezonde mensen, rokers en niet-rokers etc. – gewogen kunnen worden. Maar erkend wordt dat het onderzoek dat nodig is om deze gewichten te kunnen baseren op voorkeuren van de algemene bevolking, nog in een embryonaal stadium verkeert (Wil00). Anderzijds wordt voorgesteld om in elk geval op korte termijn in economisch evaluatieonderzoek steeds in beeld te brengen aan welke subpopulaties – gedefinieerd in termen van bijvoorbeeld leeftijd, geslacht, etniciteit, sociaal-economische status – de gezondheidsbaten van een gezondheidszorgvoorziening ten deel vallen, zodat besluitvormers de verdelingseffecten van alternatieve bestedingsopties zullen kunnen beoordelen (Sas01).

Een heel andere benadering is geïmpliceerd in het eerder genoemde *saved young life equivalent*, ofwel de SAVE (Nor92a). De SAVE kwantificeert de *sociale waarde* die wordt toegekend aan een bepaalde gezondheidsverbetering bij een persoon met een bepaalde gezondheidstoestand, een bepaalde leeftijd en eventuele andere kenmerken: hoeveel van dergelijke verbeteringen bij die categorie personen wegen op tegen het volledig herstellen van de gezondheid van één jeugdig persoon die anders een zekere en snelle dood gestorven zou zijn (cf. 3.3.1)? Aldus incorporeert de SAVE zowel *efficiency* als *equity*-overwegingen, wat hem op de kritiek komt te staan van gezondheidseconomen die juist in het helder onderscheiden van deze twee typen overwegingen één van de grote deugden van de gezondheidseconomische analyse zien (Wil00). Degenen echter die met Menzel van mening zijn dat elke poging om gezondheidswinst in QALY's te kwantificeren "will remain persistently suspicious" indien niet de verdelingsbeslissingen waarbij op QALY's een beroep wordt gedaan, in beschouwing genomen worden (Men90), zullen van deze kritiek niet onder de indruk zijn. Vanuit hun gezichtspunt is het immers alleszins aannemelijk dat het in de

gezondheidszorg eenvoudigweg onmogelijk zal blijken om *efficiency* op een geloofwaardige manier te kwantificeren lós van *equity*-overwegingen.

Tot besluit

“Het QALY-concept wordt al ruim 25 jaar toegepast en is beschreven in talloze wetenschappelijke publicaties. Aanvankelijk leverden wetenschappers nog wel kritiek op de kwaliteiten van QALY's, maar intussen is dit concept dusdanig stevig onderbouwd dat QALY's wereldwijd worden gebruikt in beleidsmodellen.” (Pol02)

“QALYs seem to be the most widely used method for capturing both quality and quantity of life. However, the QALY concept is controversial [...]” (Gre02)

Is doelmatigheid een bruikbaar criterium voor het bepalen van de samenstelling van het basispakket? Dat was een van de vragen die de toenmalige minister van VWS in haar adviesaanvraag voorlegde aan de Gezondheidsraad. De minister verwees naar het rapport *Volksgezondheidszorg* van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR). De WRR gaf in dat rapport een specifieke invulling aan het doelmatigheidscriterium en schetste een rooskleurig beeld van de bruikbaarheid ervan. De kostenutiliteit van voorzieningen, uitgedrukt in kosten per te winnen *Disability Adjusted Life Year* (DALY), zou volgens de WRR zowel in de preventieve als in de curatieve gezondheidszorg hét criterium voor de toedeling van collectieve middelen moeten zijn. De WRR tekende aan ‘dat in de voorgestelde methodiek lastige keuzen overblijven’, maar merkte op dat het daarbij gaat om ‘in zeker opzicht [...] technische vragen van uitwerking’, die geen reden zouden moeten zijn om het kostenutiliteitscriterium niet toe te passen.

In dit rapport hebben deze ‘technische vragen van uitwerking’ het volle pond gekregen. Het resultaat geeft te denken: niet zozeer voor de gezondheidseconomen, die in het algemeen goed bekend zijn met de in hun wetenschappelijke literatuur gesignaleerde problemen, als wel voor de besluitvormers, die wellicht, misschien tegen beter weten in, hoopten met de kostenutiliteitsanalyse (KUA) een relatief simpel toepasbaar instrument in handen te hebben dat hen van veel moeilijke afwegingen zou kunnen verlossen. De KUA en het *Quality Adjusted Life Year* (de QALY), de maat voor gezondheidswinst waardoor deze vorm van kosteneffectiviteitsanalyse zich van andere onderscheidt, blijken ruim 25 jaar na hun introductie nog steeds, in allerlei opzichten, ter discussie te staan.

Standaardisatie is geen oplossing voor dit probleem. De veelgehoorde roep om standaardisatie is ingegeven door precies die vaststelling die in het licht van de voorafgaande hoofdstukken onontkoombaar lijkt: dat er tussen KUA's methodologische verschillen bestaan die betekenisvolle effecten hebben op de uitkomsten en die de vergelijkbaarheid van K/U-verhoudingen discutabel maken (CCO97, CVZ99, Gol96b). Maar standaardisatie leidt niet noodzakelijkerwijs tot een juiste methodologie. Er is geen reden om aan te nemen dat één bepaalde verzameling van methodologische keuzen zou leiden tot dezelfde rangordening van voorzieningen naar kostenutiliteit als een andere, evenzeer verdedigbare, verzameling. Bovendien, al zou die overeenstemming bestaan, dan zou dat nog geen bewijs zijn voor de juistheid van de rangorde. Aanzetten tot standaardisatie zijn tot nu toe dan ook steeds gevolgd door nieuwe ronden van discussie.

Wat betekent dit voor de bruikbaarheid van de K/U-verhouding en, om terug te keren naar de eigenlijke vraag van de minister, voor de toepasbaarheid van het doelmatigheids criterium bij de afbakening van een collectief te financieren basispakket? Op dit punt gekomen is het nuttig opnieuw te onderscheiden tussen de verschillende aspecten van het begrip doelmatigheid die in dit rapport, aan de hand van de bespreking van de KUA, aan de orde zijn geweest. De problemen die we in de voorafgaande hoofdstukken tegenkwamen zijn verschillend van aard en vragen om verschillende benaderingen. Ze passeren hier opnieuw de revue, maar nu in omgekeerde volgorde, zodat we eindigen met de kwestie van de QALY.

Disconteren

Is de waarde van gezondheidswinst onafhankelijk van het moment, in het heden of in de toekomst, waarop ze geboekt wordt en, zo nee, hoe moet dan toekomstige winst in vergelijking met huidige gewogen worden? In de gezondheidseconomische literatuur zijn deze vragen uitgebreid bediscussieerd. Een eensluidend antwoord is daaruit niet voortgekomen.

Of de doelmatigheid van een gezondheidszorgvoorziening afhankelijk is van de wijze van disconteren van toekomstige gezondheidswinst is in een sensitiviteitsanalyse relatief gemakkelijk na te gaan. In de praktijk is die afhankelijkheid vooral groot bij preventieve voorzieningen en *public health* programma's (Kra93). Zolang voorzieningen vergeleken worden met vergelijkbare tijdsrelaties tussen kosten en gezondheidswinst, levert de onzekerheid over het disconteren weinig problemen op. Doelmatigheidsvergelijkingen die deze beperking te buiten gaan – bijvoorbeeld een directe kwantitatieve vergelijking van de doelmatigheid van preventieve en curatieve voorzieningen – vergen echter duidelijkheid over de juiste wijze van disconteren. Bij gebrek aan wetenschappelijke consensus zal die duidelijkheid slechts door de overheid geschapen kunnen worden (Kra93, Bro00b).

Indien de onzekerheid rond het disconteren van gezondheidswinst blijft voortbestaan, zal dat vooral een belemmering zijn voor een vergelijking van K/U-verhoudingen over de volle breedte van de gezondheidszorg. De vergelijking van K/U-verhoudingen en kosteneffectiviteitsverhoudingen binnen deelgebieden van de gezondheidszorg, waar een zekere homogeniteit bestaat ten aanzien van de tijdsrelatie tussen kosten en gezondheidswinst, zal in het algemeen van het disconteringsprobleem minder hinder ondervinden.

Kosten

De problemen die spelen bij het kwantificeren van de kosten zijn in dit rapport verre van uitputtend behandeld. Maar duidelijk is dat verschillende benaderingswijzen van de diverse kostenposten – bijvoorbeeld van de directe gezondheidszorgkosten, de productiviteitskosten, de kosten in gewonnen levensjaren, de informele zorg – tot belangrijke verschillen in de schatting van de kosten kunnen leiden. Door middel van sensitiviteitsanalyse is tot op zekere hoogte na te gaan in hoeverre een doelmatigheidsschatting van variaties in kosten afhankelijk is. Maar de sensitiviteitsanalyse is in dit geval veel minder eenvoudig dan wanneer het erom gaat alleen het effect na te gaan van de wijze van disconteren van toekomstige gezondheidswinst. Er kunnen immers veel verschillende kostenvariabelen in het geding zijn, die ook weer onderling van elkaar afhankelijk kunnen zijn (bijvoorbeeld kosten van informele zorg en directe gezondheidszorgkosten) en die voor een belangrijk deel variëren over een bereik dat niet *a priori* duidelijk is. De methodologie van het schatten van de cumulatieve invloed van de onzekerheden in verschillende, deels elkaar onderling beïnvloedende variabelen is bovendien nog volop in ontwikkeling (Bri95, Bri97, Bri98a, Bri98b, Bri00, Gar00, Hun98b, Man96).

Problemen in de schatting van de kosten spelen elke vorm van doelmatigheidsanalyse parten, maar ze zijn lastiger naarmate de ambitie van de analyse

hoger is gesteld en dus het scala aan relevante kostensoorten groter is. Vanuit het algemeen maatschappelijk perspectief geldt het principe dat alle kosten relevant zijn. Aangezien onder gezondheidseconomen niet op alle punten overeenstemming bestaat over welke keuzen en definities van kosten uit dat principe voortvloeien, lijkt het onvermijdelijk dat de overheid, wanneer zij het doelmatigheids criterium in haar besluitvorming over de afbakening van het basispakket wil betrekken, zelf sommige knopen doorhakt (Boe02). Acht zij de productiviteitsverliezen ten gevolge van ziekte relevant? Zo ja, wat verstaat zij daaronder: gaat het om de integrale productiviteitsverliezen of om de – arbeidsmarktafhankelijke – frictiekosten? Acht zij het juist dat voor de schatting van productiviteitsverliezen leeftijds- en geslachtsspecifieke gemiddelde lonen worden gebruikt? En wat betreft de kosten in gewonnen levensjaren: is de overheid geïnteresseerd in deze kosten en, zo ja, alleen in de gerelateerde of ook in de ongerelateerde medische kosten, of zelfs in de niet-medische kosten? Voor en tegen al deze opties valt veel te zeggen en in de gezondheidseconomische literatuur is dat volop gebeurd. Het maken van keuzen lijkt eerder op de weg te liggen van de overheid dan van de wetenschap.

Gezondheidswinst

Wanneer we de opbrengst van gezondheidszorg meten in termen van ‘gezondheidswinst’, dan vangen we daarmee niet alles wat gezondheidszorg waardevol maakt. In de besluitvorming over de afbakening van het basispakket zal er ruimte moeten zijn om andere aspecten van de opbrengst mee te wegen. Dat doet echter niets af aan de centrale rol van de gezondheidswinst: als component van doelmatigheid en ook, maar dat terzijde, als zelfstandig criterium.*

Hoe gezondheidswinst te definiëren? Veranderingen in levensduur én veranderingen in gezondheidstoestand zijn beide relevant, daarover is iedereen het eens. Minder vanzelfsprekend is dat die veranderingen op één algemeen toepasbare dimensie gewaardeerd zouden kunnen of moeten worden. De meest courante wijze om gezondheidswinst te meten met één universele maat, de QALY, stuit op talrijke moeilijkheden: problemen bij het integreren van ‘kwaliteit’ en ‘kwantiteit’ van leven, bij het aggregeren van QALY’s van verschillende personen, bij het op een gestandaardiseerde wijze beschrijven van gezondheidstoestanden, bij het formuleren van een theorie over voorkeuren voor gezondheidstoestanden, bij het ontwerpen van procedures voor het meten van die voorkeuren en bij de keuze van de beoordelaars.

* Stel dat voor een acute, dodelijke ziekte twee behandelingen beschikbaar zijn. Behandeling A verlengt het leven met één jaar en kost honderd euro; behandeling B geeft tien jaar levensverlenging, maar kost honderdduizend euro. De doelmatigheid van behandeling A, uitgedrukt in directe medische kosten per gewonnen levensjaar, is dan honderd keer zo hoog. Vanuit een oogpunt van effectiviteit verdient echter behandeling B de voorkeur.

Deze problemen roepen enerzijds vragen op over de interpretatie van de QALY: is de QALY wel een zinvolle meeteenheid, een entiteit waaraan een duidelijke betekenis valt toe te kennen? Anderzijds zijn ook deze problemen deels vanuit een kwantitatieve invalshoek te benaderen. In een sensitiviteitsanalyse is het in principe mogelijk om na te gaan in hoeverre de schatting van de gezondheidswinst in QALY's afhankelijk is van onzekerheden in de schatting van de kwaliteitscorrectiefactoren. Maar dit is geen sinecure, aangezien er veel verschillende gezondheidstoestanden in het spel kunnen zijn. Daar komt bij dat een goede beoordeling van het effect op de resulterende kostenutiliteitsverhouding een, methodologisch gecompliceerde, multivariate sensitiviteitsanalyse vergt waarin ook de onzekerheden betrokken worden in de schatting van de duur van de perioden die in de verschillende gezondheidstoestanden worden doorgebracht en, uiteindelijk, ook die in de schatting van de kosten. Bovendien is moeilijk in te zien hoe de meer basale theoretische problemen, zoals die bij de integratie van levenskwaliteit en levensduur en de aggregatie van gezondheidsbaten van verschillende mensen, in een kwantitatieve sensitiviteitsanalyse te hanteren zouden zijn.

Dat neemt niet weg dat het goed voorstelbaar is dat in bepaalde situaties de problemen bij de kwaliteitscorrectie van levensduur nauwelijks terzake doen, omdat de doelmatigheid in hoge mate bepaald wordt door het uitstel van sterfte en niet of nauwelijks door de kwaliteit van leven. Maar men kan zich hier makkelijk vergissen. In het geval van behandeling met een cholesterolsyntheseremmer ('statine') ter voorkoming van coronaire hartziekte, bijvoorbeeld, lijkt het aannemelijk dat de doelmatigheid sterk bepaald wordt door het uitstel van sterfte en slechts in geringe mate door vermindering van morbiditeit (Bon99). De economische analyse ten behoeve van de CBO-consensusvorming over de indicaties voor de behandeling met statines richtte zich dan ook primair op de kosten per gewonnen levensjaar (CBO98). In een latere KUA bleken niettemin de kosten van statinebehandeling per gewonnen levensjaar 20% af te wijken van die per gewonnen QALY (Pro00). Bovendien kan een gering verschil tussen de kosten per gewonnen levensjaar en per gewonnen QALY ook betekenen dat betekenisvolle effecten op de kwaliteit van leven in de KUA zijn gemist. Het Washington Panel (Gol96) wijst bijvoorbeeld op de discrepantie tussen het grote verschil dat werd aangetoond in de in 3.2 besproken Australische KUA van borstkankerscreening (Hal92) en het verschil van 5,6% dat gevonden werd in een Nederlandse KUA (Kon91).

Onderzoek dat demonstreert hoe verschillen in de methodologie van de kwaliteitscorrectie van gewonnen levensduur uitwerken in kostenutiliteits-schattingen is in de literatuur niet ruim voorradig. Afgezien van eerder genoemde voorbeelden (Hal92 in 3.2, Gab99 in 3.3.2) is er het onderzoek van Hornberger en collega's (Hor92) dat laat zien dat kostenutiliteitsverhoudingen van nierdialyse kunnen variëren van \$35.000 tot \$45.000, afhankelijk van de methode van gezondheidstoestand-waardering. Er zijn

echter heel wat voorbeelden van onderzoek dat aantoont dat gezondheidstoestand-waarderingen, afhankelijk van de gebruikte methodologie, behoorlijk kunnen verschillen (Bel01, Fry93, Nea95, Neu00b, Nor92b, Rea84). Dat maakt aannemelijk dat we bij de problemen rond de kwaliteitscorrectie van levensduur niet van doen hebben met een kwantitatief onbeduidend fenomeen. Heel wat voorzieningen, bij bijvoorbeeld chronische ziekten als maagdarfstormissen, aandoeningen van het bewegingsapparaat, longaandoeningen en psychische stoornissen, zijn tenslotte primair gericht op verbetering van de gezondheidstoestand en niet op levensverlenging. En veel levensverlengende voorzieningen hebben ook een belangrijk effect op de kwaliteit van leven. Zo bleken bijvoorbeeld in een Nederlands onderzoek de kosten van longtransplantatie per gewonnen QALY ruim 20% lager dan per gewonnen levensjaar (Enc97).

De conclusie dat de QALY geen algemeen toepasbare maat voor gezondheidswinst is waarop men zich kan verlaten bij de vergelijking van de doelmatigheid van uiteenlopende zorgvoorzieningen, lijkt al met al gewettigd.*

Alternatieven voor de QALY

Ligt er een andere universele ééndimensionale maat voor gezondheidswinst in het verschieft die de plaats van de QALY in de KUA zou kunnen innemen? Tot nu toe zijn in de literatuur drie kandidaten naar voren geschoven, de DALY, het *Healthy Year Equivalent* (HYE) en het *saved young life equivalent* (SAVE).

De DALY is primair ontwikkeld als maat om internationaal de *burden of disease* te kunnen vergelijken, maar ook wel omarmd als maat voor gezondheidswinst in KUA's (Mur96, Mur97). Het is niet veel anders dan een 'omgekeerde QALY'. De DALY combineert verloren jaren ten gevolge van voortijdig overlijden met een weging van geleefde jaren voor de mate van *disability*. Het resultaat is een maat voor gezondheidsverlies in plaats van gezondheidswinst, zoals de QALY. De problemen zijn dezelfde.

* Dit rapport staat met deze terughoudende conclusie niet alleen, integendeel. In de methodologisch georiënteerde literatuur over KUA's is het tamelijk gebruikelijk om over de geloofwaardigheid van de QALY niet hoog op te geven. Het Washington Panel schrijft bijvoorbeeld over de verschillende manieren om gezondheidstoestanden te wegen: "This diversity in how preference weights are gathered markedly constrains the ability to credibly compare analyses where the effectiveness measure is presented in QALYs" (Gol96, p 119). En de gezondheidseconomen Drummond en Torrance stellen, in artikel waarin zij waarschuwen tegen de misplaatste suggestie van vergelijkbaarheid die kan uitgaan van zogenaamde *league tables* waarin kostenutiliteitsverhoudingen afkomstig uit verschillende onderzoeken eenvoudig onder elkaar worden gezet: "[...] the results of economic evaluation studies should be presented in a disaggregated form as well as in the form of cost-effectiveness ratios. [...] the consequences of interventions in terms of their physical effects should be reported separately from utility weights used to value them." (Dru93)

Het minder bekende HYE, gebaseerd op de waardering van *lifetime health profiles* in plaats van geïsoleerde gezondheidstoestanden, is bedoeld om een oplossing te bieden voor bepaalde problemen van de QALY (zie 3.1). De theoretische meerwaarde van het HYE in vergelijking met de QALY lijkt evident, al is hij niet onomstreden (Rie98). De meting van HYE's is echter praktisch gezien geen sinecure en in empirisch onderzoek tot nu toe weinig toegepast. Niettemin lijkt de HYE-benadering nog steeds een nadere exploratie waard te zijn (Lle02).

Een radicaler alternatief voor de QALY biedt de SAVE (zie 3.3.1). Deze maat kwantificeert de sociale waarde van veranderingen in gezondheidstoestanden, met een expliciete verwijzing naar het type verdelingsbeslissingen ter ondersteuning waarvan de gezondheidswinst gemeten wordt. De SAVE ontsnapt aan veel van de problemen die in hoofdstuk 3 geïdentificeerd werden en lijkt tevens een manier te bieden om op een directe, inzichtelijke wijze om te gaan met de verdelingsaspecten die verbonden zijn aan de waardering van de opbrengst van gezondheidszorgvoorzieningen. Meer nog dan het HYE lijkt de SAVE een maat die in de Nederlandse context een nadere exploratie waard is.

Er is nog een andere veelbelovende benadering van het probleem van de waardering van gezondheidswinst, die echter in het voorafgaande niet aan de orde kwam omdat ze de KUA te buiten gaat: de *willingness to pay* (WTP) benadering. De WTP, ook wel *contingent valuation* genoemd, waardeert gezondheidswinst in termen van geld en kan dus kan worden ingezet in de door gezondheidseconomen hogelijk gewaardeerde, want stevig in de welvaartseconomische theorie gegrondveste, *cost-benefit analysis* (Dru97). Met name de WTP-vraagstelling die respondenten uitnodigt aan te geven hoeveel verzekeringspremie zij bereid zijn extra te betalen om een bepaalde voorziening aan het verstrekkingenpakket toe te voegen (de *ex ante insurance based WTP*) verwoordt met een plezierige, onverbloemde directheid de onaangename vraag waar het bij de toedeling van (collectieve) middelen aan gezondheidszorgvoorzieningen nu eenmaal om gaat: hoeveel hebben we er voor over? De WTP-benadering heeft bovendien het voordeel dat andere dimensies van de opbrengst van gezondheidszorg dan alleen de gezondheidswinst relatief eenvoudig in de waardering te betrekken zijn (Bir99, Lee97, Neu94). Ongetwijfeld kent ook de WTP aanzienlijke problemen (Ble01, Klo99, Ols01). Wie de ontwikkeling van het gezondheidseconomisch evaluatie-onderzoek overziet vraagt zich niettemin af of de afkeer van het waarderen van gezondheid in termen van geld, die zo'n belangrijke rol heeft gespeeld in de (over)waardering van de KUA ten koste van de *cost-benefit analysis* (Gar96, p 28; Wei80, p 240), niet een slechte raadgever is geweest.* Sinds enkele jaren mag de WTP zich in een toenemende

* Hetzelfde geldt voor het argument dat de WTP op een onwenselijke wijze beïnvloed zou worden door het inkomen van de respondenten (Ble01, Gar96, Ols01): voor zover in een onderzoek een relatie zou blijken te bestaan met de hoogte van het inkomen, is de WTP immers met behulp van regressie-analyse eenvoudig naar elk gewenst inkomensniveau te standaardiseren.

belangstelling verheugen in de internationale gezondheidseconomische literatuur (Bir99, Dru97, Klo99, OBr96). Ook in Nederland lijkt een herwaardering aanstaande.*

Hoe interessant met name de SAVE en de WTP ook zijn, geen van de genoemde vier benaderingen van het waarderen van gezondheidswinst is op dit moment, en waarschijnlijk ook niet op een termijn van enkele jaren, te beschouwen als een werkelijk alternatief voor de QALY. Dat betekent dat een geloofwaardige manier om 'gezondheidswinst' met één algemeen toepasbare maat te meten op de korte termijn niet beschikbaar is. De overheid zou zich niettemin kunnen bekennen tot één welomschreven methodologie voor het meten van gezondheidswinst in QALY's. Maar gezien de aard van de problemen die in de literatuur over de QALY aan de orde zijn en de enorme maatschappelijke druk waaronder de besluitvorming over de verdeling van collectieve middelen in de gezondheidszorg moet plaatsvinden, lijkt dit een weinig kansrijke strategie. Het zou de overheid niet meevallen om zich te verweren tegen het verwijt dat de QALY in essentie "a technical solution to a political problem" is (Car91).

Hoe verder met doelmatigheid?

Er lijkt weinig anders op te zitten dan te accepteren dat voor de vergelijking van de doelmatigheid van gezondheidszorgvoorzieningen voorlopig geen algemeen toepasbare technische middelen voorhanden zijn. De vaak veelsoortige effecten van voorzieningen op de gezondheidstoestand, de effecten op de levensduur en de kosten zullen door de besluitvormende instantie zelf in de besluitvorming gewogen moeten worden. Bovendien zullen naast de doelmatigheid nog andere overwegingen in de besluitvorming tot hun recht moeten komen:** effectiviteit, invloed op anderen dan de patiënt en op de maatschappij als geheel, andere aspecten van de opbrengst van voorzieningen, de mate waarin (de doelmatigheid van) een voorziening zich verder zou kunnen ontwikkelen, het beschikbare budget, praktische overwegingen, verdelende

* Anders dan in bijvoorbeeld het bekende boek van Drummond en collega's (Dru97), komt in twee recente Nederlandse boeken over gezondheidseconomie en gezondheidseconomisch evaluatie-onderzoek de WTP-benadering nauwelijks (Uyl00) of summier (Blei01) aan de orde. Maar ook in deze boeken wordt een meer prominente rol van de WTP in de toekomst niet uitgesloten.

** In de gezondheidseconomische literatuur wordt dit in het algemeen niet bestreden (Blei01, Bro00a, Bus00, Rus96, Wei97b). Hurley wijst erop dat gezondheidseconomen dit vaak aanvoeren ter verdediging, wanneer zij geconfronteerd worden met kritiek op hun ambitie alle belangrijke effecten van gezondheidszorg te vangen in de QALY. Zij antwoorden dan "that the results of economic evaluations are only one piece of information intended to aid decision making and that other relevant considerations will enter via other avenues at the time a decision is taken." Daarmee, betoogt Hurley, verdampt echter veel van de aantrekkingskracht van het perspectief alle mogelijke gezondheidszorgvoorzieningen op één dimensie van kostenutiliteit te kunnen ordenen. In het besluitvormingsproces zullen immers hoe dan ook de resultaten van de doelmatigheidsanalyse gecombineerd moeten worden met andersoortige gegevens. En als dat zo is, vervolgt hij, "why should the entire structure of the economic evaluation be distorted so as to obtain a single number at the end of the analysis?" (Hur98)

rechtvaardigheid in relatie tot dimensies als ernst van de ziekte, prognose, afhankelijkheid en leeftijd, en andere ethische en juridische aspecten. En dan zal vaak ook nog niet het eenvoudige al of niet vergoeden van voorzieningen aan de orde zijn, maar het specificeren van de indicaties waarbij die voorzieningen vergoed moet worden* – indicaties die bovendien continu aan verandering onderhevig zijn, door nieuwe wetenschappelijke en technologische ontwikkelingen en de opkomst van concurrerende of additionele interventies.** Het sluitstuk van de besluitvorming over de toedeling van schaarse middelen in de gezondheidszorg, ten slotte, rust in de handen van zorgverleners, die zullen moeten beslissen of de vastgestelde indicaties voor vergoeding bij individuele patiënten aanwezig zijn.

Hoe om te gaan met deze overweldigende complexiteit? Het is duidelijk dat we hier voor een groot probleem staan, in termen van uitvoerbaarheid, maar bovenal van legitimiteit. Uit de internationaal gevoerde discussies en de ervaringen van de afgelopen vijftien jaar wordt duidelijk dat alleen discussie over verdelingsprincipes onvoldoende is. Naast principes zijn actieve participatie van alle betrokken partijen en deskundigheden, heldere verdeling van verantwoordelijkheden en transparante besluitvormings- en beroepsprocedures essentieel om meningsverschillen over prioriteitstelling in de zorg op te lossen en een collectief maatschappelijk leerproces op gang te brengen (Bur98, Dan93, Dan97, Dan00, Day98, Ham97, Hol98, Kle98, Leg02, Nee00).

Afbakenen van deelgebieden in de gezondheidszorg waarbinnen de prioriteitstelling zijn beslag moet krijgen, lijkt een van de noodzakelijke mechanismen om de complexiteit hanteerbaar te maken. Hoe die sectoren precies omschreven moeten worden is vers twee. Maar het is duidelijk dat door het afbakenen van deelgebieden als de zorg voor geestelijk gehandicapten, ambulante geestelijke gezondheidszorg, preventie van hart- en vaatziekten en solide-tumoroncologie, een zekere mate van homogeniteit van typen problemen en afwegingen, en van overzichtelijkheid van relevante kennis en ervaring, te creëren is die node wordt gemist wanneer in één besluitvormingsprocedure de middelenverdeling over de gehele gezondheidszorg zijn beslag zou moeten krijgen.***

* Cf. de laatste voetnoot bij hoofdstuk 4.

** Diezelfde dynamiek leidt ook tot snelle veranderingen in wat het aangewezen alternatief is waarmee, bijvoorbeeld in een KUA, een op haar doelmatigheid te beoordelen voorziening vergeleken dient te worden (Ger93, Rut00a). Alleen al die snelle veranderlijkheid van het te onderzoeken contrast, lijkt een belangrijk obstakel op de weg naar een prominente rol van de bewerkelijke, methodologisch veeleisende en data-intensieve KUA bij de afbakening van een collectief te financieren voorzieningenpakket.

*** Keerzijde van deze reductie van complexiteit is natuurlijk dat op die manier de verdeling van middelen over de verschillende deelgebieden niet geregeld is. Hier lijkt een onvervreembare verantwoordelijkheid te liggen van de landelijke overheid, die daaraan waarschijnlijk in samenwerking met de sectoren inhoud zal moeten geven.

Niettemin zal ook binnen deze deelgebieden het bereiken van overeenstemming over de indicaties waarbij voorzieningen vergoed moeten worden een enorme opgave zijn. Een hulpmiddel als de QALY zal daarbij waarschijnlijk goede diensten kunnen bewijzen. Het ligt immers voor de hand te veronderstellen dat de interpretatieproblemen die QALY's omgeven minder zwaar zullen wegen naarmate de diversiteit van de vergeleken voorzieningen, gezondheidsproblemen, patiëntenpopulaties en gezondheidsuitkomsten geringer is (Rus96). Zeker is dat het geval wanneer het erom gaat de doelmatigheid te vergelijken van de toepassing van eenzelfde voorziening in verschillende stadia van eenzelfde aandoening. Te denken valt aan de vergelijking van de doelmatigheid van een operatie vroeger of later in de ontwikkeling van prostaatklasten of van het gebruik van cholesterolsyntheseremmers afhankelijk van de kans op coronaire hartziekte. Overigens zullen bij doelmatigheidsvergelijkingen in zulke homogene domeinen QALY's juist vaak niet nodig zijn en zal volstaan kunnen worden met specifieke uitkomstmaten, zoals gewonnen levensjaren of ziektevrije levensjaren. Ook de SAVE en de WTP zullen wellicht bij de besluitvorming in de deelgebieden van nut kunnen zijn.

De legitimiteit van de besluitvorming in een bepaald zorggebied zal echter primair moeten berusten op een zorgvuldig vormgegeven besluitvormingsprocedure, waarin de partijen die dat deel van de gezondheidszorg van binnenuit kennen, zorgverleners en patiënten voorop, actief participeren. KUA en op termijn misschien ook *cost-benefit analysis* kunnen in dat besluitvormingsproces hulpmiddelen zijn om overzicht en overeenstemming te bereiken.

Literatuur

-
- Aar96 Aaronson NK, Cull AM, Kaasa S, e.a. The EORTC modular approach to quality of life assessment in oncology. In: Spilker B, red. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 179-89.
- Ana97 Anand S, Hanson K. Disability-adjusted life years: a critical review. *J Health Econ* 1997; 16: 685-702.
- Arn99 Arnesen T, Nord E. The value of DALY life: problems with ethics and validity of disability adjusted life years. *Br Med J* 1999; 319: 1423-5.
- Ash89 Ashmore M, Mulkay M, Pinch T. Health and efficiency. A sociology of health economics. Buckingham: Open University Press, 1989.
- Ash00 Ashton T, Cumming J, Devlin N. Priority-setting in New Zealand: translating principles into practice. *J Health Serv Res Policy* 2000; 5: 170-5.
- Bad01 Badia X, Roset M, Herdman M, e.a. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making* 2001; 21: 7-16.
- Bar98 Barendregt JJ, Bonneux L. Degenerative disease in an aging population. Models and conjectures (Proefschrift). Rotterdam: Erasmus Universiteit, 1998.
- Bar99 Barendregt JJ, Bonneux L. The trouble with health economics. *Eur J Public Hlth* 1999; 9: 309-12.
- Bel01 Bell CM, Chapman RH, Stone PW, e.a. An off-the-shelf help list: a comprehensive catalog of preferences scores from published cost-utility analyses. *Med Decis Making* 2001; 21: 288-94.
- Ber73 Berg RL, red. Health status indexes. Chicago: Hospital Research and Educational Trust, 1973.
- Ber76 Bergner M, Bobbitt RA, Kressel S, e.a. The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv* 1976; 6: 393-415.
- Ber89 Bergner M. Quality of life, health status, and clinical research. *Med Care* 1989; 27 (suppl): S148-56.
-

- Ber98 Berg M, Bezemer NM, van den Burg M. Normatieve aspecten van richtlijnen en kosteneffectiviteitsanalyses. Deelrapport 2 KNMG-project Passende medische zorg. Utrecht: KNMG, 1998.
- Bir99 Birch S, Gafni A, O'Brien B. Willingness to pay and the valuation of programmes for the prevention and control of influenza. *Pharmacoeconomics* 1999; 16 (suppl 1): 55-61.
- Ble95 Bleichrodt H. QALYs and HYE: Under what conditions are they equivalent? *J Health Econ* 1995; 14: 17-37.
- Ble96 Bleichrodt H. Applications of utility theory in the economic evaluation of health care (Proefschrift, Erasmus Universiteit Rotterdam). Ridderkerk: Ridderprint, 1996.
- Ble97a Bleichrodt H. Health utility indices and equity considerations. *J Health Econ* 1997; 16: 65-91.
- Ble97b Bleichrodt H, Quiggin J. Characterizing QALYs under a general rank dependent utility model. *J Risk Uncertainty* 1997; 15: 151-65.
- Ble97c Bleichrodt H, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. *J Risk Uncertainty* 1997; 15: 107-14.
- Ble97d Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of qalys. *J Health Econ* 16, 1997: 155-175.
- Ble00 Bleichrodt H, Brouwer W. Disconteren. In: Rutten-van Mólken MPMH, Van Bussbach JJ, Rutten FFH, red. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 123-9.
- Blei01 Bleichrodt H, Koopmanschap M. Economische evaluatie. In: Lapré R, Rutten F, Schut E, red. Algemene economie van de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2001: 273-94.
- Ble02 Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ* 2002; 11: 447-56.
- Boe98 Boers M. Kunst van doelmatig onderzoek (Inaugurele rede). Amsterdam: Vrije Universiteit, 1998.
- Boe02 Boer A. Onderzoek op maat. Een verkenning van factoren voor het gebruik van Technology Assessment. (Proefschrift). Rotterdam: Erasmus Universiteit, 2002.
- Bon91 Bonsel GJ. Methods of medical technology assessment with an application to liver transplantation (Proefschrift). Rotterdam: Erasmus Universiteit, 1991.
- Bon92 Bonsel GJ. QALY's: normatieve toepassing van gezondheidsmeting. In: Gunning-Scheepers LJ, Mootz M, red. Gezondheidsmeting. Houten: Bohn, 1992: 122-33.
- Bon93 Bonsel GJ. Kosteneffectiviteitsanalyse: over wikken en beschikken. In: Dillmann RJM, van Leeuwen E, Kingsma GK. Ethiek in de medische praktijk. Utrecht: Bunge, 1993: 167-78.
- Bon99 Bonneux L, Barendregt JJ. Ook met de consensus over verlaging van de serumcholesterolconcentratie blijft primaire preventie duur. *Ned Tijdschr Geneesk* 1999; 143: 772-5.
- Bon01a Bonsel G. Wat heet beter? (Inaugurele rede). Amsterdam: Vossiuspers UvA, 2001.
- Bon01b Bonneux L, Birnie E. The discount rate in the economic evaluation of prevention: a thought experiment. *J Epidemiol Community Health* 2001; 55: 123-25.
- Bor95 Borst-Eilers E. Beleidsbrief Medische Technology Assessment (MTA) en doelmatigheid van zorg. Tweede Kamer, 1995-96, 24126, nr 9.

- Bos00 Bosch JL, Hunink MGM. Comparison of the Health Utilities Index mark 3 (HUI3) and the EuroQol EQ-5D in patiënts treated for intermittent claudication. *Qual Life Res* 2000; 9: 591-601.
- Boy90 Boyd NF, Sutherland HJ, Heasman KZ, e.a. Whose utilities for decision analysis? *Med Decis Making* 1990; 10: 58-67.
- Bra99 Brazier J, Deverill M, Green C, e.a. A review of the use of health status measures in economic evaluation. *Health Technol Assessment* 1999; 3(9): 1-164.
- Bri95 Briggs A, Sculpher M. Sensitivity analysis in economic evaluation: a review of published studies. *Health Econ* 1995; 4: 355-71.
- Bri97 Briggs A, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Econ* 1997; 6: 327-40.
- Bri98a Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. *J Health Serv Res Policy* 1998; 3: 233-45.
- Bri98b Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ* 1998; 7: 723-40.
- Bri00 Briggs A. Handling uncertainty in cost-effectiveness models. *Pharmacoeconomics* 2000; 17: 479-500.
- Bro93 Broome J. Qalys. *J Public Economics* 1993; 50: 149-67.
- Bro95 Brock DW. Justice and the ADA: does prioritizing and rationing health care discriminate against the disabled? *Social Phil & Policy* 1995; 12: 159-85.
- Bro97a Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs measurement through quality of life? A response to the recommendation of the Washington Panel. *Health Econ* 1997; 6: 253-9.
- Bro97b Brouwer WBF, Koopmanschap MA, Rutten FFH. Productivity costs in cost-effectiveness analysis: numerator or denominator: a further discussion. *Health Econ* 1997; 6: 511-14.
- Bro00a Brouwer WBF, Koopmanschap MA. On the economic foundations of CEA. Ladies and gentlemen, take your positions! *J Health Econ* 2000; 19: 439-59.
- Bro00b Brouwer W, van Hout B, Rutten F. A fair approach to discounting future effects: taking a societal perspective. *J Health Serv Res Policy* 2000; 5: 114-8.
- Bro01 Brouwer WBF, Rutten FH. Health Economics. A bridge over troubled water. *Eur J Public Health* 2001; 11: 234-6.
- Bur98 van den Burg M, ter Meulen RHJ. Prioriteiten binnen de gezondheidszorg. Een vergelijkende studie naar normatieve overwegingen van prioriteitenstelling in Zweden, Engeland en Nederland. Deelrapport 1 KNMG-project Passende medische zorg. Utrecht: KNMG, 1998.
- Bus00 van Bussbach JJ. Waarderend kwaliteit van leven-onderzoek. In: Rutten-van Mólken MPMH, van Bussbach JJ, Rutten FFH, red. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 77-94.
- Cal87 Callahan D. Setting limits. Medical goals in an aging society. New York: Simon & Schuster, 1987.
- Cal90 Callahan D. What kind of life. The limits of medical progress. New York: Simon & Schuster, 1990.
- Car89 Carr-Hill RA. Assumptions of the QALY procedure. *Soc Sci Med* 1989; 29: 469-77.
- Car91 Carr-Hill RA. Allocation of resources to health care: is the QALY (Quality Adjusted Life Year) a technical solution to a political problem? *Int J Health Serv* 1991; 21: 351-63.
-

- CBO98 Centraal Begeleidingsorgaan voor de Intercollegiale Toetsing (CBO). Behandeling en preventie van coronaire hartziekten door verlaging van de plasmacholesterolconcentratie. Consensus Cholesterol, tweede herziening. Utrecht: CBO, 1998.
- CCO97 Canadian Coordination Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals: Canada. 2e editie. Ottawa: CCOHTA, 1997.
- Cul98 Culyer AJ. How ought health economists to treat value judgments in their analyses? In: Barer ML, Getzen TE, Stoddart GL, red. Health, health care and health economics: perspectives on distribution. Chicester: John Wiley & Sons, Ltd, 1998: 363-71.
- CVZ99 College voor zorgverzekeringen. Richtlijnen voor farmaco-economisch onderzoek. Amstelveen: CVZ, 1999.
- CVZ01 College voor zorgverzekeringen. Het basispakket: inhoud en grenzen. Amstelveen: CVZ, 2001.
- Dan88 Daniels N. Am I my parents' keeper? An essay on justice between the young and the old. New York: Oxford University Press, 1988.
- Dan93 Daniels N. Rationing fairly: programmatic considerations. *Bioethics* 1993; 7(2/3): 224-33.
- Dan97 Daniels N, Sabin J. Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philos Public Aff* 1997; 4: 303-50.
- Dan00 Daniels N. Accountability for reasonableness. Establishing a fair process for priority setting is easier than agreeing on principles. *Br Med J* 2000; 321: 1300-1.
- Day98 Day P, Klein R. The dilemmas of choice. *Odyssey* 1998; 4(2): 8-13.
- Dol96 Dolan P. The effect of experience of illness on health state valuations. *J Clin Epidemiol* 1996; 49: 551-64.
- Dol99 Dolan P. Whose preferences count? *Med Decis Making* 1999; 19: 482-6.
- Dol00 Dolan P. The measurement of health-related quality of life for use in resource allocation decisions in health care. In: Culyer AJ, Newhouse JP, red. *Handbook of health economics*. Volume 1. Amsterdam: Elsevier, 2000: 1724-60.
- Don88 Donaldson C, Atkinson A, Bond J, e.a. Should QALYs be programme-specific? *J Health Econ* 1988; 7: 239-57.
- Don97 Donaldson C, Shackley P. Does "proce utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med* 1997; 44: 699-707.
- Dru93 Drummond M, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? *Soc Sci Med* 1993; 37: 33-40.
- Dru97 Drummond MF, O'Brien B, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 2e editie. Oxford: Oxford University Press, 1997.
- Edg98 Edgar A, Salek S, Shickle D, e.a. *The Ethical Qaly: Ethical issues in healthcare resource allocations*. Haslemere: Euromed Communications, 1998.
- Enc97 van Enkevort P, Ten Vergert E, Geertsma B. Kosteneffectiviteitsanalyse van longtransplantatie. *Medisch Contact* 1997; 52: 439-41.
- Eme89 Emery DD, Schneiderman LJ. Cost-effectiveness analysis in health care. *Hastings Center Report* 1989; 19: 8-13.
-

- Ess95 Essink-Bot M L. Health status as a measure of outcome of disease and treatment (Proefschrift). Rotterdam: Erasmus Universiteit, 1995.
- Ess00 Essink-Bot ML. Beschrijvend kwaliteit van leven onderzoek. In: Rutten-van Mólken MPMH, van Bussbach JJ, Rutten FFH, red. Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg. Maarssen: Elsevier gezondheidszorg, 2000: 63-76.
- Eur90 EuroQuol Group. EuroQuol – a new facility for the measurement of health-related quality of life. Health Policy 1990; 16: 199-208.
- Fee96 Feeny DH, Torrance GW, Furlong WJ. Health Utilities Index. In: Spilker B, red. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 239-52.
- Fei87 Feinstein AR. Clinimetrics. New Haven: Yale University Press, 1987.
- Fis91 Fischhoff B. Value elicitation. Is there anything there? American Psychologist 1991; 46: 835-47.
- Fow95 Fowler FJ, Cleary PD, Massagli MP, e.a. Describing and measuring the values of health states. The role of reluctance to give up life in the measurement of the values of health states. Med Decis Making 1995; 15: 195-200.
- Fri80 Fries JF, Spitz P, Kraines RG, e.a. Measurement of patient outcome in arthritis. Arthritis Rheum 1980; 27: 1344-52.
- Fro89a Froberg DG, Kane RL. Methodology for measuring health-state preferences. I. Measurement strategies. J Clin Epidemiol 1989; 42: 675-85.
- Fro89b Froberg DG, Kane RL. Methodology for measuring health-state preferences. II. Scaling methods. J Clin Epidemiol 1989; 42: 675-85.
- Fro89c Froberg DG, Kane RL. Methodology for measuring health-state preferences. III. Population and context effects. J Clin Epidemiol 1989; 42: 675-85.
- Fro89d Froberg DG, Kane RL. Methodology for measuring health-state preferences. IV. Progress and research agenda. J Clin Epidemiol 1989; 42: 675-85.
- Fry93 Fryback DG, Dasbach EJ, Klein R, e.a. The Beaver Dam Health Outcomes Study: initial catalog of health-state quality factors. Med Decis Making 1993; 13: 89-102.
- Gab99 Gabriel SE, Kneeland TS, Melton LJ, e.a. Health related quality of life in economic evaluations for osteoporosis: Whose values should we use ? Med Decis Making 1999; 19: 141-8.
- Gaf95 Gafni A, Birch S. Preferences for outcomes in economic evaluation: an economic approach to addressing economic problems. Soc Sci Med 1995; 40: 767-76.
- Gar96 Garber AM, Weinstein MC, Torrance GW, Kamlet MS. Theoretical foundations of cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 25-53.
- Gar97 Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. J Health Econ 16, 1997: 1-31.
- Gar99 Garber AM. Realistic rigor in cost-effectiveness methods. Med Decis Making 1999; 19: 378-9.
- Gar00 Garber AM. Advances in cost-effectiveness analysis of health interventions. In: Culyer AJ, Newhouse JP, red. Handbook of health economics. Volume 1. Amsterdam: Elsevier, 2000: 181-221.
- Ger93 Gerard K, Mooney G. QALY league tables: handle with care. Health Econ 1993; 2: 59-64.

- Gli99 Glick HA, Polsky D, Willke RJ, e.a. A comparison of preference assessment instruments used in a clinical trial: Responses to the visual analog scale from the EuroQol EQ-5D and the Health Utilities Index. *Med Decis Making* 1999; 19: 265-75.
- Gol96a Gold MR, Patrick DL, Torrance GW, e.a. Identifying and valuing outcomes. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996: 82-134.
- Gol96b Gold MR, Siegel JE, Russell LB, Weinstein MC, red. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996.
- Goo96 Goold SD. Allocating health care: cost-utility analysis, informed democratic decision making, or the veil of ignorance? *J Health Politics Policy Law* 1996; 21: 69-98.
- GR86 Gezondheidsraad. *Grenzen van de gezondheidszorg*. Den Haag: Gezondheidsraad, 1986; publicatie nr 1986/29.
- GR91 Gezondheidsraad. Beraadsgroep Geneeskunde. *Medisch handelen op een tweesprong*. Den Haag: Gezondheidsraad, 1991; publicatie nr 1991/23.
- Gre01 Green C. On the societal value of health care: what do we know about the person trade-off technique? *Health Econ* 2001; 10: 233-43.
- Gre02 Greenberg D, Pliskin JS. Preference-based outcome measures in cost-utility analyses. A 20-year overview. *Int J Technol Assess* 2002; 18: 461-6.
- Gui99 Guillemin F. The value of utility: Assumptions underlying preferences and quality adjusted life years. *J Rheumatol* 1999; 26: 1861-3.
- Guy96 Guyatt GH, Jeaschke R, Feeny DH, e.a. Measurements in clinical trials: choosing the right approach. In: Spilker B, red. *Quality of life and pharmacoeconomics in clinical trials*. 2e editie. Philadelphia: Lippincott-Raven, 1996: 41-8.
- Had91 Hadorn DC. Setting health care priorities in Oregon. Cost-effectiveness meets the rule of rescue. *JAMA* 1991; 265: 2218-25.
- Hal92 Hall J, Gerard K, Salkeld G, e.a. A cost utility analysis of mammography screening in Australia. *Soc Sci Med* 1992; 34: 993-1004.
- Ham97 Ham C. Priority setting in health care: learning from international experience. *Health Policy* 1997; 42: 49-66.
- Har87 Harris J. QALYfying the value of life. *J Med Ethics* 1987; 13: 117-23.
- Har01 Hart W, Walvoort HC. 'Troosten: altijd'; codoleancebetuigingen van artsen aan de nabestaanden van hun patiënten. *Ned Tijdschr Geneesk* 2001; 145: 2185-7.
- Hey90 Heyink J. Adding years to your life or adding life to your years. QALY (Quality Adjusted Life Year): theoretical and practical issues. *Int J Health Sciences* 1990; 1: 45-9.
- Hol95 Hollingworth W, Mackenzie R, Todd CJ, e.a. Measuring changes in quality of life following magnetic resonance imaging of the knee: SF-36, EuroQol or Rosser index? *Qual Life Res* 1995; 4: 325-34.
- Hol98 Holm S. Goodbye to the simple solutions: the second phase of priority setting in health care. *Br Med J* 1998; 317: 1000-2.
-

- Hor92 Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *J Clin Epidemiol* 1992; 45: 505-12.
- Hou98 van Hout BA. Discounting costs and effects: a reconsideration. *Health Econ* 1998; 7: 581-594.
- Hou01 van Hout BA. *Geen Sinterklaas zonder Zwarte Piet. Over keuzen in de gezondheidszorg (Inaugurele rede)*. Utrecht: Universiteit Utrecht, 2001.
- Hun80 Hunt SM, McEwen J. The development of a subjective health indicator. *Sociol Health Illness* 1980; 2: 231-46.
- Hun98a Hunt SM. Cross-cultural issues in the use of quality of life measures in randomized controlled trials. In: Staquet MJ, Hays RD, Fayers PM, red. *Quality of life assessment in clinical trials*. Oxford: Oxford University Press, 1998: 51-67.
- Hun98b Hunink MGM, Bult JR, de Vries J, e.a. Uncertainty in decision models analyzing cost-effectiveness. *Med Decis Making* 1998; 18: 337-46.
- Hur98 Hurley J. Welfarism, extra-welfarism and evaluative economic analysis in the health sector. In: Barer ML, Getzen TE, Stoddart GL, red. *Health, health care and health economics: perspectives on distribution*. Chichester: Wiley, 1998: 373-95.
- Hur00 Hurley J. An overview of the normative economics of the health sector. In: Culyer AJ, Newhouse JP, red. *Handbook of health economics*. Vol 1. Amsterdam: Elsevier, 2000: 56-118.
- Hut96 Hutchinson JF. Quality of life in ethnic groups. In: Spilker B, red. *Quality of life and pharmacoeconomics in clinical trials*. 2e editie. Philadelphia: Lippincott-Raven, 1996: 587-93.
- Joh96 Johannesson J, Jönsson B, Karlsson G. Outcome measurement in economic evaluation. *Health Econ*, 1996; 5: 279-96.
- Joh97 Johannesson M, Meltzer D, O'Connor RM. Incorporating future costs in medical cost-effectiveness analysis: implications for the cost-effectiveness of the treatment of hypertension. *Med Decis Making* 1997; 17: 382-9.
- Jon86 Jonsen AR. Bentham in a box: technology assessment and health care allocation. *Law Med Health Care* 1986; 14(3-4): 172-4.
- Kah79 Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica* 1979; 47: 263-91.
- Kar48 Karnofsky DA, Abelman WH, Craver LF, e.a. The use of nitrogen mustards in the palliative treatment of carcinoma. *Cancer* 1948; 1: 634-56.
- Kar49 Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: MacLeod CM, red. *Evaluation of chemotherapeutic agents*. New York: Columbia University Press, 1949: 191-205.
- Kat63 Katz S, Ford AB, Moskowitz RW, e.a. Studies of illness in the aged. The index of ADL. *JAMA* 1963; 185: 914-9.
- Kee83 Keeler EB, Cretin S. Discounting of life-saving and other nonmonetary effects. *Management Science* 1983; 29: 300-6.
- Keu91 Commissie Keuzen in de zorg. *Kiezen en delen*. 1. Den Haag: DOP, 1991.
- Kin96 Kind P. The EuroQol instrument: an index of health-related quality of life. In: Spilker B, red. *Quality of life and pharmacoeconomics in clinical trials*. 2e editie. Philadelphia: Lippincott-Raven, 1996: 191-201.

- Kir00 Kirsch J, McGuire A. Establishing health state valuations for disease specific states: an example from heart disease. *Health Econ* 2000; 9: 149-58.
- Kle93 Klein R. Health care reform: the global search for Utopia. *Br Med J* 1993; 307: 752.
- Kle98 Klein R. Puzzling out priorities. Why we must acknowledge that rationing is a political process. *Br Med J* 1998; 317: 959-60.
- Klo99 Klose T. The contingent valuation method in health care. *Health Policy* 1999; 47: 97-123.
- Kon91 de Koning HJ, van Ineveld BM, van Oortmarssen GJ, e.a. Breast cancer screening and cost-effectiveness: Policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer* 1991; 49: 531-7.
- Koo94 Koopmanschap MA, Rutten FFH. The impact of indirect costs on outcomes of health care programs. *Health Econ* 1994; 3: 385-93.
- Koo95 Koopmanschap MA, Rutten FFH, Ineveld BM, e.a. The friction cost method for measuring indirect costs of disease. *J Health Econ* 1995; 14: 171-89.
- Koo98 Koopmanschap MA, Rutten FFH. Berekening van kosten van zorg. Vaak onderschat in economische evaluatiestudies. *Tijdschr Gezondheidswetens* 1998; 76: 83-8.
- Kor99 Korczyn AD, Davidson M. Quality of life in Alzheimer's disease. *Eur J Neurol* 1999; 6: 487-9.
- Kra93 Krahn M, Gafni A. Discounting in the economic evaluation of health care interventions. *Med Care* 1993; 31: 403-18.
- Kra98a Krabbe PFM, Bonsel GJ. Sequence effects, health profiles, and the QALY model: in search of realistic modelling. *Med Decis Making* 1998; 18: 178-86.
- Kra98b Krabbe PFM. The valuation of health outcomes. A contribution to the QALY approach (Proefschrift). Rotterdam: Erasmus Universiteit, 1998.
- Lee97 Lee SJ, Neumann PJ, Churchill WH, e.a. Patients' willingness to pay for autologous blood donation. *Health Policy* 1997; 40: 1-12.
- Leg02 Legemaate J. De (dubbel)rol van de arts. In: *Zorg, schaarste en recht. Preadvies uitgebracht ten behoeve van het Jubileumsymposium van de Vereniging voor Gezondheidsrecht*. Vereniging voor Gezondheidsrecht, 23 mei 2002: 34-63.
- Len99 Lenert LA, Treadwell JR, Schwartz CE. Associations between health status and utilities. Implications for policy. *Med Care* 1999; 37: 479-89.
- Len00 Lenert L, Kaplan RM. Validity and Interpretation of preference-based measures of health-related quality of life. *Med Care* 2000; 38 (suppl II): II-138-145.
- Lep97 Leplège A, Hunt S. The problem of quality of life in medicine. *JAMA* 1997; 278: 47-50.
- Leu93 Leussink AB. Kiezen in de zorg: het meten van het onmeetbare. QALY of SAVE. *Graadmeter* 1993; 9: 5-7.
- Lip96 Lipscomb J, Weinstein MC, Torrance GW. Time preference. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996: 214-46.
- Lle02 Llewellyn-Thomas HA, Arshinoff R, Bell M, e.a. Healthy year equivalents in major joint replacement. Can patients provide meaningful responses? *Int J Technol Assess Health Care* 2002; 18: 467-84.
- Loo89 Loomes G, McKenzie L. The use of QALYs in health care decision making. *Soc Sci Med* 1989; 28: 299-308.
-

- Luc96 Luce BR, Manning WG, Siegel JE, e.a. Estimating costs in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 176-213.
- Maa95 van der Maas PJ. Volksgezondheidsonderzoek. In: van der Maas PJ, Mackenbach JP, red. Volksgezondheid en gezondheidszorg. Utrecht: Bunge, 1995: 17-75.
- Mac01 Macran S, Kind P. "Death" and the valuation of health-related quality of life. *Med Care* 2001; 39: 217-27.
- Man96a Mandelblatt JS, Fryback DG, Weinstein MC, e.a. Assessing the effectiveness of health interventions. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 135-75.
- Man96b Manning WG, Fryback DG, Weinstein MC. Reflecting uncertainty in cost-effectiveness analysis. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996: 247-75.
- McK98 McKie J, Richardson J, Singer P, e.a. The allocation of health care resources. An ethical evaluation of the 'Qaly' approach. Aldershot: Ashgate/Dartmouth, 1998.
- Meh89 Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-year equivalents. *Med Decis Making* 1989; 9: 142-9.
- Mel75 Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975; 1: 277-99.
- Mel97 Meltzer D. Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ* 1997; 16: 33-64.
- Mel99a Meltzer D, Johannesson M. Inconsistencies in the "societal perspective" on costs of the Panel on cost-effectiveness in health and medicine. *Med Decis Making* 1999; 19: 371-7.
- Mel99b Meltzer D, Johannesson M. On the role of theory in cost-effectiveness analysis – a response to Garber, Russell, and Weinstein. *Med Decis Making* 1999; 19: 383-4.
- Mel00 Meltzer D, Egleston B, Stopffel D, e.a. Effect of future costs on cost-effectiveness of medical interventions among young adults. The example of intensive therapy for type 1 diabetes mellitus. *Medical Care* 2000; 38: 679-85.
- Men90 Menzel PT. Strong medicine. The ethical rationing of health care. New York, Oxford: Oxford University Press, 1990.
- Men99 Menzel P, Gold MR, Nord E, e.a. Toward a broader view of values in cost-effectiveness analysis of health. *Hastings Center Report* 1999; May-June: 7-15.
- Men02 Menzel P, Dolan P, Richardson J, e.a. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Soc Sci Med* 2002; 55: 2149-58.
- Miy85 Miyamoto JM, Eraker SA. Parameter estimates for a QALY utility model. *Med Decis Making* 1985; 5: 191-213.
- Moo94a Mooney G. What else do we want from our health services? *Soc Sci Med* 1994; 39: 151-4.
- Moo94b Mooney G. Key issues in health economics. New York: Harvester Wheatsheaf, 1994.
- Moo00 Mooney G, Wiseman V. Burden of disease and priority setting. *Health Econ* 2000; 9: 369-72.
- Mul97 Mulder JH. Terug naar de zorg. Dunning gekanteld. *Med Contact* 1997; 52: 1219-20.

- Mur96 Murray CJL, Lopez AD. The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Cambridge, MA: Harvard University Press, 1996.
- Mur97 Murray CJL, Acharya AK. Understanding DALYs. *J Health Econ* 1997; 16: 703-30.
- Mus87 Musschenga AW, Borst-Eilers E. Ethische criteria voor prioriteitenstelling in de gezondheidszorg. In: Musschenga AW, de Neeling JND, red. *Verdeling van schaarse middelen in de gezondheidszorg*. Amsterdam: VU Uitgeverij, 1987: 101-22.
- Nea95 Nease RF, Kneeland T, O'Connor GT, e.a. Variation in patient utilities for outcomes of the management of chronic stable angina: implications for clinical practice guidelines. *JAMA* 1995; 272: 1185-90.
- Nee91 de Neeling JND. *Quality of life. Het onderzoek naar welzijnseffecten van medische behandelingen*. Utrecht: Bunge, 1991.
- Nee92 de Neeling JND. Kwaliteit van leven: een meetbaar begrip? *Medisch Contact* 1992; 47: 143-5.
- Nee00 de Neeling JND, namens de projectgroep Passende Medische Zorg. *Passende medische zorg. Over keuzen en richtlijnen. Eindrapport KNMG-project Passende medische zorg*. Utrecht: KNMG, 2000.
- Neu44 von Neumann J, Morgenstern O. *Theory of games and economic behavior*. Princeton: Princeton University Press, 1953 (1944).
- Neu94 Neumann PJ, Johannesson M. The willingness to pay for in vitro fertilization: a pilot study using contingent valuation. *Med Care* 1994; 32: 686-99.
- Neu97 Neumann PJ, Zinner DE, Wright JC. Are methods for estimating QALYs in cost-effectiveness analyses improving? *Med Decis Making* 1997; 17: 402-8.
- Neu00a Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health* 2000; 21: 587-611.
- Neu00b Neumann PJ, Sandberg EA, Araki SS, e.a. A comparison of HUI2 and HUI3 utility scores in Alzheimer's disease. *Med Decis Making* 2000; 20: 413-22.
- Nor92a Nord E. An alternative to QALYs: the saved young life equivalent (SAVE). *Br Med J* 1992; 305: 875-7.
- Nor92b Nord E. Methods for quality adjustment of life years. *Soc Sci Med* 1992; 34: 559-69.
- Nor93 Nord E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. *Health Policy* 1993; 24: 227-38.
- Nor94 Nord E. The QALY – a measure of social value rather than individual utility? *Health Econ* 1994; 3: 89-93.
- Nor95 Nord E. The person-trade-off approach to valuing health care programs. *Med Decis Making* 1995; 15: 201-8.
- Nor99 Nord E, Pinto JL, Richardson J, e.a. Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Econ* 1999; 8: 25-39.
- Nor01 Nord E. The desirability of a condition versus the well being and worth of a person. *Health Econ* 2001; 10: 579-81.
- OBr96 O'Brien B, Gafni A. When do the "dollars" make sense? Toward a conceptual framework for contingent valuation studies in health care. *Med Decis Making* 1996; 16: 288-99.
- Ols01 Olsen JA, Smith RD. Theory versus practice: a review of 'willingness-to-pay' in health and health care. *Health Econ* 2001; 10: 39-52.
-

- Oos00 Oostenbrink JB, Koopmanschap MA, Rutten FFH. Handleiding voor kostenonderzoek. Methoden en richtlijnrijzen voor economische evaluaties in de gezondheidszorg. Amstelveen: College voor Zorgverzekeringen, 2000.
- Pat89 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989; 27 (suppl): S217-32.
- Pin97 Pinto Prades JL. Is the person trade-off a valid method for allocating health care resources? *Health Econ* 1997; 6: 71-81.
- Pli80 Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. *Opns Res* 1980; 28: 206-24.
- Pol02 Poley MJ, Stolk EA, Brouwer WBF, van Busschbach JJ. Ziekte last als uitwerking van het criterium 'noodzakelijkheid' bij het maken van keuzen in de zorg. *Ned Tijdschr Geneesk* 2002; 146: 2312-5.
- Pro00 Prosser LA, Stinnett AA, Goldman PA, e.a. Cost-effectiveness of cholesterol-lowering therapies according to selected patient characteristics. *Ann Int Med* 2000; 132: 769-79.
- Raw89 Rawles J. Castigating QALYs. *J Med Ethics* 1989; 15: 143-7.
- Rea84 Read JL, Quinn RJ, Berwick DM, e.a. Preferences for health outcomes: comparisons of assessment methods. *Med Decis Making* 1984; 4: 315-29.
- RG001 Raad voor Gezondheidsonderzoek. Rapport van het Platform HTA. Bijlage 1. Rapport van de commissie Methodologie. Den Haag: Raad voor Gezondheidsonderzoek, 2001.
- Ric94 Richardson J. Cost utility analysis: what should be measured? *Soc Sci Med* 1994; 39: 7-21.
- Ric96 Richardson J, Hall J, Salkeld G. The measurement of utility in multiphase health states. *Int J Technol Assess Health Care* 1996; 12: 151-62.
- Rie98 Ried W. QALYs versus HYE's – what's right and what's wrong. A review of the controversy. *J Health Econ* 1998; 17: 607-25.
- Rig86 Rigger H. Wie volgt? Over selectie en wachtlijsten in de gezondheidszorg. Den Haag: Gezondheidsraad 1986; publicatie nr 1986/28.
- Rus96 Russell LB, Siegel JE, Daniels N, e.a. Cost-effectiveness analysis as a guide to resource allocation in health: roles and limitations. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, red. *Cost-effectiveness in health and medicine*. New York: Oxford University Press, 1996: 3-24.
- Rus99 Russell LB. Improving the panel's recommendations. *Med Decis Making* 1999; 19: 379-80.
- Rut98 Rutten FFH, van Busschbach JJ, van Hout BA, e.a. Economische evaluatie van gezondheidszorgprogramma's. Principes en instrumentarium. *Tijdschr Gezondheidswetens* 1998; 76: 74-82.
- Rut00a Rutten FFH. Voor de praktijk. Dwalingen in de methodologie. XIX. Kosteneffectiviteitsratio's. *Ned Tijdschr Geneesk* 2000; 144: 1311-4.
- Rut00b Rutten FFH, Oostenbrink JB, Koopmanschap MA. Directe kosten van gezondheidszorg. In: Rutten-van Mólken MPMH, Van Busschbach JJ, Rutten FFH, red. *Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg*. Maarssen: Elsevier gezondheidszorg, 2000: 31-41.
- Sac78 Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. *J Chron Dis* 1978; 31: 697-704.

- Sas01 Sassi F, Archard L, Le-Grand J. Equity and the economic evaluation of healthcare. *Health Technol Assessment* 2001; 5(3): 1-138.
- Sen91 Sen A. Utility. Ideas and terminology. *Economics and Philosophy* 1991; 7: 277-83.
- Shi97 Shiell A. Health outcomes are about choices and values: an economic perspective on the health outcomes movement. *Health Policy* 1997; 39: 5-15.
- Slo95 Slovic P. The construction of preference. *American Psychologist* 1995; 50: 364-71.
- Smi87 Smith A. Qualms about QALY's. *Lancet* 1987; May 16: 1134-6.
- Smi93 Smith R, Dobson M. Measuring utility values for QALYs: two methodological issues. *Health Econ*, 1993; 2: 349-55.
- Spi90 Spilker B, Molinek FR, Johnston KA, e.a. Quality of life bibliography and indexes. *Med Care* 1990; 28 (suppl): DS1-77.
- Spi96 Spilker B, red. *Quality of life and pharmacoeconomics in clinical trials*. 2e editie. Philadelphia: Lippincott-Raven, 1996.
- Ste00 Stewart AL, Napoles-Springer A. Health-related quality-of-life assessments in diverse population groups in the United States. *Med Care* 2000; 38 (suppl II): 102-24.
- Sto00a Stouthard MEA, Essink-Bot ML, Bonsel GJ, e.a. Disability weights for diseases. A modified protocol and results for a Western European region. *Eur J Public Health* 2000; 10: 24-30.
- Sto00b Stone PW, Chapman RH, Sandberg EA, e.a. Measuring costs in cost-utility analyses. Variations in the literature. *Int J Technol Assessment Health Care* 2000; 16: 111-24.
- Sug78 Sugden R, Williams A. *The principles of practical cost-benefit analysis*. Oxford: Oxford University Press, 1978.
- Tor86 Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Econ* 1986; 5: 1-30.
- Tor89 Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technology Assessment Health Care*, 1989; 5: 559-75.
- Tor99 Torgersen D, Raftery J. Measuring outcomes in economic evaluations. *Br Med J* 1999; 318: 1413.
- Tsu00 Tsuchiya A. QALYs and ageism: philosophical theories and age weighting. *Health Econ* 2000; 9: 57-68.
- Ube99 Ubel PA, Richardson J, Prades J LP. Life-saving treatments and disabilities. Are all QALYs created equal? *Int J Technol Assessment Health Care* 1999; 15: 738-48.
- Ube00a Ubel PA. *Pricing life. Why it's time for health care rationing*. Cambridge, MA: MIT Press, 2000.
- Ube00b Ubel PA, Richardson J, Menzel P. Societal value, the person trade-off, and the dilemma of whose values to measure for cost-effectiveness analysis. *Health Econ* 2000; 9: 127-36.
- Ube00c Ubel PA, Nord E, Gold M, e.a. Improving value measurement in cost-effectiveness analysis. *Med Care* 2000; 38: 892-901.
- Uyl00 Uyl-de Groot CA. Uitgangspunten en opzet van onderzoek. In: Rutten-van Mólken MPMH, van Bussbach JJ, Rutten FFH, red. *Van kosten tot effecten. Een handleiding voor evaluatiestudies in de gezondheidszorg*. Maarssen: Elsevier gezondheidszorg, 2000: 17-30.
- Ver01 Verkerk MA, van Busschbach JJ, Karssing ED. Health-related quality of life research and the capability approach of Amartya Sen. *Qual Life Res* 2001; 10: 49-55.
- Vle90 Vlek CAJ. *Beslissen over risico-acceptatie*. Den Haag: Gezondheidsraad, 1990; publicatie nr A90/10.
-

- Wag00 Wagemakers A, van der Wijk P, Francissen P, red. Wanneer is het leven te kostbaar? Leeftijd en kosteneffectiviteit in de gezondheidszorg. Amsterdam: Boom, 2000.
- Wag91 Wagstaff A. QALYs and the equity-efficiency trade-off. *J Health Econ* 1991; 10: 21-41.
- War96 Ware JE Jr. The SF-36 health survey. In: Spilker B, red. Quality of life and pharmacoeconomics in clinical trials. 2e editie. Philadelphia: Lippincott-Raven, 1996: 337-45.
- Wei77 Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977; 296: 716-21.
- Wei80 Weinstein MC, Fineberg HV, red. Clinical decision analysis. Philadelphia: WB Saunders, 1980.
- Wei88 Weinstein MC. A QALY is a QALY is a QALY – or is it? *J Health Econ* 1988; 7: 289-90.
- Wei96 Weinstein MC, Siegel JE, Gold MR, e.a. Recommendations of the panel on cost-effectiveness in health and medicine. *JAMA* 1996; 276: 1253-8.
- Wei97a Weinstein MC, Siegel JE, Garber AM, e.a. Productivity costs, time costs and health-related quality of life: a response to the Erasmus group. *Health Econ* 1997; 6: 505-10.
- Wei97b Weinstein MC, Manning WG. Theoretical issues in cost-effectiveness analysis. *J Health Econ* 1997; 16: 121-8.
- Wei99 Weinstein MC. Theoretically correct cost-effectiveness analysis. *Med Decis Making* 1999; 19: 381-2.
- Wil81 Williams A. Welfare economics and health status measurement. In: van der Gaag J, Perlman M, red. Health, economics and health economics. Amsterdam: North Holland, 1981.
- Wil92 Williams A. Cost-effectiveness analysis: is it ethical? *J Med Ethics* 1992; 18: 7-11.
- Wil93 Williams A. Cost-benefit analysis: applied welfare economics or general decision aid. In: Williams A, Giardina E, red. Efficiency in the public sector. London: Edward Elgar, 1993.
- Wil96 Williams A. QALYs and ethics: a health economist's perspective. *Soc Sci Med* 1996; 43: 1795-1804.
- Wil00 Williams A, Cookson R. Equity in health. In: Culyer AJ, Newhouse JP, red. Handbook of health economics. Volume 1. Amsterdam: Elsevier, 2000: 1863-1910.
- Wit00 de Wit GA, van Busschbach JJ, de Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ* 2000; 9: 109-26.
- WRR97 Wetenschappelijke Raad voor het Regeringsbeleid. Volksgezondheidszorg. Den Haag: Sdu, 1997.
- Zec76 Zeckhauser R, Shepard D. Where now for saving lives? *Law and Contemporary Problems* 1976; 40: 5-45.

A Verantwoording

Bijlage

Verantwoording

Dit rapport is in concept besproken in de volgende Gezondheidsraadcommissies:

- Commissie Contouren van het Basispakket
- Kerncommissie MTA
- Beraadsgroep Geneeskunde
- Beraadsgroep Gezondheidsethiek en Gezondheidsrecht

Voorts hebben de volgende externe deskundigen het concept becommentarieerd:

- dr H Bleichrodt, econoom; Erasmus Universiteit Rotterdam
- dr JJ van Busschbach, psycholoog; Erasmus Universiteit Rotterdam
- prof. dr BA van Hout, econoom; Universitair Medisch Centrum Utrecht
- prof. dr FFH Rutten, econoom; Erasmus Universiteit Rotterdam

De verantwoordelijkheid voor de tekst berust geheel bij de schrijver, die secretaris is van de Beraadsgroep Geneeskunde.

