



Zorginstituut Nederland

Richtlijn

voor het uitvoeren van economische evaluaties in de gezondheidszorg

Herziene versie 2024

| Van goede zorg verzekerd |

Inhoudsopgave

Voorwoord	4
Afkortingen	5
Inleiding	7
1 Standaardanalyse	9
2 Raamwerk van de economische evaluatie	11
2.1 Doelstelling en gebruiker	11
2.2 Perspectief	11
2.3 PICOTS	11
2.3.1 Patient (P)	11
2.3.2 Intervention (I)	11
2.3.3 Comparison (C)	11
2.3.4 Outcomes (O)	12
2.3.5 Time (T)	12
2.3.6 Setting (S)	12
2.4 Analysetechniek	12
3 Gegevens	14
3.1 Effectiviteit	14
3.1.1 Systematisch literatuuronderzoek	14
3.1.2 Studiedesigns	14
3.1.3 Indirecte vergelijkingen	14
3.2 Kosten	15
3.2.1 Soorten kosten en volumemeting	15
3.2.2 Waardering	17
3.2.3 Inflatiecorrectie/prijsindexering	17
3.3 Kwaliteit van leven	17
3.3.1 Meten van kwaliteit van leven	17
3.3.2 Disutiliteiten	18
3.3.3 Kwaliteit van leven van mantelzorgers	18
3.4 Raadplegen van experts	18
3.4.1 Expert opinion	19
3.4.2 Expert elicitation	19
4 Methode	20
4.1 Mogelijke onderzoeksdesigns	20
4.1.1 Empirische economische evaluaties	20
4.1.2 Modelmatige economische evaluaties	20
4.2 Disconteren	20
4.3 Missende data	21
4.4 Corrigeren voor bias	21
4.4.1 Corrigeren voor bias in gerandomiseerde studies	21
4.4.2 Corrigeren voor bias in niet-gerandomiseerde studies	22
4.5 Extrapolatietechnieken	22
4.6 Subgroepanalyses	23
4.7 Onzekerheidsanalyses	24
4.7.1 Onzekerheidsanalyses in empirische economische evaluaties	24
4.7.2 Onzekerheidsanalyses in modelmatige economische evaluaties	24
4.8 Validatie	27
4.8.1 Validatie bij een modelmatige economische evaluatie	27

4.8.2	Validatie van een empirische economische evaluatie	28
5	Rapportage	29
5.1	Rapportage van gegevens	29
5.1.1	Effectiviteit	29
5.1.2	Kosten	29
5.1.3	Kwaliteit van leven	29
5.1.4	Expertraadpleging	29
5.2	Rapportage van de methode	29
5.2.1	Missende data	29
5.2.2	Extrapolaties	30
5.2.3	Validatie	30
5.3	Rapportage van de resultaten	30
5.3.1	Base-case analyse	30
5.3.2	Subgroepanalyses	31
5.3.3	Deterministische en scenarioanalyses	31
5.3.4	Visualisatie van probabilistische analyse	31
5.3.5	Value of information	31
6	Aandachtspunten	33
6.1	Sequentiële elementen	33
6.2	Stapsgewijze innovatie	33
6.3	Leercurve	33
6.4	Alternatieve uitkomsten	33
7	Referenties	35
	Bijlage I	40



Voorwoord

Sinds de eerste publicatie van de Nederlandse richtlijn voor farmaco-economisch onderzoek, vijftig jaar geleden, hebben zowel de wetenschap als het Zorginstituut niet stilgestaan. Actualisaties van die eerste richtlijn volgden in 2006 en 2016, telkens ingegeven door nieuwe wetenschappelijke inzichten, bijvoorbeeld over de rapportage van onzekerheid rondom de uitkomsten van economische evaluaties. Binnen het Zorginstituut is het belang van kosteneffectiviteit bij de advisering over de inhoud van het basispakket in de loop van de jaren steeds groter geworden. Kosteneffectiviteit is een van de vier pakketcriteria die gehanteerd worden. Omdat het Zorginstituut de verantwoordelijkheid heeft voor systematische en transparante afwegingen rondom het basispakket, is het belang van standaardisatie van economische evaluaties over de jaren steeds groter geworden. In de voorlaatste versie van de richtlijn uit 2016, werd daarom voor het eerst een 'standaardanalyse' (reference case) geformuleerd.

Economische evaluaties die voldoen aan de voorgeschreven standaardanalyse kunnen onderling vergeleken worden op dit pakketcriterium. In deze derde actualisatie van de richtlijn wordt de standaardanalyse weer verder aangescherpt, bijvoorbeeld ten aanzien van de zorgkosten in gewonnen levensjaren, die nu voor het eerst integraal onderdeel vormen van de kosteneffectiviteitsanalyse. Wat al die jaren onveranderd is gebleven is de keuze voor het maatschappelijke perspectief van de economische evaluatie. Deze keuze is door de richtlijncommissie ook deze keer niet ter discussie gesteld. In internationaal verband is Nederland een van de landen die hierin traditioneel gezien uitgesproken keuzes heeft gemaakt, omdat weinig landen zo'n veelomvattend maatschappelijk perspectief hanteren. De keuze voor dit perspectief is ook mogelijk omdat Nederlandse wetenschappers veel onderzoek doen naar het meten en waarderen van maatschappelijke kosten en effecten, zodat deze ook daadwerkelijk meegenomen kunnen worden in de economische evaluatie.

De richtlijncommissie bestond uit acht leden die allen goed bekend zijn met de beoordelingsprocedures van het Zorginstituut. Deze commissie werd ondersteund door deskundige medewerkers van het Zorginstituut. Hoewel de richtlijn uitgegeven wordt door het Zorginstituut, heeft de commissie bij alle keuzes ook de vele economische evaluaties die in Nederland uitgevoerd worden buiten het bestek van het Zorginstituut in het achterhoofd gehouden. Nederland is een land waar het gezondheidseconomische vakgebied al lang geleden tot bloei kwam en op hoog niveau beoefend wordt. Onderzoek naar doelmatigheid van nieuwe medische ontwikkelingen wordt bijvoorbeeld in alle UMC's uitgevoerd en is vaak een voorwaarde voor toekenning van onderzoekssubsidies door ZonMw of de onderzoeksprogramma's van de Europese Unie. De richtlijncommissie heeft de ambitie en de verwachting dat deze herziene richtlijn ook voor dergelijke economische evaluaties gebruikt kan worden.

Met deze geactualiseerde richtlijn kunnen zowel het Zorginstituut als Nederlandse gezondheidseconomen de komende jaren blijven bijdragen aan een zorgvuldige onderbouwing van het gezondheidsbeleid en daarmee aan passende zorg.

Namens de commissie richtlijnherziening,

Prof. dr. Ardine de Wit
Voorzitter commissie richtlijnherziening

Afkortingen

AdVISHE	Assessment of the Validation Status of Health-Economic decision models
AIC	Akaike information criterion
BIC	Bayesian information criterion
CEAC	Cost-effectiveness acceptability curve
CE-plane	Cost-effectiveness plane
CPI	Consumentenprijsindex
CVZ	College voor Zorgverzekeringen
DCE	Discrete choice experiment
ENBS	Expected net benefit of sampling
EQ-5D	EuroQol-5 dimensions
EVPI	Expected value of perfect information
EVPII	Expected value of partial perfect information
EVSI	Expected value of sample information
GCP	Good clinical practice
GRADE	Grading of recommendations, assessment, development and evaluations
ICER	Incremental cost-effectiveness ratio
IKB	Intersectorale kosten en baten
ISPOR	The International Society for Pharmacoeconomics and Outcomes Research
ITC	Indirect treatment comparison
KEA	Kosteneffectiviteitsanalyse
KNAW	Koninklijke Nederlandse Akademie van Wetenschappen
KUA	Kostenutiliteitsanalyse
MAR	Missing at random
MAIC	Matching adjusted indirect comparison
MCAR	Missing completely at random
MCDA	Multi-criteria decision analysis
MKBA	Maatschappelijke kostenbatenanalyse
MTC	Mixed treatment comparison
NMA	Netwerk meta-analyse
NMB	Net monetary benefit
MNAR	Missing not at random
PAID	Practical Application to Include future Disease costs

PICOTS	Patient, Intervention, Comparison, Outcome, Time, Setting
QALY	Quality adjusted life year
RCT	Randomised controlled trial
SHELF	Sheffield elicitation framework
STC	Simulated treatment comparison
VOI	Value of Information

Inleiding

Doel van de richtlijn

Deze richtlijn heeft als doel om duidelijke kaders te stellen voor uitvoerders van economische evaluaties in de gezondheidszorg. Daarbij gaat het specifiek om economische evaluaties bedoeld voor besluitvorming over de invoering of vergoeding van nieuwe zorginterventies, of besluitvorming over het stoppen van de vergoeding van bestaande zorginterventies. Het uitgangspunt bij het opstellen van deze richtlijn is daarom een besiskundige benadering: methodologische keuzes worden gebaseerd op het uitgangspunt dat uiteindelijk een beslissing kan worden genomen binnen de context van de Nederlandse zorgsector.

Historie

In 1999 heeft het toenmalige College voor Zorgverzekeringen (CVZ) de richtlijnen voor farmaco-economisch onderzoek uitgebracht. In 2006 is dit document herzien en sindsdien werd deze gebruikt als beoordelingskader voor farmaco-economische evaluaties die onderdeel werden van vergoedingsdossiers bij geneesmiddelen. Aan de hand van de versie uit 2006 werd nagegaan of de doelmatigheid van een geneesmiddel, waarvoor vergoeding werd aangevraagd, voldoende was onderbouwd. In 2008 werd de leidraad uitkomstenonderzoek gepubliceerd en in 2010 de handleiding voor kostenonderzoek (ook wel: kostenhandleiding), waardoor er drie richtlijnen beschikbaar waren. In 2016 zijn deze documenten samengebracht in één nieuwe richtlijn: de richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. In tegenstelling tot de drie losse richtlijnen, was de richtlijn van 2016 niet alleen gericht op geneesmiddelen, maar kwam er ook aandacht voor andere soorten zorginterventies. Daarnaast werd de richtlijn aangepast naar de laatste wetenschappelijke ontwikkelingen.

Wegens continue methodologische ontwikkelingen op het gebied van economische evaluaties, is de richtlijn uit 2016 anno 2024 weer toe aan een herziening. In **tabel 1** wordt een overzicht gegeven van de historie van de richtlijn.

Tabel 1: Historie van de richtlijn

Jaar	Titel/versie	Samenvatting van de belangrijkste wijzigingen ten opzichte van vorige versie
1999	Richtlijnen voor farmaco-economisch onderzoek	n.v.t.
2006	Herziene versie richtlijnen voor farmaco-economisch onderzoek	Actualisatie
2008	Leidraad uitkomstenonderzoek	n.v.t.
2016	Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg	<ul style="list-style-type: none">- Samenvoeging van verschillende richtlijnen: richtlijnen voor farmaco-economisch onderzoek (2006) en de leidraad uitkomstenonderzoek (2008)- Grote inhoudelijke herziening- Toevoeging verdiepingmodules:<ul style="list-style-type: none">- Kostenhandleiding*- QALY en kwaliteit-van-leven-metingen- Onzekerheid en <i>Value of information</i> analyses
2024	Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg, herziene versie 2024	Actualisatie

*Zie de **verdiepingsmodule 'kostenhandleiding'** voor de historie van de verschillende herzieningen daarvan.

Wijzigingen ten opzichte van de richtlijn van 2016

De belangrijkste wijzigingen in de huidige versie ten opzichte van de vorige versie van de richtlijn zijn als volgt:

- De disconteringsvoet voor kosten is aangepast naar 3%;
- Medische kosten in gewonnen levensjaren worden in de base-case analyse opgenomen;
- De resultaten van zowel de base-case analyse als scenarioanalyses dienen gebaseerd te worden op de probabilistische analyse;
- *Value of information (VOI) analyse*, in de vorm van de *expected value of perfect information (EVPI)* en *expected value of partial perfect information (EVPPPI)*, is een verplicht onderdeel bij modelmatige economische evaluaties;
- Kwaliteit van leven van mantelzorgers dient meegenomen te worden in een scenarioanalyse wanneer relevant;
- De EQ-5D-Y moet gebruikt worden voor het meten van kwaliteit van leven bij kinderen van 8-12 jaar oud.

Daarnaast bevat de richtlijn meer sturing omtrent:

- empirische economische evaluaties;
- extrapolatie van *time-to-event data*;
- subgroepanalyses;
- onzekerheidsanalyses;
- validatie;
- *expert opinion* en *expert elicitation*.

Tijdelijke commissie

Om Zorginstituut Nederland te ondersteunen bij het herzien van de richtlijn is een tijdelijke commissie ingesteld, bestaande uit prof. dr. G.A. de Wit (voorzitter), dr. H.M. Blommestein, prof. dr. J.E. Bosmans, prof. dr. W.B.F. Brouwer, prof. dr. J.J. van Busschbach, prof. dr. T.L. Feenstra, prof. dr. H. Koffijberg en prof. dr. M.A. Joore. Vanuit Zorginstituut Nederland waren betrokken bij de actualisatie: dr. H.A. Geuzinge (projectleider/secretaris), dr. M. El Alili, dr. J.J. Enzing, en dr. S. Knies.

Leeswijzer

Opbouw richtlijn

De richtlijn bestaat uit verschillende delen: het hoofddocument met daarin de feitelijke richtlijn en daarnaast een aantal verdiepingsmodules. Het hoofddocument bestaat uit zes hoofdstukken. In hoofdstuk 1 wordt een overzicht gegeven van de standaardanalyse waaraan alle economische evaluaties moeten voldoen. Hoofdstuk 2 gaat in op het raamwerk van een economische evaluatie. Hoofdstuk 3 besteedt aandacht aan de verschillende gegevens die nodig zijn voor een economische evaluatie. Hoofdstuk 4 is gericht op de methode. Het rapporteren van een uitgevoerde evaluatie is uitgewerkt in hoofdstuk 5. In hoofdstuk 6 worden specifieke aandachtspunten behandeld.

Verdiepingsmodules

Er bestaan drie verschillende verdiepingsmodules, namelijk:

- De kostenhandleiding
- QALY en kwaliteit-van-leven-metingen
- Onzekerheid en Value of Information analyses

De verdiepingsmodules gaan verder in op specifieke onderdelen van de richtlijn, en maken integraal onderdeel uit van de richtlijn.

1 Standaardanalyse

Om de vergelijkbaarheid en kwaliteit van Nederlandse economische evaluaties te vergroten, start deze richtlijn met een uiteenzetting van de standaardanalyse waaraan alle economische evaluaties, die worden uitgevoerd ter ondersteuning van besluitvorming, ten minste moeten voldoen (zie **tabel 2**). In scenarioanalyses kunnen er eventueel andere keuzes worden gemaakt, indien dit van toegevoegde waarde is. Deze keuzes moeten goed gemotiveerd en duidelijk beschreven worden. Eventuele additionele analyses vervangen de standaardanalyse dus niet. In **hoofdstuk 6** wordt ingegaan op verschillende aandachtspunten die relevant kunnen zijn bij diverse toepassingsgebieden.

Tabel 2: Uiteenzetting van de standaardanalyse

Onderdeel	Standaardanalyse	Zie paragraaf
Perspectief	Maatschappelijk perspectief	2.2
PICOTS	P: Beoogde patiëntenpopulatie I: De te beoordelen interventie C: Standaard en/of meest gebruikelijke interventie in Nederland (indien dit niet de meest kosteneffectieve interventie is, dient ook met de meest kosteneffectieve interventie vergeleken te worden) O: Zie 'effecten' en 'kosten' in deze tabel T: Bij modelmatige evaluaties een levenslange tijdshorizon. Bij empirische evaluaties zodanig dat een geldige uitspraak kan worden gedaan over de effectiviteit en kosten S: Context waarin de zorg wordt geleverd in Nederland	2.3
Type economische evaluatie	Kostenutiliteitsanalyse	2.4
Effectiviteitsdata	Systematische review/meta-analyse, klinische studie en ander passend bewijs	3.1
Kosten	Alle kosten binnen de gezondheidszorg (ook in gewonnen levensjaren), patiënt- en familiekosten, en kosten in andere sectoren. Productiviteitsverliezen: frictiekostenmethode. Gebruik referentieprijzen uit de verdiepingsmodule 'kostenhandleiding'	3.2
Effecten	Uitgedrukt in QALYs, bepaald met de EQ-5D-5L (of de EQ-5D-Y bij kinderen van 8-12 jaar oud) met Nederlandse waarderingen	3.3
Disconteren	Kosten 3% en effecten 1,5%	4.2
Subgroepanalyses	Subgroepanalyses bij vooraf bekende heterogeniteit in behandel-effect of in het baseline risico op basis van patiënt- of ziektekaracteristieken	4.6
Onzekerheidsanalyses	Bij modelmatige evaluaties: deterministische onzekerheidsanalyses, probabilistische analyse, scenarioanalyses en <i>value of information</i> analyse Bij empirische evaluaties: deterministische onzekerheidsanalyse, <i>bootstrapping</i> , scenarioanalyses	4.7

Validatie	Bij modelmatige evaluaties: het conceptuele model, de inputgegevens, de softwarematige implementatie en modeluitkomsten. Bij empirische evaluaties: voorbereidingen van de data, aannames rondom zorggebruik, de statistische analyse en de uitkomsten.	4.8
Rapporteren resultaten	Totale kosten en effecten voor de interventie en vergelijkende interventie(s), incrementele kosten en effecten, ICER, NMB	5.3.1
	Deterministische onzekerheidsanalyse: tornado diagram en tabel Scenarioanalyse: tabel Probabilistische of <i>bootstrap</i> analyse: <i>CE-plane</i> en CEAC <i>Value of information</i> (bij modelmatige economische evaluaties): patiënt- en populatie-EVP(P)I en EVP(P)I curve	5.3.3 t/m 5.3.5

ICER: incrementele kosteneffectiviteitsratio, NMB: net monetary benefit, CE-plane: cost-effectiveness plane; CEAC: cost-effectiveness acceptability curve; EVPI: expected value of perfect information; EVPPI: expected value of partial perfect information

2 Raamwerk van de economische evaluatie

Voordat een economische evaluatie van een zorginterventie wordt uitgevoerd, moet deze worden afgebakend en vormgegeven. Voor alle onderstaande aspecten geldt dat ze duidelijk beschreven moeten worden in de rapportage en dat alle keuzes moeten worden gemotiveerd.

2.1 Doelstelling en gebruiker

De eerste stap van een economische evaluatie is het definiëren van de doelstelling. De doelstelling moet aansluiten bij het beslisprobleem van de toekomstige gebruiker(s) van de evaluatie. Daarom is het belangrijk om de gebruiker(s) van de evaluatie te definiëren. Dat kan Zorginstituut Nederland zijn, maar bijvoorbeeld ook een zorgverzekeraar of zorginstelling.

2.2 Perspectief

De economische evaluatie moet uitgevoerd en gerapporteerd worden vanuit het maatschappelijk perspectief. Dit houdt in dat alle relevante maatschappelijke kosten en baten, ongeacht wie de kosten draagt of aan wie de baten toevallen, meegenomen moeten worden in de evaluatie.

Ter aanvulling op het maatschappelijke perspectief kunnen in scenarioanalyses de resultaten gepresenteerd worden vanuit andere perspectieven, zoals het gezondheidszorgperspectief (zie **paragraaf 3.2**).

2.3 PICOTS

Nadat er duidelijkheid is over de doelstelling, de gebruiker en het perspectief, volgt het formuleren van de specifieke vraagstelling voor de economische evaluatie. De evaluatie wordt verricht aan de hand van een vraagstelling die is opgesteld volgens de PICOTS-methode:^[1]

- **Patient** = de beoogde patiëntenpopulatie waarop de interventie zich richt;
- **Intervention** = de te beoordelen interventie;
- **Comparison** = de interventie(s) waarmee wordt vergeleken;
- **Outcome** = de relevante uitkomsten/uitkomstmaten;
- **Time** = de relevante tijdsperiode waarover effecten en kosten moeten worden meegenomen.
- **Setting** = de context waarin de zorg wordt geleverd, wanneer het voor het effect van de interventie uit kan maken. Bijvoorbeeld de eerste lijn of tweede lijn.

In onderstaande subparagrafen worden deze elementen nader toegelicht.

2.3.1 Patient (P)

De patiëntenpopulatie of doelpopulatie is de groep waarvoor de interventie is bedoeld in de Nederlandse praktijk. Nederlandse epidemiologische gegevens moeten worden gebruikt voor het beschrijven van de patiëntenpopulatie en de relevante subgroepen. Leeftijd, geslacht en ziektekenmerken kunnen verschillen tussen en binnen populaties en een grote invloed hebben op de resultaten. In **paragraaf 4.6** wordt toegelicht wanneer er subgroepanalyses uitgevoerd dienen te worden.

2.3.2 Intervention (I)

De te beoordelen interventie is een interventie zoals die in de Nederlandse praktijk kan worden toegepast voor de gedefinieerde patiëntenpopulatie. Er moet nauwkeurig worden beschreven hoe de patiënt (of hulpverlener of naaste) de interventie zal gebruiken. Daarnaast dient beschreven te worden of de interventie een bestaande interventie vervangt of het een toevoeging is aan een al bestaande interventie.

2.3.3 Comparison (C)

De te beoordelen interventie dient vergeleken te worden met de op dat moment geldende standaard- of gebruikelijke interventie in de Nederlandse praktijk.^[1] Indien dit niet de meest kosteneffectieve interventie is van de op dat moment in Nederland beschikbare alternatieven, dient ook de meest kosteneffectieve interventie als vergelijking toegevoegd te worden.

De standaardinterventie is de interventie die volgens klinische richtlijnen of praktijk wordt gezien als eerste keus.

Een vergelijking met de gebruikelijke interventie is aan de orde als er geen standaardinterventie is. Het gaat dan om de interventie die het vaakst wordt toegepast bij de gedefinieerde patiëntenpopulatie en waar op dat moment in de dagelijkse praktijk de meeste ervaring mee is. De gebruikelijke interventie kan ook bestaan uit geen behandeling en een advies om af te wachten al dan niet in combinatie met verlichting van klachten (ondersteunende of palliatieve zorg).

Indien er meerdere standaard- of gebruikelijke interventies zijn, dienen er afzonderlijke vergelijkingen te worden gemaakt. Dit kan door middel van een *fully incremental* analyse (zie **paragraaf 5.3.1**). Indien de standaard- of gebruikelijke interventie niet de meest kosteneffectieve interventie is, zal er dus ook met meerdere interventies vergeleken moeten worden, en dient dit ook door middel van een *fully incremental* analyse uitgevoerd te worden.

2.3.4 Outcomes (O)

De uitkomsten van een kosteneffectiviteitsanalyse zijn de incrementele kosten en incrementele effecten van een interventie ten opzichte van een of meerdere vergelijkende interventies. Kosten worden onder andere beïnvloed door het zorggebruik, de inzet van de betreffende interventies, prijzen, maar ook door bijvoorbeeld productiviteitsverliezen. Meer informatie over te includeren kosten(categorieën) en het bepalen van zorggebruik en prijzen staat beschreven in **paragraaf 3.2** en de **verdiepingsmodule 'kostenhandleiding'**.

De effecten worden (in het geval van een kostenutiliteitsanalyse, zie **paragraaf 2.4**) uitgedrukt in QALYs. QALYs worden berekend door gezondheidsgelateerde kwaliteit-van-leven-waarderingen (vaak 'utiliteiten' genoemd) te vermenigvuldigen met levensjaren. Kwaliteit van leven is een belangrijke patiëntgeoriënteerde uitkomst in economische evaluaties. Deze wordt gemeten met behulp van het generieke meetinstrument EQ-5D-5L (zie **paragraaf 3.3** en de **verdiepingsmodule 'QALY en kwaliteit-van-leven-metingen'**).

2.3.5 Time (T)

De lengte van de tijdshorizon van een economische evaluatie moet zodanig zijn dat alle relevante kosten en effecten meegenomen worden. Bij levensverlengende zorginterventies wordt vaak een modelmatige economische evaluatie uitgevoerd, waarbij de tijdshorizon levenslang is. Indien een levenslange tijdshorizon niet relevant is (bijvoorbeeld bij een kortdurende ziekte), kan beargumenteerd een kortere tijdshorizon worden gehanteerd. Bij empirische economische evaluaties zal de tijdshorizon vaker korter dan levenslang zijn.

2.3.6 Setting (S)

Zowel bij de interventie als bij de vergelijkende interventie is het van belang rekening te houden met de implementatie ervan in de Nederlandse praktijk. Zowel de organisatorische setting, als karakteristieken van zorgverleners en patiënten zijn hierbij van belang. De uitkomsten van zorginterventies, zowel effecten als kosten, kunnen namelijk sterk contextafhankelijk zijn. Zo kan bijvoorbeeld de behandellijn een rol spelen of de plaats waar de interventie gebruikt wordt (bijvoorbeeld in de thuissituatie of in een zorginstelling). Daarnaast kunnen er verschillende gebruikelijke interventies worden ingezet in verschillende behandelcentra. Bij karakteristieken van zorgverleners kan bijvoorbeeld gedacht worden aan het specialisme en aan de hoeveelheid ervaring met het toepassen van de interventie (zie ook **paragraaf 6.3** over leereffecten).

2.4 Analysetechniek

Er bestaan verschillende typen economische evaluaties. De meest gebruikte typen zijn de kosteneffectiviteitsanalyse (KEA) en de kostenutiliteitsanalyse (KUA). In de volksmond worden deze termen echter vaak door elkaar gebruikt. Bij zowel een KEA als KUA wordt het verschil in kosten (de incrementele kosten) gedeeld door het verschil in effecten (de incrementele effecten) van de te vergelijken interventies. Bij een KEA worden (klinische) effectmaten gebruikt (bijvoorbeeld bloeddruk, *event free survival*, of gewonnen levensjaren). Bij een KUA worden de effecten uitgedrukt in QALYs, door waarden van kwaliteit van leven te vermenigvuldigen met de

levensjaren. Het voordeel van een KUA is dat het de vergelijkbaarheid van de uitkomsten van analyses tussen verschillende patiëntenpopulaties en zorginterventies vergroot. Bij een economische evaluatie ten behoeve van een vergoedingsbeslissing dient standaard een KUA te worden uitgevoerd (zie **hoofdstuk 1**).

De economische evaluatie kan worden uitgevoerd als onderdeel van een empirische studie, of door gebruik te maken van een simulatiemodel. Meer informatie over empirische economische evaluaties en modelmatige economische evaluaties is terug te vinden in **paragraaf 4.1**.

3 Gegevens

3.1 Effectiviteit

Indien er data uit de literatuur nodig zijn met betrekking tot de relatieve effectiviteit van een interventie, is een systematisch literatuuronderzoek verplicht. Indien nodig, kan op basis van de gevonden studies een indirecte vergelijking uitgevoerd worden.

3.1.1 Systematisch literatuuronderzoek

In het literatuuronderzoek wordt op systematische wijze gezocht naar beschikbare relevante klinische studies die zoveel mogelijk aansluiten op de PICOTS, volgens een protocol dat vooraf is vastgesteld. De zoekstrategie met relevante databases en tijdsrestrictie moet worden gerapporteerd (in een bijlage), zodat deze reproduceerbaar is. Er moet een overzicht worden gemaakt van alle geselecteerde studies en de daarvoor gebruikte criteria. Daarnaast moet gerapporteerd worden waarom studies geïncludeerd dan wel geëxcludeerd zijn voor gebruik in de economische evaluatie. Het literatuuronderzoek wordt niet beperkt tot een bepaald studiedesign. De effectiviteitsgegevens moeten gebaseerd worden op studies die het grootste vertrouwen geven over de effecten in de beschreven patiëntenpopulatie.^[1]

3.1.2 Studiedesigns

Als er RCTs beschikbaar zijn, worden deze bij voorkeur als uitgangspunt genomen. Het is dan van belang dat de RCTs van goede kwaliteit zijn en dat deze aansluiten op de PICOTS. Een goed opgezette en uitgevoerde RCT geeft in principe de minste kans op vertekening van het effect en kan derhalve in principe de hoogste mate van zekerheid geven over de causale relatie tussen de interventie en het waargenomen effect.^[1] Indien het mogelijk en wenselijk is, zoals vaak bij geneesmiddelen, is een RCT bij voorkeur (drie)dubbel blind uitgevoerd.

Het is ook mogelijk om de *evidence* voor de economische evaluatie te baseren op observationeel onderzoek. Hierbij moet echter rekening worden gehouden met een grotere kans op bias (bijvoorbeeld *confounding*, informatie- en selectiebias) dan bij studies met een experimenteel design. Daarentegen kan de externe validiteit van een observationele studie wel groter zijn dan die van een RCT. Zie **paragraaf 4.4** voor methoden om te corrigeren voor bias.

Het komt voor dat een nieuwe interventie enkel onderzocht is in een *single-arm* studie. In een *single-arm* studie krijgen alle studiedeelnemers de nieuwe interventie toegewezen. Het is hierdoor niet mogelijk om enkel op basis van een *single-arm* studie betrouwbaar de relatieve effectiviteit te bepalen. Deze studies kunnen daarom enkel worden gebruikt in combinatie met een controle-arm uit een extern cohort in een indirecte vergelijking. Deze vorm van indirect vergelijken brengt in de regel zeer veel onzekerheid met zich mee en heeft dus nooit de voorkeur.

3.1.3 Indirecte vergelijkingen

Bij een indirecte vergelijking worden resultaten van studiearmen uit verschillende onderzoeken met elkaar vergeleken. Dit zijn bij voorkeur studiearmen uit gerandomiseerde studies voor een *anchored comparison*. Ook kunnen er gegevens uit niet-gerandomiseerde studies, zoals observationele cohorten en *single-arm* studies, gebruikt worden. Bij het uitvoeren van indirecte vergelijkingen is een zorgvuldige selectie van studies van belang om bias zoveel mogelijk te voorkomen.

Een indirecte vergelijking kan Bayesiaans^[2-7] dan wel frequentistisch worden uitgevoerd.^[8] Bij voorkeur wordt er gebruik gemaakt van een netwerk meta-analyse (NMA) (een *indirect treatment comparison* (ITC) of een *mixed treatment comparison* (MTC)).^[9-12] Het is echter niet altijd mogelijk om een NMA uit te voeren, bijvoorbeeld wanneer de benodigde vergelijkingen niet beschikbaar zijn in de literatuur (*incomplete evidence network*) of wanneer er sprake is van te veel heterogeniteit tussen studies. Daarnaast kan het problemen opleveren wanneer er veel tussenvergelijkingen nodig zijn in het netwerk van een MTC. Een indirecte vergelijking kan ook worden uitgevoerd door middel van een *simulated treatment comparison* (STC) of een *matching adjusted indirect comparison* (MAIC).^[12] Voor iedere methode geldt dat de voorwaarden en

aannames met betrekking tot die methode duidelijk moeten worden beschreven en gerechtvaardigd.

3.2 Kosten

Economische evaluaties moeten worden uitgevoerd vanuit het maatschappelijk perspectief. Dit betekent dat alle relevante kosten en effecten, ongeacht wie ze draagt, onderdeel zijn van de analyse. Het gaat hierbij om kosten binnen de gezondheidszorg (bijvoorbeeld ziekenhuisopnames en zorgkosten in gewonnen levensjaren), kosten voor de patiënt en de familie (bijvoorbeeld mantelzorg) en kosten die optreden in andere maatschappelijke sectoren (zoals productiviteitsverliezen, onderwijs of justitie).

Kosten kunnen berekend worden met behulp van de volgende formule: $\text{kosten} = \text{volume} \times \text{prijs}$. Om tot de volumecomponent te komen moeten eenheden gemeten worden, en met behulp van de prijs per eenheid kunnen de eenheden gewaardeerd worden om de totale kosten te berekenen. Voor identificatie, meting en waardering van de kosten, moet gebruik gemaakt worden van methoden zoals beschreven in de **verdiepingsmodule 'kostenhandleiding'**. Dit borgt de uniformiteit en standaardisatie van kostenidentificatie, -meting en -waardering in economische evaluaties van zorginterventies.

3.2.1 Soorten kosten en volumemeting

In **tabel 3** staat een overzicht van de relevante kostencategorieën behorend bij het maatschappelijk perspectief. Alle kostencategorieën moeten afzonderlijk worden beschreven en toegelicht.

Voor elke kostencategorie gelden verschillende methoden om de volumina (de hoeveelheden) te meten. Het meten hiervan bestaat uit het inventariseren van de inzet van mensen en middelen tijdens, na en soms ook voor de interventie. Idealiter wordt gebruik gemaakt van Nederlandse bronnen aangezien zorggebruik erg kan verschillen tussen landen. Indien gebruik wordt gemaakt van volumegegevens uit internationale onderzoeken, moeten deze worden gevalideerd voor gebruik in de Nederlandse situatie. Daarnaast moet er in de gaten worden gehouden dat er geen dubbeltellingen optreden tussen verschillende kostencategorieën.

Tabel 3: Perspectieven en bijbehorende kostencategorieën

Perspectief	Kostencategorieën	Specificatie kosten
Maatschappelijk perspectief	Kosten binnen de gezondheidszorg	<ul style="list-style-type: none"> - Alle zorgkosten die rechtstreeks verband houden met de aandoening - Alle zorgkosten die zich voordoen in gewonnen levensjaren
	Kosten van patiënten en familie	Bijvoorbeeld: <ul style="list-style-type: none"> - Reiskosten - Eigen betalingen - Tijkosten - Kosten van mantelzorg
	Kosten in andere sectoren	Kosten gemaakt in sectoren buiten de zorg, bijvoorbeeld door gemeenten, in het onderwijs of door vrijwilligers Productiviteitsverliezen: kosten van verzuim (absenteïsme) of improductiviteit tijdens werk (presenteïsme) en onbetaald werk
Gezondheidszorg-perspectief	Kosten binnen de gezondheidszorg	<ul style="list-style-type: none"> - Alle zorgkosten die rechtstreeks verband houden met de interventie - Alle zorgkosten die zich voordoen in gewonnen levensjaren

Bron: Drummond et al. (2015)^[13]

3.2.1.1 *Kosten binnen de gezondheidszorg*

In de categorie 'kosten binnen de gezondheidszorg' vallen alle zorgkosten die verband houden met of beïnvloed worden door de betreffende aandoening en/of interventie. Het gaat hierbij om alle kosten gedurende de gehele tijdshorizon van de economische evaluatie. De kosten van het behandelen van eventuele bijwerkingen of complicaties behoren ook tot deze kostencategorie.

Zorgkosten die zich voordoen in gewonnen levensjaren dienen ook meegenomen te worden in de *base-case* analyse. Het betreft hier alle zorgkosten als gevolg van een levensverlengend effect van een interventie. Dit kunnen zowel zogenaamde 'gerelateerde kosten' zijn (bijvoorbeeld bloedverduunners na een hartoperatie) of 'niet gerelateerde kosten' (bijvoorbeeld de kosten van een gebroken heup in de gewonnen levensjaren na een harttransplantatie). Met behulp van de *Practical Application to Include future Disease costs* (PAID) kunnen deze kosten bepaald worden.^[14]

3.2.1.2 *Kosten van patiënt en familie*

Kosten van patiënten en familie zijn kosten die optreden buiten de formele gezondheidszorg maar die wel een directe relatie hebben met de aandoening en/of interventie. Deze kosten kunnen onder andere bestaan uit reiskosten en eigen betalingen maar ook uit tijdskosten.

Indien het aantal bezoeken aan een zorginstelling of zorgverlener wordt beïnvloed door een interventie, moeten reiskosten worden meegenomen in de economische evaluatie. De frequentie van het aantal bezoeken aan een zorginstelling of zorgverlener is vaak goed te bepalen doordat dit doorgaans geregistreerd wordt. Voor de gemiddelde afstand die patiënten afleggen naar een zorginstelling en voor de monetaire waardering van reiskosten, wordt verwezen naar de **verdiepingsmodule 'kostenhandleiding'**.

Een van de belangrijkste voorbeelden van tijdskosten is het bieden van informele zorg (mantelzorg) door een naaste. Onder mantelzorg wordt verstaan: veelal onbetaalde en vaak langdurige zorg voor zieke familieleden of vrienden. Dit kan verzorging zijn of hulp bij dagelijkse activiteiten. De tijd die aan mantelzorg wordt besteed kan ten koste gaan van betaald werk, onbetaald werk en vrije tijd. Door middel van de *iMTA Valuation of Informal Care Questionnaire* (iVICQ) kan worden bepaald hoeveel uur iemand aan mantelzorg besteedt.^[15] De monetaire waardering van mantelzorguren dient gebaseerd te worden op de vervangingskosten van huishoudelijke zorg. Meer informatie over het waarderen van mantelzorg te vinden in de **verdiepingsmodule 'kostenhandleiding'**.

3.2.1.3 *Kosten binnen andere sectoren*

Welke kosten gemaakt worden in andere sectoren is sterk afhankelijk van de te evalueren interventie en de betreffende patiëntenpopulatie. Er kunnen bijvoorbeeld ook kosten worden gemaakt voor speciaal onderwijs en justitie. Wanneer op deze gebieden relevante kostenverschillen tussen de te vergelijken interventies verwacht worden, dienen deze onderdeel te zijn van de analyse. Meer informatie over het waarderen van kosten binnen andere sectoren is beschikbaar in de **verdiepingsmodule 'kostenhandleiding'**. Meer informatie is ook te vinden in de 'handleiding intersectorale kosten en baten'.^[16] De resultaten van het PECUNIA project bieden voorbeelden van te overwegen methoden voor het waarderen van kosten binnen andere sectoren.^[17] Daarnaast biedt ook de 'werkwijzer voor kosten-batenanalyse in het sociale domein' een analysebenadering.^[18]

Interventies die (ook) gericht zijn op mensen die tot de beroepsbevolking behoren, kunnen de werkgerelateerde productiviteit beïnvloeden. Indien de ziekte of de interventie mensen belemmert om productief te zijn, leidt dit tot kosten. Dergelijke kosten zijn onderdeel van de standaardanalyse. Productiviteit gerelateerd aan zowel betaald werk als onbetaald werk is daarbij van belang. In beide gevallen moet worden bepaald welke productiviteitsverliezen (of –winsten) er optreden en moeten deze worden gewaardeerd. Bij betaald werk is daarbij een

onderscheid te maken tussen verminderde productiviteit op het werk ('presenteïsme') en verzuim van het werk ('absenteïsme'). Beide kunnen van belang zijn. Een veelgebruikt instrument om productiviteitsverliezen te meten is de iMTA Productivity Cost Questionnaire (iPCQ).^[19]

Voor het waarderen van productiviteitsverliezen moet gebruik worden gemaakt van de frictiekostenmethode. Hierbij is het uitgangspunt dat zieke werknemers bij langdurige afwezigheid vervangen worden. Productiviteitsverliezen treden daardoor voornamelijk op gedurende de periode die een werkgever nodig heeft om een zieke werknemer te vervangen: de frictieperiode. Meer informatie over en toelichting op de wijze van berekenen van productiviteitsverliezen met de frictiekostenmethode is beschikbaar in de **verdiepingsmodule 'kostenhandleiding'**.

3.2.2 Waardering

De geïdentificeerde en gemeten volume-eenheden moeten vervolgens gewaardeerd worden in monetaire eenheden. De **verdiepingsmodule 'kostenhandleiding'** bevat, naast richtlijnen voor het zelf uitvoeren van kostprijsonderzoek, referentieprijzen voor een aantal veelvoorkomende eenheden. Indien specifieke benodigde kosten niet in de kostenhandleiding zijn gerapporteerd, kunnen alternatieve bronnen worden gebruikt (zie **verdiepingsmodule 'kostenhandleiding'**).

3.2.3 Inflatiecorrectie/prijsindexering

Bij het bepalen van kosten in economische evaluaties moet altijd vermeld worden op welk jaar de vastgestelde prijzen betrekking hebben. Wanneer kostprijzen van verschillende eenheden of volumes niet uit hetzelfde kalenderjaar komen, dan moeten de prijzen gecorrigeerd worden voor inflatie naar eenzelfde basisjaar. Dit wordt gedaan met de Nederlandse consumentenprijsindex (CPI), te vinden binnen Statline van het CBS. Meer informatie en toelichting zijn beschikbaar in de **verdiepingsmodule 'kostenhandleiding'**.

3.3 Kwaliteit van leven

In een KUA worden de effecten uitgedrukt in QALYs door gezondheidsgelateerde kwaliteit van leven (door sommigen ook wel 'utiliteiten' genoemd) te vermenigvuldigen met de levensjaren. In deze paragraaf wordt beschreven hoe kwaliteit van leven gemeten moet worden. In de **verdiepingsmodule 'QALY en kwaliteit-van-leven-metingen'** is daarover meer informatie te vinden.

3.3.1 Meten van kwaliteit van leven

De kwaliteit van leven moet worden vastgesteld middels gevalideerde generieke kwaliteit-van-leven-vragenlijsten, die door patiënten worden ingevuld. De gevalideerde vragenlijsten, in combinatie met een zogeheten waarderingssset of tarief, maken het mogelijk de gezondheidstoestand van de patiënt te waarderen. Zo'n waardering voor een gezondheidstoestand wordt ook wel een 'utiliteit' genoemd. De tarieven voor de waardering moeten worden bepaald bij het algemeen publiek (middels steekproeven). De patiënt beschrijft dus zijn of haar gezondheidstoestand middels een gevalideerde vragenlijst, en de resulterende gezondheidstoestand krijgt een waardering met behulp van een vastgesteld tarief op basis van voorkeuren in de algemene bevolking. Deze stappen worden verder uitgelegd in de **verdiepingsmodule 'QALY en kwaliteit-van-leven-metingen'**.

Er bestaan veel verschillende vragenlijsten voor het meten van de kwaliteit van leven, maar er zijn er maar een paar waarvoor tarieven op basis van voorkeuren in de algemene bevolking beschikbaar zijn. Vanwege de mogelijke verschillen in de waardering van kwaliteit van leven, die ontstaan bij het gebruik van verschillende vragenlijsten, dient de kwaliteit van leven in elk geval gemeten te worden met de EQ-5D-5L, met waardering op basis van Nederlandse tarieven.^[20] Voor kinderen van 8 tot 12 jaar oud is de EQ-5D-Y-3L vragenlijst beschikbaar.^[21] Voor kinderen onder de 8 jaar en voor mensen die geestelijk of lichamelijk niet in staat zijn om hun kwaliteit van leven aan te geven, kan een verzorger een *proxy* versie van de EQ-5D vragenlijst invullen.^[22]

Gebruik van de EQ-5D-5L vragenlijst is mogelijk niet in alle gevallen toereikend. Alternatieve vragenlijsten en andere methoden om de kwaliteit van leven in kaart te brengen mogen náást de standaardanalyse worden aangeboden als aannemelijk is gemaakt dat de EQ-5D-5L in de betreffende situatie niet toereikend is. Generieke uitkomstmaten hebben hierbij de voorkeur boven ziekte-specifieke uitkomstmaten. De alternatieven worden in de **verdiepingsmodule 'QALY en kwaliteit-van-leven-metingen'** besproken.

3.3.2 Disutiliteiten

Bij bepaalde interventies, zoals geneesmiddelen of operaties, kan er sprake zijn van bijwerkingen of complicaties die mogelijk ook een effect hebben op de kwaliteit van leven. Het is belangrijk dat hierbij onderscheid wordt gemaakt tussen bijwerkingen of complicaties met langdurige gevolgen en bijwerking of complicaties die kortdurend van aard zijn.

Voor het meten van effecten van complicaties en bijwerkingen op de kwaliteit van leven - meestal disutiliteiten genoemd - moet de EQ-5D-5L (of EQ-5D-Y) vragenlijst gebruikt worden. Dit betekent dat de vragenlijst moet worden afgenomen op het moment dat de bijwerkingen of complicaties voorkomen. Indien dit niet mogelijk is, dient de hoogte van de waarden (disutiliteiten), alsmede de bijbehorende tijdsduren, op wetenschappelijke literatuur gebaseerd te worden.

3.3.3 Kwaliteit van leven van mantelzorgers

Interventies hebben in sommige gevallen niet alleen effect op patiënten, maar ook op hun mantelzorgers. Mantelzorgers kunnen een essentiële rol vervullen in de zorg voor een patiënt. Onder mantelzorg wordt verstaan: veelal onbetaalde en vaak langdurige zorg voor zieke familieleden of vrienden. Dit kan verzorging zijn of hulp bij dagelijkse activiteiten. De kwaliteit van leven van mantelzorgers dient meegenomen te worden in een scenarioanalyse indien de tijdsduur van de mantelzorg en/of de kwaliteit van leven van de mantelzorgers beïnvloed wordt door de interventie. De maximale duur die hierbij aangehouden wordt is de levensduur van de patiënt.

Er zijn meerdere instrumenten en methoden beschikbaar om de kwaliteit van leven van mantelzorgers te meten en te waarderen. Er bestaat echter niet altijd afdoende evidentie over de validiteit en sensitiviteit van de beschikbare instrumenten bij gebruik in deze specifieke context. Het Zorginstituut raadt aan om effecten op gezondheidsgerelateerde kwaliteit van leven van mantelzorgers primair middels de EQ-5D-5L vragenlijst tijdens een klinische studie in kaart te brengen. Indien dit niet mogelijk is, dient gebruik gemaakt te worden van waarderingen uit de literatuur (gemeten met de EQ-5D vragenlijst). Door gebruik van de EQ-5D vragenlijst voor het meten van de kwaliteit van leven van zowel de patiënt als de mantelzorgers, kunnen de uitkomsten gecombineerd worden. Indien de kwaliteit van leven van mantelzorgers gebaseerd wordt op waarderingen uit de literatuur, kan dit als functie van de gezondheidstoestand van de patiënt worden meegenomen in de economische evaluatie.^[23]

3.4 Raadplegen van experts

Indien relevante data ontbreken, kan gebruik gemaakt worden van expertraadpleging om gegevens te verkrijgen voor een economische evaluatie. Het Zorginstituut benadrukt dat zo veel mogelijk gepubliceerde gegevens gebruikt moeten worden en dat er alleen gebruik gemaakt mag worden van expertraadpleging wanneer er geen ander bewijs beschikbaar is. Expertraadpleging wordt dan ook gezien als de laagste vorm van bewijs. Onderdelen van de evaluatie die gebaseerd zijn op expertraadpleging dienen daarom in gevoeligheidsanalyses onderzocht te worden.

Een expertpanel bestaat uit ten minste 5 onafhankelijke personen die aan diverse centra verbonden zijn, omdat opinies vaak variëren onder experts. Bij zeer zeldzame ziektes kan beargumenteerd van dit aantal worden afgeweken. De experts dienen toereikende kennis te hebben over het onderwerp waarover zij geraadpleegd worden. Experts kunnen om die reden klinische experts zijn, maar bijvoorbeeld ook ervaringsdeskundigen.

Het vergaren en rapporteren van informatie van experts dient op een gestructureerde manier gedaan te worden (zie ook **paragraaf 5.1.4** en **tabellen B1-B3 in bijlage I**). Er wordt onderscheid gemaakt tussen *expert opinion* en *expert elicitation*.

3.4.1 Expert opinion

Bij *expert opinion* gaat het om het verkrijgen van kwalitatieve informatie van experts.^[24] Voorbeelden hiervan zijn de plaatsbepaling van een nieuwe medische interventie, informatie over de huidige klinische praktijk en het beloop van een ziekte. Er moet begonnen worden met het vergaren van de informatie bij elke expert afzonderlijk. Dit kan gedaan worden door een persoonlijk interview of door het uitsturen van een vragenlijst. Bij ieder antwoord dienen de experts een onderbouwing te geven. In het geval dat de antwoorden van de individuele experts niet met elkaar in overeenstemming zijn, is streven naar consensus vereist. Dit kan door een bijeenkomst met de geraadpleegde experts te organiseren of door toepassing van de Delphi methode.^[25, 26] In het geval van een bijeenkomst, dienen duidelijke notulen aangeleverd te worden waarin terug te lezen is wat de meningen van de individuele experts zijn.

3.4.2 Expert elicitation

Bij *expert elicitation* worden kwantitatieve waarden van experts verkregen.^[24] Zij kunnen bijvoorbeeld een schatting geven van een bepaalde inputparameter van een economische evaluatie, zoals de gebruiksduur van een interventie of een verdeling van verschillende vervolgbehandelingen. Er bestaan verschillende gestructureerde methoden voor *expert elicitation*, zoals de *Sheffield Elicitation Framework* (SHELF) methode, de Delphi methode en de Cooke methode.^[26, 27, 25, 28] Het is een vereiste om een gestructureerde methode te gebruiken. Deze methoden beginnen met individuele *elicitation*, door middel van een interview of vragenlijst. Voor het uitvragen van parameterwaarden kan een *fixed interval* methode of een *variable interval* methode gebruikt worden.^[29] Deze methoden leveren, naast een puntschatting, ook een verdeling rondom de parameterwaarden. Er zijn diverse tools beschikbaar voor het uitvragen van schattingen bij experts.^[30, 27]

Experts moeten door middel van een aantal voorbeeldvragen voorbereid worden op het toepassen van verdelingen en kansen zodat zij in staat zijn de vragen goed te beantwoorden.^[31] Bij elke uitgevraagde parameter dient geverifieerd te worden of de uitkomsten daadwerkelijk de gedachten van de experts reflecteren.^[31] In het handboek van Horscroft et al. staan verschillende manieren beschreven waarop dit gedaan kan worden.^[32]

Na het verzamelen van de schattingen van individuele experts, dienen de uitkomsten samengevoegd, oftewel geaggregeerd, te worden. Dit kan door middel van wiskundige aggregatie (*mathematical aggregation*) of gedragsaggregatie (*behavioral aggregation*).^[29] Bij wiskundige aggregatie worden de individuele schattingen samengevoegd op basis van een algoritme. Bij gedragsaggregatie vindt er interactie tussen de experts plaats om vervolgens tot een uitkomst te komen door middel van consensus. Dit kan door middel van feedback op basis van de antwoorden van de andere experts (de Delphi methode) of door het organiseren van een expertpanel (zoals bij de SHELF methode). Bij beide methoden dient niet alleen een puntschatter geaggregeerd te worden maar ook een verdeling. Deze verdelingen dienen vervolgens in de deterministische onzekerheidsanalyses en probabilistische analyse meegenomen te worden (zie **paragraaf 4.7**).

4 Methode

4.1 Mogelijke onderzoeksdesigns

De economische evaluatie kan worden uitgevoerd als onderdeel van een empirische studie, of door gebruik te maken van een simulatiemodel. Bij een empirische economische evaluatie wordt er hoofdzakelijk gebruik gemaakt van kosten en effecten op patiëntniveau uit een gecontroleerde studie (waarbij de gegevens niet geëxtrapoleerd worden over de tijd). In het geval er geen empirisch onderzoek voorhanden is dat voldoet aan de PICOTS (zie **paragraaf 2.3**), dan is men aangewezen op een model. Bij een modelmatige economische evaluatie worden de verwachte kosten en effecten van interventies geschat door middel van een simulatiemodel.

De keuze voor een empirische aanpak dan wel een modelmatige aanpak, dient goed te worden gemotiveerd.

4.1.1 Empirische economische evaluaties

Bij een empirische aanpak is het van belang dat de klinische studie pragmatisch is opgezet zodat een valide schatting van de kosten en effecten van de interventies gemaakt kan worden. Aanbevelingen voor de opzet van empirische economische evaluaties zijn te vinden in de ISPOR-richtlijnen.^[33]

Bij de statistische analyse van de data moet er rekening worden gehouden met een aantal complexiteiten, zoals scheve verdelingen van de data, gecorreleerde kosten en effecten en clustering van data.^[34, 33] Indien regressietechnieken worden gebruikt om gemiddelde kosten en effecten te bepalen, dient de keuze voor de regressietechniek en het gekozen regressiemodel duidelijk te worden beargumenteerd. Er moet rekening worden gehouden met de scheve verdeling van de data door bijvoorbeeld het gebruik van lineaire regressie in combinatie met *non-parametric bootstrapping* of *generalized linear models*. Om rekening te houden met de correlatie tussen kosten en effecten kan er gebruik worden gemaakt van *seemingly unrelated regression*. Bij geclusterde data dienen *multilevel* regressiemodellen of *generalized estimating equations* gebruikt te worden. Toelichting op methoden om rekening te houden met scheefverdeeldheid van data, gecorreleerde kosten en effecten en geclusterde data is te vinden in een *scoping review*.^[34]

Naast bovengenoemde complexiteiten kan er bijvoorbeeld ook sprake zijn van verschillen in baselinekarakteristieken (bias) tussen de gerandomiseerde groepen en van missende data. In **paragraaf 4.4** wordt verder ingegaan op hoe er gecorrigeerd kan worden voor verschillende vormen van bias. In **paragraaf 4.3** staat beschreven hoe met missende data kan worden omgegaan.

4.1.2 Modelmatige economische evaluaties

Bij een modelmatige aanpak moet er een keuze voor een type model gemaakt worden. Deze keuze is sterk afhankelijk van onder meer de onderzoeksvraag en het ziektebeeld. Veelgebruikte modellen zijn *state-transition* modellen (ook bekend als Markov modellen), *partitioned survival* modellen, *discrete event* simulaties en *dynamic transmission* modellen. Er bestaan verschillende documenten/tools die gebruikt kunnen worden bij het kiezen van het juiste type model.^[35-37] De inputparameters voor het model worden geschat op basis van empirische data (bijvoorbeeld een klinische studie). Hierbij dienen aanbevolen statistische methoden gebruikt te worden.^[38]

4.2 Disconteren

Wanneer gegevens over de effecten en kosten over een periode langer dan één jaar verzameld of gemodelleerd worden, moeten deze na het eerste jaar gedisconteerd worden. Voor de standaardanalyse moeten toekomstige kosten met een constante disconteringsvoet van 3% gedisconteerd worden en toekomstige effecten met een constante disconteringsvoet van 1,5%.

De disconteringsvoet van 3% voor toekomstige kosten berust op de voor de maatschappelijke kosten-batenanalyse (MKBA) geldende disconteringsvoet van 2,25%,^[39] en een bijstelling naar

boven rekening houdend met actuele marktontwikkelingen. Vanuit de gedachte dat de waarde van gezondheid groeit over de tijd, en dat hier niet op een andere wijze rekening mee wordt gehouden, is de disconteringsvoet voor effecten lager dan de disconteringsvoet voor kosten. De disconteringsvoet voor effecten is berekend als de disconteringsvoet voor kosten minus de groeivoet van de consumptiewaarde van gezondheid.^[40, 41] Het jaarlijkse groeipercentage van de consumptiewaarde van gezondheid is in Nederland geraamd op 0,6% tot 2,9%.^[42] Er wordt een groeipercentage van 1,5% aangehouden, wat iets onder het midden van de genoemde *range* ligt en tevens de bovengrens is van een eerdere Duitse schatting.^[43] Uitgaande van een disconteringsvoet van 3% voor kosten, komt de disconteringsvoet voor effecten daarmee uit op 1,5% (3% minus 1,5%).

4.3 Missende data

Er zijn verschillende factoren die ertoe kunnen leiden dat observaties ontbreken. Het omgaan met missende data vormt een integraal onderdeel van de statistische analyse in een empirische analyse én bij het analyseren van empirische data voor het schatten van modelparameters.

Het is belangrijk om rekening te houden met missende data bij de data-analyse omdat dit anders kan resulteren in vertekende schattingen en overmatige of verminderde precisie.^[44, 45] In de literatuur wordt onderscheid gemaakt tussen naïeve methoden (zoals *complete cases*, *mean imputation* en *last value carried forward*) en meer geavanceerde statistische methoden (zoals *multiple imputation* en *expectation-maximization algoritme*).^[44, 34]

Het is belangrijk om eerst na te gaan welke missende data mechanismes een rol spelen. Bij kleine proporties missende data (<5% missende data) of *Missing Completely At Random* (MCAR) voldoen in de meeste gevallen naïeve methoden, zoals *complete case analysis*. Vaak is de proportie missende data echter groter en/of wordt aangenomen dat missende data *Missing At Random* (MAR) zijn. In deze gevallen moet *multiple imputation* gebruikt worden om missende data te imputeren. Aangezien het niet mogelijk is om onderscheid te maken tussen MAR of *Missing Not At Random* (MNAR), wordt daarnaast aanbevolen scenarioanalyses uit te voeren om de robuustheid van de MAR-assumptie te testen. Hiervoor zijn diverse methoden beschikbaar zoals o.a. selectie-modellen, *pattern-mixture* modellen en *reference-based multiple imputation*.^[46]

Het kan ook voorkomen dat er geen data beschikbaar zijn om een parameterwaarde te schatten. In dat geval kan er gebruik worden gemaakt van *expert elicitation* (zie **paragraaf 3.4.2**).

4.4 Corrigeren voor bias

In vergelijkend onderzoek kan er sprake zijn van bias (vertekening) in het gevonden effect. De kans hierop is het kleinst wanneer er gebruik wordt gemaakt van een gerandomiseerd, vergelijkend onderzoek (een RCT). Bij een RCT zijn er echter toch vormen van bias die kunnen optreden. Het risico op bias is vooral groot bij niet-gerandomiseerd vergelijkend onderzoek. In deze paragraaf worden verschillende analysemethoden genoemd om te corrigeren voor bias in de resultaten.

4.4.1 Corrigeren voor bias in gerandomiseerde studies

In RCTs kunnen verschillende vormen van bias optreden. Belangrijke vormen van bias waarvoor gecorrigeerd moet worden in de analyse zijn selectiebias en *attrition* bias.^[47] Als gevolg van selectiebias kan er sprake zijn van verschillen in baselinekarakteristieken tussen de gerandomiseerde groepen. In dit geval moet hiervoor gecorrigeerd worden door gebruik te maken van regressietechnieken. Dit reduceert bias in de resultaten en vergroot de precisie van de schattingen.^[34, 33] *Attrition* bias ontstaat als patiënten die uitvallen tijdens een studie verschillen van patiënten die wel in de studie blijven. *Attrition* bias kan gereduceerd worden door gebruik te maken van imputatie (zie **paragraaf 4.3**). Daarnaast moet er altijd gebruik worden gemaakt van *intention-to-treat* analyse. Indien het nodig is om te corrigeren voor bias, moet een scenario worden toegevoegd dat laat zien wat de uitkomsten zijn bij van de niet-gecorrigeerde analyses.

4.4.2 Corrigeren voor bias in niet-gerandomiseerde studies

Bij niet-gerandomiseerd vergelijkend onderzoek is het risico op selectiebias groot. Om voor selectiebias te corrigeren dient er gebruik gemaakt te worden van technieken zoals *propensity score* analyse of *matching*. Verder is *confounding* een grote uitdaging bij het analyseren van de data. Door middel van adequate statistische methoden moet hier voor gecorrigeerd worden.^[48] Er kan onderscheid worden gemaakt tussen *measured* en *unmeasured confounding*.^[48]

Om te corrigeren voor *measured confounding* kan gebruik gemaakt worden van *matching* of *inverse probability weighting*, met als doel om de verschillende behandelgroepen vergelijkbaar te maken.^[49] Daarnaast kan er gecorrigeerd worden voor *confounding* door middel van bijvoorbeeld *regression adjustment*, multivariate regressie of *propensity score adjustment*.^[50]

Om voor *unmeasured confounding* te corrigeren kan gebruik gemaakt worden van methoden die vallen onder instrumentele variabele analyse. Hiervoor zijn verschillende opties mogelijk, zoals *two-stage least squares* regressie en *instrumental propensity scores*.^[50] Een alternatieve benadering is om randomisatie na te bootsen door te corrigeren voor de invloed van het effect van *confounders* door middel van *difference-in-differences* en regressie-discontinuïteit analyses. Er zijn nog meer methoden om het risico op bias te minimaliseren bij niet-gerandomiseerd vergelijkend onderzoek.^[49, 50]

Indien het nodig is om voor bias te corrigeren, dient de keuze voor de methode duidelijk te worden beargumenteerd. Het is van belang dat er rekening wordt gehouden met onderliggende aannames van de betreffende methoden; deze moeten dan ook duidelijk worden beschreven en waar mogelijk getoetst. Indien er gecorrigeerd wordt voor het optreden van bias, moet een scenario worden toegevoegd dat laat zien wat de uitkomsten zijn bij gebruik van niet-gecorrigeerde analyses.

4.5 Extrapolatietechnieken

Wanneer de tijdshorizon van een klinische studie korter is dan de relevante tijdshorizon zoals gedefinieerd in de PICOTS, moeten extrapolatietechnieken gebruikt worden (in een modelmatige economische evaluatie). Longitudinale *time-to-event* data (Kaplan-Meier curves) worden hoofdzakelijk geëxtrapolerd door een parametrische verdeling voor de data te veronderstellen. Door de parameters van deze verdeling te schatten op basis van de geobserveerde data, kan een schatting gemaakt worden van het verloop van de data na observatie. Hierbij kunnen er verschillende aannames gemaakt worden over het verloop van de *events* over tijd. Alle standaard parametrische verdelingen dienen geëvalueerd en gerapporteerd te worden: exponentieel, Weibull, Gompertz, log-logistisch en lognormaal, gamma en *generalized gamma*. Flexibelere methoden zoals *spline* en *cure* modellen zijn ook mogelijk, maar mogen alleen gebruikt worden indien de standaard parametrische verdelingen niet volstaan.^[51, 52] Dit moet dan goed onderbouwd worden.

Aan de hand van de volgende stappen wordt bepaald wat de best passende verdeling voor de extrapolatie van geobserveerde data is:

- 1) Evaluatie van visuele fit: in hoeverre volgt de geschatte curve de geobserveerde Kaplan-Meier curve. Hierbij dient een voldoende gedetailleerde schaal gehanteerd te worden voor de verticale en horizontale as.
- 2) Evaluatie van de *proportional hazards* assumptie: de *proportional hazards* assumptie dient getoetst te worden om te bepalen of er afhankelijke of onafhankelijke parametrische verdelingen gefit moeten worden op de *time-to-event* data. Hiervoor worden *log-cumulative hazards* plots gebruikt, die de keuze voor een *proportional hazards* model ondersteunen. De voorkeur gaat uit naar het fitten van afhankelijke parametrische verdelingen, tenzij aangetoond wordt dat er niet aan de *proportional hazards* assumptie wordt voldaan. Bij het fitten van afhankelijke parametrische verdelingen, wordt er één parametrische curve voor beide interventies gefit met de interventie als verklarende variabele. Bij onafhankelijke parametrische verdelingen worden de waarden voor de parametrische curves voor de interventies onafhankelijk van elkaar gefit. Hierbij is het dus mogelijk dat er verschillende parametrische verdelingen worden gekozen voor de interventies. In beide opties is het van belang dat, ook op basis

van een klinische rationale, de gekozen aanpak en impliciete aannames onderbouwd worden.

- 3) Evaluatie van de statistische fit: de statistische fit moet geëvalueerd worden aan de hand van *Akaike Information Criterion* (AIC) en het *Bayesian Information Criterion* (BIC).^[53]

Bovenstaande stappen richten zich op de interne validiteit van de gebruikte parametrische verdelingen maar geven geen indicatie over hoe passend deze verdelingen zijn na de follow-up duur van een klinische studie (klinische en externe validatie). Om die reden zijn de klinische en externe plausibiliteit van de extrapolaties belangrijker dan de interne validiteit. De volgende stap is daarom van groot belang:

- 4) Klinische en externe validatie: om de klinische plausibiliteit van de geëxtrapoleerde curves te evalueren, is het mogelijk om andere databronnen te gebruiken, zoals vergelijkbare klinische studies met langere follow-up duur en registers.^[54] Indien er grote verschillen tussen de extrapolatie en externe data zijn, is het belangrijk om te evalueren of de verschillen het gevolg zijn van een suboptimale extrapolatie of van beperkingen in de externe data. Indien er geen externe data beschikbaar zijn, volstaat validatie door klinische experts (zie **paragraaf 4.8**). Het is noodzakelijk om het type verdeling en de onderliggende verdeling van events over tijd (toenemend, afnemend of constante hazards) klinisch te valideren. Daarnaast is het belangrijk om de verdeling van het incrementele effect over de tijd klinisch te evalueren (treedt het grootste deel van het incrementele effect op tijdens de follow-up duur van een studie of na observatie), zie **paragraaf 5.2.2**. Indien de extrapolatie resulteert in een langere overleving dan die van de algemene Nederlandse bevolking, dient hiervoor gecorrigeerd te worden.

Bij immature data (wanneer de mediane overleving niet is bereikt) wordt er een substantiële extrapolatie van de data gedaan.^[51, 52] Dit brengt extra onzekerheid met zich mee, naast de al bestaande onzekerheid rondom de extrapolatie van data. In dit soort situaties dienen er conservatieve aannames gedaan te worden. Dit houdt in dat van de best passende en klinisch plausibele extrapolatiecurves de meest conservatieve curve (d.w.z. resulterend in het kleinste incrementele verschil in levensverwachting tussen interventies) gekozen dient te worden in de base-case analyse.

De onzekerheid rondom de geschatte parameters dient vervolgens in de probabilistische analyse geïncorporeerd te worden (zie **paragraaf 4.7.2.2**). Verder moet in scenarioanalyses onderzocht worden hoe de resultaten veranderen bij andere relevante verdelingen en bij kortere lengtes van de tijdshorizon (zie **paragraaf 4.7.2.3**).

4.6 Subgroepanalyses

Binnen een geïndiceerde patiëntenpopulatie kan er sprake zijn van heterogeniteit. Dit kan leiden tot verschillen in incrementele kosten en effecten tussen subgroepen. Er bestaan verschillende vormen van heterogeniteit die relevant zijn voor het uitvoeren van subgroepanalyses.^[55]

Belangrijke vormen, die op dit moment kunnen worden meegenomen in vergoedingsbeslissingen, zijn gerelateerd aan patiënt- of ziektekenmerken die een relatie hebben met de effectiviteit van een interventie of het baseline risico op het optreden van events. Indien er sprake is van deze vormen van heterogeniteit, die berust op een klinische rationale, waarbij subgroepen in de klinische praktijk zijn te onderscheiden, én waarbij het ethisch en conform het gelijkheidsbeginsel is om op basis van heterogeniteit een verschillende interventie wel of niet te vergoeden, dienen er subgroepanalyses uitgevoerd te worden. Empirische studies moeten dan voldoende statistische *power* hebben voor deze subgroepanalyses. Als uiteindelijk blijkt dat de statistische *power* toch niet voldoende is, is dit echter geen valide reden om enkel de gemiddelde effecten in de totale populatie te bepalen. De gevonden effecten dienen dan alsnog per subgroep toegepast te worden in subgroepanalyses. Alle andere parameters, zoals complicatiekansen, kwaliteit-van-leven-waarden en zorggebruik, dienen dan ook specifiek voor de subgroepen te worden bepaald.

De kosteneffectiviteitsanalyse dient per subgroep uitgevoerd te worden. De onzekerheid rondom de schattingen van de subgroepen dient daarbij mee te worden genomen in de onzekerheidsanalyses.

4.7 Onzekerheidsanalyses

De uitkomsten van een economische evaluatie zijn omringd door onzekerheid. Onzekerheid kan in de eerste plaats worden veroorzaakt door methodologische keuzes: methodologische onzekerheid. Deze onzekerheid wordt grotendeels ondervangen door het volgen van deze richtlijn en het gebruik van de standaardanalyse. Er zijn echter meer vormen van onzekerheid. In onderstaande paragrafen wordt ingegaan op de verschillende vormen van onzekerheid en hoe dit te identificeren bij empirische en modelmatige economische evaluaties.

4.7.1 Onzekerheidsanalyses in empirische economische evaluaties

In empirische economische evaluaties kan er, naast methodologische onzekerheid, ook sprake zijn van statistische onzekerheid en onzekerheid in vaste parameterwaarden (bijvoorbeeld schattingen van kostprijzen).

Om de onzekerheid rond totale en incrementele kosten, effecten en de ICER te kwantificeren (statistische onzekerheid), moeten valide statistische methoden gebruikt worden. Hierbij moet rekening worden gehouden met de scheve verdeling van kosten en de eventuele correlatie tussen kosten en effecten. *Bootstrappen* is de meest gebruikte methode om onzekerheid rondom totale en incrementele kosten en effecten en de ICER te schatten.^[56] Het aantal gebruikte trekkingen bij het bootstrappen dient aantoonbaar te leiden tot een stabiel resultaat.^[57, 58] Uit de literatuur blijkt dat dit vaak het geval is bij minimaal 5.000 replicaties.^[58]

Onzekerheid in vaste parameterwaarden moet worden getoond in deterministische onzekerheidsanalyses. Waardes van parameters dienen gevarieerd te worden binnen een *range* op basis van de standaardfout. Indien er geen informatie over de standaardfout beschikbaar is, en de parameters veel invloed heeft op de uitkomsten, moet er gebruik worden gemaakt van *expert elicitation* (zie **paragraaf 3.4.2**). Indien er geen informatie beschikbaar is en de parameter weinig invloed heeft op de uitkomsten, volstaat het aannemen van een *range* van +/- 20% (rekening houdend met de logische grenzen van de betreffende parameter).

4.7.2 Onzekerheidsanalyses in modelmatige economische evaluaties

Bij een modelmatige economische evaluatie is er vaak sprake van onzekerheid door een gebrek aan data over een belangrijke uitkomstmaat, een kleine steekproefomvang en immaturiteit van data. Een tool om de verschillende onzekerheden te identificeren is TRUST.^[59]

Bij microsimulatiemodellen is het van belang dat inzichtelijk wordt gemaakt bij welke omvang van de gesimuleerde populatie de resultaten stabiel zijn.^[60] Dit kan bijvoorbeeld visueel met behulp van een convergentieplot waarin de relevante uitkomst(en) worden afgezet tegen een toenemende omvang van de gesimuleerde populatie, eventueel met 95% betrouwbaarheidsinterval erbij, of door het schatten van de standaardfout of betrouwbaarheidsinterval van de relevante uitkomst(en) op basis van een vooraf gedefinieerde (grote) omvang van de gesimuleerde populatie. Indien parametrische verdelingen worden gebruikt om variabiliteit te reflecteren, moet ook de onzekerheid in de parameters van deze verdelingen worden meegenomen. Bij microsimulatiemodellen wordt aanbevolen deze onzekerheid door middel van *bootstrapping* mee te nemen, omdat hiermee ook de correlatie tussen verschillende parametrische verdelingen op basis van dezelfde empirische dataset kan worden behouden.^[61] Verder is het van belang dat de variabiliteit duidelijk wordt onderscheiden van de overige onzekerheid.

Om de invloed van onzekerheid in kaart te brengen wordt er voor modelmatige economische evaluaties onderscheid gemaakt tussen deterministische en *threshold* analyses, probabilistische analyses en scenarioanalyses. Om de consequenties van onzekerheid te schatten, worden VOI-analyses gebruikt.

4.7.2.1 *Deterministische onzekerheidsanalyses*

Met behulp van een deterministische onzekerheidsanalyse wordt inzicht gegeven in de invloed van parameteronzekerheid op de resultaten. De simpelste vorm is de univariate analyse waarin de waarde van één enkele parameter wordt gevarieerd binnen een *range* (op basis van de standaardfout) en de rest van de parameters constant wordt gehouden. Indien er geen informatie over de standaardfout beschikbaar is, en de parameter veel invloed heeft op de uitkomsten, moet er gebruik worden gemaakt van *expert elicitation* (zie **paragraaf 3.4.2**). Indien er geen informatie beschikbaar is maar de parameter weinig invloed heeft op de uitkomsten, volstaat het aannemen van een *range* van +/-20% (rekening houdend met de logische grenzen van de betreffende parameter). In een multivariate analyse worden twee of meer parameters tegelijk gevarieerd.

Een *threshold* analyse kan worden gebruikt om te bepalen welke waarde een parameter moet aannemen om een beslissing te veranderen, of om voor een parameter de waarde te identificeren die leidt tot een vooraf gedefinieerd resultaat.^[38]

4.7.2.2 *Probabilistische analyse*

Het doel van een probabilistische analyse is het schatten van de impact van alle parameteronzekerheid tezamen op de uitkomsten van het model. Dit wordt gedaan door alle onzekere inputparameters tegelijkertijd te variëren en aan de hand van Monte Carlo simulatie een groot aantal schattingen te maken van de incrementele effecten en kosten.

De gebruikte verdelingen waarmee de onzekerheid rondom de inputparameter wordt geschat, moeten gerapporteerd en verantwoord worden. De keuze van de verdeling hangt daarbij af van de logische beperkingen van de desbetreffende parameter.^[62] De gebruikte *ranges* in de probabilistische analyse moeten hetzelfde zijn als in de deterministische analyse (zie voorgaande paragraaf).

Daarnaast moet er rekening worden gehouden met mogelijke correlaties tussen parameters. Onder meer bij parameters die onderdeel zijn van de gefitte parametrische verdelingen op de Kaplan-Meier curves (zie **paragraaf 4.5**), dient de correlatie gehandhaafd te worden. Dit kan bijvoorbeeld door middel van de Cholesky decompositie methode.^[62]

Bij microsimulatiemodellen moet ervoor worden gewaakt dat de variabiliteit/heterogeniteit en de onzekerheid niet worden vermengd.^[38, 63, 64]

Het aantal gebruikte iteraties in de probabilistische analyse dient aantoonbaar te leiden tot een stabiel resultaat. Methoden om dit te evalueren, worden beschreven in een artikel van Hatswell et al.^[65] Een goed functionerende *random* getal-generator en een vaste *random seed value* is noodzakelijk zodat de modeluitkomsten kunnen worden gereproduceerd.^[66] De ICER dient berekend te worden door de gemiddelde incrementele kosten en effecten over alle iteraties van de probabilistische analyse op elkaar te delen.

Meer informatie over de probabilistische analyse is te vinden in de **verdiepingsmodule 'Onzekerheid en Value of Information analyses'**.

4.7.2.3 *Scenarioanalyses*

Door middel van scenarioanalyses dient de invloed van methodologische keuzes en van belangrijke aannames in kaart te worden gebracht. De volgende scenario's moeten altijd uitgevoerd worden, bij zowel empirische als modelmatige economische evaluaties:

- i) geen discontering;
- ii) gezondheidszorgperspectief (zie **paragraaf 3.2**), zonder de indirecte kosten in gewonnen levensjaren.

Bij modelmatige economische analyses moeten ook de volgende scenario's uitgevoerd worden:

- i) alternatieve lengtes van de tijdshorizon (ten minste: de duur van de klinische studie en halverwege de geëxtrapoleerde tijdshorizon);
- ii) andere relevante verdelingen voor extrapolatie van *time-to-event* data (indien dit

onderdeel is van een modelmatige economische evaluatie).

De volgende scenario's dienen uitgevoerd te worden in de beschreven situaties:

- iii) andere bronnen voor kosten en zorggebruik (indien er plausibele alternatieve bronnen zijn en de gekozen waarden onzeker zijn);
- iv) andere bronnen voor de kwaliteit-van-leven-gegevens (indien er plausibele alternatieve bronnen zijn en de gekozen waarden onzeker zijn);
- v) niet meenemen van niet-gerelateerde toekomstige medische kosten (in het geval van levensverlengende interventies)
- vi) meenemen van kwaliteit van leven van mantelzorgers (indien dit beïnvloed wordt door de interventie, zie **paragraaf 3.3.3**).

Andere relevante scenarioanalyses kunnen additioneel worden toegevoegd. Alle scenarioanalyses dienen probabilistisch uitgevoerd te worden, zodat de resultaten ervan vergeleken kunnen worden met die van de standaard probabilistische analyse.

4.7.2.4 Value of information analyse

Met *Value of Information* (VOI) analyse wordt inzicht verkregen in de invloed van de onzekerheid op de besluitvorming.^[67] Een VOI-analyse wordt uitgevoerd op basis van de resultaten van de probabilistische analyse (zie **paragraaf 4.7.2.2**). De analyse richt zich op het kwantificeren van de consequenties van onzekerheid die zijn opgenomen in de analyse en ondersteunt daarmee besluitvorming over al dan niet uitstel het vergoeden van een nieuwe interventie en eventueel aanvullend onderzoek.

Als onderdeel van de standaardanalyse dienen de *expected value of perfect information* (EVPI) en *expected value of partial perfect information* (EVPPPI) bepaald te worden. De EVPI is de verwachte waarde van perfecte informatie over alle gemodelleerde aspecten van het beslisprobleem, wat gelijk is aan de verwachte kosten van de onzekerheid gerelateerd aan het nemen van een beslissing met het huidige imperfecte bewijs.^[67] Daarbij is de EVPI ook een risicomaat die de consequenties van onzekerheid laat zien. De EVPPPI laat de invloed van individuele parameters of een groep parameters zien op de consequenties van de totale beslisonzekerheid. Er bestaan verschillende benaderingsmethoden voor het bepalen van de EVP(P)I (zie **verdiepingsmodule 'Onzekerheid en Value of Information analyses'**).

Alle parameters die in de probabilistische analyse zijn meegenomen dienen ook in de EVPI te worden meegenomen. In de EVPPPI kunnen parameters gecombineerd worden in groepen op basis van de onzekerheid rondom deze groep van parameters. Bij de keuze voor groepen parameters is het van belang dat die parameters een logische samenhang vertonen, zijnde dat informatie over de parameters in een soortgelijk onderzoek kan worden verzameld. Voorbeelden van groepen van parameters zijn: waarderingen voor de kwaliteit van leven, alle kostenparameters, de effectschattingen en de extrapolaties van *time-to-event* data.

De EVPI en EVPPPI worden vervolgens vermenigvuldigd met de omvang van de populatie die baat heeft bij het onderzoek, om zo de populatie-EVP(P)I te berekenen. Het is daarom nodig om de grootte van de populatie te schatten waarop de beslissing betrekking heeft. Voor de tijdshorizon dient in ieder geval een periode van 5 jaar gehanteerd te worden. Daarnaast kan, indien goed beargumenteerd, ook een andere tijdshorizon worden meegenomen. Dit kan onderbouwd worden met historisch bewijs en verwachte toekomstige veranderingen (bijvoorbeeld patentverloop, ander onderzoek en toekomstige nieuwe interventies voor dezelfde populatie).

Wanneer de populatie-EVP(P)I klein is (lager dan de kosten van aanvullend onderzoek), geeft dat aan dat het uitvoeren van aanvullend onderzoek waarschijnlijk niet optimaal is. Besluitvorming kan dan plaatsvinden op basis van de huidige informatie. Bij een hogere populatie-EVP(P)I (wat betekent dat er substantiële gevolgen zijn van beslisonzekerheid) kan er voor gekozen worden om aanvullende VOI-analyses uit te voeren. Aanvullende VOI-analyses (*expected value of sample information* (EVSPI) en *expected net benefit of sampling* (ENBS)) kunnen worden gebruikt om de vraag te beantwoorden of, en welk, aanvullend onderzoek zinvol is en of de definitieve beslissing al dan niet moet worden uitgesteld om de resultaten van verder

onderzoek af te wachten. De noodzaak om deze aanvullende analyses uit te voeren is afhankelijk van de grootte van de gevolgen van de beslisonzekerheid en de haalbaarheid van het doen van aanvullend onderzoek. Deze aanvullende analyses zijn dus geen verplicht onderdeel van de standaardanalyse. Zie voor meer informatie de **verdiepingsmodule 'Onzekerheid en Value of Information analyses'**.

4.8 Validatie

4.8.1 Validatie bij een modelmatige economische evaluatie

Bij een modelmatige evaluatie is het van belang dat het model op verschillende vlakken gevalideerd wordt. Het doel van de validatie is om de bruikbaarheid en betrouwbaarheid van de resultaten vast te stellen. De validatie dient zowel het conceptuele model, de inputgegevens, de softwarematige implementatie (codeverificatie/technische validatie) als de modeluitkomsten (operationele validatie) te beslaan. De ISPOR-SMDM *task force* noemt verschillende soorten validatietesten.^[68] In deze paragraaf wordt per onderdeel beschreven welke validatietesten uitgevoerd moeten worden. De AdViSHE-checklist moet gebruikt worden voor de rapportage van de validatie.^[69]

4.8.1.1 Validatie van het conceptuele model

Voor de validatie van het conceptuele model moet tenminste *cross* validatie uitgevoerd worden.

Bij *cross* validatie wordt het model vergeleken met andere modellen.^[68] Er moet hierbij een literatuuronderzoek gedaan worden naar relevante studies waarin vergelijkbare indicaties en/of interventies gemodelleerd zijn. Hierbij dient niet alleen naar gepubliceerde wetenschappelijke literatuur gekeken te worden, maar ook naar eerdere beoordelingen van het Zorginstituut indien beschikbaar. Wanneer er verschillen zijn, moet er in worden gegaan op mogelijke verklaringen.

Indien er sprake is van verschillen tussen het betreffende model en de modellen in de gevonden studies en eerdere beoordelingen, of wanneer er geen vergelijking met andere modellen kan worden gedaan, moet ook *face* validatie uitgevoerd worden. Bij *face* validatie worden de modelonderdelen (bijvoorbeeld de modelstructuur) beoordeeld door meerdere (klinische) experts. Het is belangrijk dat de manier waarop de te valideren onderdelen gepresenteerd worden, past bij de kennis en ervaring van de experts. Een beschrijving van de uitgevoerde *face* validatie moet aangeleverd worden waarin duidelijk terug te lezen is wat de meningen van de individuele experts zijn over de voorgelegde onderdelen. De te raadplegen experts dienen inzicht te geven in mogelijke (conflicterende) belangen, door middel van de belangenverklaring van de KNAW.

4.8.1.2 Validatie van de inputdata

Bij validatie van de inputdata worden de voorbereidingen van ruwe data voor het schatten van modelparameters grondig gecontroleerd.^[38, 69] Bij statistische analyses (bijvoorbeeld regressieanalyses) dienen de gebruikelijke testen te worden uitgevoerd (analyse van afwijkingen tussen statistisch model en observaties, check op normaliteit waar relevant etc.).

4.8.1.3 Technische validatie

Bij de technische validatie van een model (ook wel (code)verificatie genoemd) dient gecontroleerd te worden of alle berekeningen in het model juist zijn. *Techver* is een instrument dat hiervoor gebruikt kan worden.^[70]

4.8.1.4 Operationele validatie

Operationele validatie betreft de validatie van de modeluitkomsten. Welke uitkomsten gevalideerd moeten worden, wordt bepaald op grond van hun klinische relevantie en hun invloed op de uiteindelijke resultaten van de evaluatie. Hierbij zijn zowel interne als externe validatie van belang. Bij interne validatie worden de resultaten van het model vergeleken met inputdata. Bij externe validatie worden de resultaten van het model vergeleken met empirische data, die niet gebruikt zijn als inputdata. Net als bij *cross* validatie, is er bij externe validatie een literatuuronderzoek nodig naar relevante publicaties. Indien de resultaten van het model

verschillen met wat er in de literatuur naar voren is gekomen, moet uitgebreid ingegaan worden op mogelijke verklaringen voor deze verschillen.

Verder moeten de modeluitkomsten gevalideerd worden door middel van *cross validatie*. Wanneer er verschillen worden geconstateerd tussen de uitkomsten van het model en de uitkomsten van andere modellen en empirische data, moet ook *face validatie* worden uitgevoerd.

4.8.2 Validatie van een empirische economische evaluatie

De kwaliteit van empirische studies is gewaarborgd in regelgeving omtrent het uitvoeren van deze studies (*good clinical practice (GCP)*). Toch is het ook van belang dat bij empirische economische evaluaties de voorbereidingen van de data, aannames rondom zorggebruik, gebruikte parameters, de statistische analyses en de uitkomsten gevalideerd worden zoals beschreven in voorgaande paragraaf. Dit houdt in dat ook voor empirische economische evaluaties *cross validatie* en *face validatie* moet worden uitgevoerd waarbij onder andere wordt gecontroleerd of juiste parameterwaarden (bijvoorbeeld kostprijzen) zijn gebruikt. Daarnaast kan er worden geëvalueerd of de gedane aannames in de empirische economische evaluatie overeenkomen met aannames in reeds gepubliceerde vergelijkbare economische evaluaties. Verder dient de voorbereiding van de ruwe data, zoals het opschonen van de data en de bijbehorende aannames gevalideerd te worden (*validatie inputdata*). Daarnaast moeten de statistische analyses die ten grondslag liggen aan de uitkomsten gevalideerd worden (*technische validatie*). Ten slotte dienen de uitkomsten van een empirische economische evaluatie ook vergeleken te worden met andere empirische data (*operationele validatie; externe validatie*).

5 Rapportage

5.1 Rapportage van gegevens

5.1.1 Effectiviteit

De resultaten van de klinische studies, die de basis vormen voor de effectiviteitsschattingen in een modelmatige economische evaluatie, moeten gerapporteerd worden. Bij alle economische evaluaties moeten de karakteristieken van de onderzochte patiëntengroep (leeftijd, geslacht, baseline gegevens van klinisch relevante parameters, et cetera), alsook de effectiviteit van de interventie en de vergelijkende interventie gespecificeerd worden.

Indien schattingen zijn gebaseerd op systematische reviews, (netwerk) meta-analyses en indirecte vergelijkingen, dienen de gebruikte methoden helder gerapporteerd en gerechtvaardigd te worden.

5.1.2 Kosten

De prijzen en volumes van alle meegenomen kosten dienen afzonderlijk te worden gerapporteerd. Daarnaast moet vermeld worden op welk jaar de gebruikte prijzen betrekking hebben.

5.1.3 Kwaliteit van leven

In de rapportage moet per waardering van kwaliteit van leven vermeld worden welk instrument of techniek gebruikt is, wat de nationaliteit van de respondenten is, en wat het perspectief is (patiënt of maatschappelijk). Daarnaast moet worden beschreven op welke momenten de kwaliteit van leven gemeten is. In empirische economische evaluaties moet voor elk van deze momenten de kwaliteit van leven gerapporteerd worden in een tabel of grafiek inclusief de bijbehorende *ranges*.

Bij een modelmatige economische evaluatie moeten de waarderingen per gezondheidstoestand worden weergegeven. Wanneer deze waarderingen berusten op de literatuur, moet per waardering het volgende vermeld worden: de gebruikte vragenlijst of waarderingstechniek; de kenmerken van de respondenten en het perspectief: patiëntwaarderingen of maatschappelijke waarderingen.

Wanneer wordt afgeweken van de EQ-5D vragenlijst (EQ-5D-5L of EQ-5D-Y met Nederlandse waardering^[20, 21]) dient dit beargumenteerd te worden. Indien kwaliteit-van-leven-data uit de literatuur worden gebruikt, moet onderbouwd worden waarom dit te generaliseren is naar de doelpopulatie. Als daarnaast kwaliteit-van-leven-uitkomsten van additionele instrumenten gepresenteerd worden, moet de validiteit daarvan worden onderbouwd aan de hand van wetenschappelijke publicaties.

5.1.4 Expertraadpleging

Er gelden eisen aan de rapportage van *expert opinion* en *expert elicitation*. In **tabellen B1-B3 in bijlage I** is een overzicht weergegeven van de aan de rapportage gestelde eisen.^[24] Van iedere geraadpleegde expert dient de KNAW-belangenverklaring aangeleverd te worden.

5.2 Rapportage van de methode

In de rapportage moeten de elementen van de PICOTS en het onderzoeksdesign (empirische of modelmatige economische evaluatie) helder worden geformuleerd en onderbouwd.

5.2.1 Missende data

In elke economische evaluatie waarbij gebruik is gemaakt van empirische data, hoort beschreven te worden of er sprake is van missende data, en hoe hier rekening mee is gehouden in de analyses. Hierbij is het ook van belang te rapporteren over de hoeveelheid ontbrekende data en of patiënten met missende data afwijken van patiënten zonder missende data (welke missende data mechanismen er spelen; *censoring*, MCAR, MAR of MNAR, zie **paragraaf 4.3**).

Indien missende data geïmputeerd zijn, dienen de gebruikte methoden gerapporteerd te worden.

5.2.2 Extrapolaties

Indien *time-to-event* data geëxtrapoleerd zijn (zie **paragraaf 4.5**), dient het volgende ten minste gerapporteerd te worden, zodat beoordeeld kan worden of de extrapolaties valide zijn:

- visuele fit van de geëxtrapoleerde curve op de geobserveerde data;
- evaluatie van de *proportional hazards* assumptie;
- statistische fit aan de hand van de AIC en BIC voor elke verdeling die is onderzocht;
- klinische validatie van de extrapolatie;
- het percentage van de patiënten in leven op verschillende tijdstippen bij gebruik van verschillende verdelingen.

Daarnaast moet ook het percentage geschatte effect dat tijdens follow-up duur optreedt worden gerapporteerd. Dit wordt berekend door de ratio's te berekenen tussen de incrementele QALYs over de follow-up duur van de studie (de tijdsperiode waarover *time-to-event* data beschikbaar is) en de incrementele QALYs over de volledige tijdshorizon van het model:

$$Ratio = \frac{\Delta QALY \text{ over de follow-up duur}}{\Delta QALY \text{ over de volledige tijdshorizon van het model}}$$

Eenzelfde berekening moet ook gedaan worden om het percentage geschatte effect te rapporteren nadat patiënten stoppen met het gebruik van de interventie (indien hier sprake van is). Dit wordt berekend door de incrementele QALYs tijdens de behandelperiode te delen door de totale incrementele QALYs over de volledige tijdshorizon van het model.

5.2.3 Validatie

Voor de rapportage van de uitgevoerde validatie van een modelmatige economische evaluatie dient de AdVISHE checklist ingevuld en aangeleverd te worden.^[69]

Bij empirische economische evaluaties moeten ten minste de volgende elementen gerapporteerd worden:

- Validatie van input data (onderdeel B1 van AdVISHE)
- Validatie van uitkomsten (onderdeel D1, D2 en D3 van AdVISHE)

Daarnaast moeten bij de statistische validatie de resultaten van de gebruikelijke testen worden gerapporteerd (analyse van afwijkingen tussen statistisch model en observaties, check op normaliteit waar relevant etc.).

De experts die geraadpleegd zijn voor *face* validatie, dienen de KNAW-belangenverklaring ingevuld te hebben.

5.3 Rapportage van de resultaten

5.3.1 Base-case analyse

De rapportage dient de absolute effecten en kosten te bevatten, alsmede de incrementele effecten en kosten, voor zowel de interventie als de vergelijkende interventie(s). Bij deze uitkomsten dienen ook betrouwbaarheidsintervallen gerapporteerd te worden. Behalve de ICER (uitgedrukt in kosten per gewonnen QALY), dient ook de *net monetary benefit* (NMB) te worden gerapporteerd. De resultaten moeten gebaseerd zijn op de probabilistische analyse. De absolute en incrementele kosten dienen uitgesplitst te worden naar kostencategorie (zie **tabel 3** in **paragraaf 3.2.1**) en als totaal gepresenteerd te worden. De absolute en incrementele gezondheidseffecten moeten bij modelmatige economische evaluaties ook uitgesplitst worden naar gezondheidstoestand.

Voor de effecten geldt dat in ieder geval QALYs gepresenteerd moeten worden, ongeacht de lengte van de tijdshorizon. Bij een lange tijdshorizon waarin overleving een rol speelt, moeten ook de levensjaren gepresenteerd worden.

Voor de berekening van de NMB is een referentiewaarde vereist. Deze kan worden bepaald aan

de hand van de ziektelast die van toepassing is op de betreffende indicatie, zie daarvoor het rapport '**kosteneffectiviteit in de praktijk**'.^[71]

Wanneer er meer dan twee interventies met elkaar worden vergeleken, dient een *fully incremental* analyse uitgevoerd te worden voor het rapporteren van de uitkomsten. Bij een *fully incremental* analyse worden de resultaten van de verschillende interventies in een tabel gerangschikt van lage naar hoge kosten of QALYs. Vervolgens wordt aangegeven welke interventies sterk en zwak gedomineerd worden.^[72] Van de interventies die niet gedomineerd worden, kunnen vervolgens de ICERs berekend worden. De resultaten kunnen visueel gepresenteerd worden door middel van een *efficiency frontier*. Dit moet ook gedaan worden wanneer de optimale inzet van een specifieke interventie of strategie onderzocht wordt, zoals het optimale screeningsinterval of de optimale afkapwaarde van een diagnostische test. Hierbij is het belangrijk om een voldoende groot aantal waarden voor de te optimaliseren parameter(s) te evalueren.^[73]

Om te bepalen wat de meest kosteneffectieve interventie is, kan er ook gebruik worden gemaakt van rangschikking op basis van de NMB.

5.3.2 Subgroepanalyses

Alle modeluitkomsten uit **paragraaf 5.3.1** dienen separaat voor elke subgroep gepresenteerd te worden.

5.3.3 Deterministische en scenarioanalyses

Bij de deterministische onzekerheidsanalyses moeten de effecten van variatie in de inputparameters overzichtelijk worden weergegeven. De weergave van de deterministische analyses voor parameteronzekerheid hoort de minimale en maximale ICER (gedisconteerd) te bevatten bij de gekozen spreiding van de parameter, alsmede de incrementele kosten en effecten waaruit deze ICERs zijn opgebouwd. Daarbij moet inzichtelijk worden gemaakt welke *ranges* voor de spreiding zijn gekozen en waarom. Aanbevolen wordt om de resultaten in een tabel te presenteren en daarnaast de meest invloedrijke parameters grafisch weer te geven in de vorm van een tornado diagram. Resultaten worden altijd gepresenteerd en vergeleken met resultaten uit de base-case analyse.

De resultaten van scenarioanalyses worden bij voorkeur in tabelvorm gepresenteerd met daarin de incrementele kosten en effecten en de ICER.

5.3.4 Visualisatie van probabilistische analyse

De resultaten van de probabilistische analyse van een modelmatige economische evaluatie of van een (*bootstrap*)analyse van de statistische onzekerheid in een empirische economische evaluatie, moeten grafisch worden weergegeven in een *cost-effectiveness plane (CE-plane)*. Daarbij moet worden aangegeven hoeveel procent van de (*bootstrap*)iteraties in elk van de vier kwadranten valt.

De resultaten van de probabilistische analyse moeten tevens worden weergegeven in de vorm van een *cost-effectiveness acceptability curve (CEAC)*.

5.3.5 Value of information

De resultaten van de VOI-analyse dienen te worden gepresenteerd in een EVP(P)I-curve. De EVP(P)I moet in ieder geval voor de populatie worden gerapporteerd. De grootte van de populatie waarop de beslissing betrekking heeft, alsmede de relevante levensduur van de interventie moeten worden gerapporteerd en onderbouwd.

De EVP(P)I curve moet verschillende referentiewaarden bevatten. Indien de ICER hoger is dan de geldende referentiewaarde,^[71] moet de *range* van waarden op de x-as ook de waarde van de ICER bevatten.

Indien er EVSI en ENBS berekeningen zijn gedaan, moeten de resultaten van verschillende onderzoekdesigns en voor de verschillende referentiewaarden worden gepresenteerd. Zie voor

meer informatie de **verdiepingsmodule 'Onzekerheid en Value of Information analyses'**. Meer informatie over het rapporteren van een VOI-analyse is te vinden in de CHEERS-Value of Information Reporting Standards.^[74]

6 Aandachtspunten

Economische evaluaties voor besluitvorming in Nederland moeten worden uitgevoerd volgens de standaardanalyse (zie **hoofdstuk 1**). Voor een aantal situaties gelden specifieke aandachtspunten. Deze aandachtspunten worden in dit hoofdstuk besproken.

6.1 Sequentiële elementen

De inzet van een bepaalde interventie kan de keuze van een vervolgactie beïnvloeden. In het geval van diagnostiek en screening kunnen verschillende testresultaten bijvoorbeeld leiden tot andere soorten vervolgonderzoek en andere behandelstrategieën. Daarnaast kan de inzet van een bepaalde interventie de effectiviteit van een opvolgende interventie beïnvloeden. Dergelijke sequentiële gevolgen moeten dan ook onderdeel zijn van de economische evaluatie.

In het geval van een modelmatige economische evaluatie van diagnostiek waarbij er meerdere opeenvolgende testen worden ingezet, is het belangrijk om conditionele kansen op testuitslagen te gebruiken in plaats van onafhankelijke kansen.^[75] Niet alleen kan een test zorgen voor een (verdere) selectie van testen (bij een negatieve test volgen geen verdere testen; bij een positieve test wel), de kans op een uitslag van een vervolgtest is vaak afhankelijk van de uitslagen van één of meerdere eerdere testen (de kans op een positieve CT-scan bij verdenking op longkanker is bijvoorbeeld groter na een positieve röntgenfoto dan na een negatieve röntgenfoto).

6.2 Stapsgewijze innovatie

Bij bepaalde zorginterventies is de verwachting dat ze onderhevig zijn aan stapsgewijze innovatie gedurende hun levenscyclus. Dit is bijvoorbeeld bij medische hulpmiddelen eerder regel dan uitzondering, maar kan ook aan de orde zijn bij andere interventies. De interventie wordt mogelijk verder ontwikkeld tijdens of na de uitvoering van een economische evaluatie ervan. Hierdoor kan het noodzakelijk zijn om de effectiviteit en de kosten dynamisch te modelleren. Het is van belang dat de specificaties van de onderzochte zorginterventie en eventuele aannames over (stapsgewijze) innovatie nauwkeurig worden gerapporteerd. De invloed van eventuele aannames dient onderzocht te worden in scenarioanalyses.

6.3 Leercurve

Bij veel zorginterventies, zoals medische hulpmiddelen, screening en chirurgische interventies, bestaat een leercurve die invloed heeft op de uitkomsten. Dit kan de externe validiteit van de studieresultaten beïnvloeden, bijvoorbeeld wanneer de uitkomsten van een studie met ervaren gebruikers wordt gegeneraliseerd naar de algehele populatie. Of andersom, in het geval dat uitkomsten van een kortdurende studie met onervaren gebruikers worden geëxtrapoleerd over een langere tijdshorizon. Het is van belang dat de invloed van een eventueel aanwezige leercurve inzichtelijk wordt gemaakt via een scenario- en/of onzekerheidsanalyse waarbij de leerperiode wordt geëxcludeerd.

6.4 Alternatieve uitkomsten

Het primaire doel van interventies in de gezondheidszorg is niet altijd het verbeteren van de gezondheidsgerelateerde kwaliteit van leven of het verlengen van de levensduur van de patiënt. In het geval van sommige interventies, zoals diagnostische tests en medische hulpmiddelen, kunnen bijvoorbeeld effecten worden behaald op bredere waardecomponenten.^[76, 77] Het kan daarbij bijvoorbeeld gaan om gemak voor de zorgverlener, of verkorting van de proceduretijd. Bij dergelijke interventies kan het relevant zijn om die andere waardecomponenten ook mee te nemen in de economische evaluatie. Voor het in kaart brengen en kwantificeren van die waardering via het rechtstreeks raadplegen van patiënten en gebruikers zijn methoden beschikbaar, zoals een *discrete choice experiment* (DCE) en *Multi-Criteria Decision Analysis* (MCDA).^[78-82]

In de langdurige zorg is het primaire doel van de te evalueren interventie niet altijd gezondheidsgerelateerde kwaliteit van leven zoals gemeten met de EQ-5D-5L) en/of

levensverlenging, maar welzijn. Hierdoor kan het gebruik van welzijn als uitkomstmaat passend zijn. Voorbeelden van instrumenten om welzijn te meten zijn de ASCOT, ICECAP-A, WiX, WOOP of de EQ-HWB.^[83-88] Het wordt aanbevolen om, naast de EQ-5D, in een scenarioanalyse een gevalideerde welzijnsvragenlijst te gebruiken met een beschrijvingsystematiek die resulteert in een utiliteitswaarde (op basis van Nederlandse tarieven).

Voor mantelzorgers geldt ook dat de effecten van interventies breder kunnen zijn dan gezondheid alleen. Instrumenten voor het meten van bredere uitkomstmaten bij mantelzorgers zijn bijvoorbeeld de CES, ASCOT-carer of CarerQoL.^[89-91] Deze kunnen in scenarioanalyses worden gepresenteerd.

Veel interventies in de jeugdzorg of GGZ zijn gericht op het verbeteren van gedrag of het verminderen van psychische problemen. Dergelijke interventies kunnen aangrijpen op de mentale gezondheid van de 'patiënt', maar ook op het beïnvloeden van de omgeving of familieomstandigheden. Daarmee overstijgen de effecten van interventies veelal de gezondheid van de patiënt. Het verbeteren van het gezinsfunctioneren of het verminderen van maatschappelijke overlast of drugsgebruik kunnen bijvoorbeeld een belangrijk doel zijn. Dit vergt extra aandacht voor het zorgvuldig meten van de effecten. Bredere uitkomstmaten (zoals hierboven beschreven) of uitkomstmaten gericht op mentale gezondheid (bijvoorbeeld te meten met de MHQoL of ASC-TASI),^[92, 93] kunnen hierbij passend zijn. Meer informatie is te vinden in de 'handleiding intersectorale kosten en baten'^[16] en in de 'werkwijzer voor kosten-batenanalyse in het sociale domein'.^[18] Het is hierbij van belang om dubbeltellingen te vermijden.

7 Referenties

1. Zorginstituut Nederland. Beoordeling stand van de wetenschap en praktijk 2023. Toegankelijk via <https://www.zorginstituutnederland.nl/publicaties/publicatie/2023/04/11/beoordeling-swp-2023>.
2. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 1: introduction. *Med Decis Making* 2013; 33: 597-606.
3. Dias S, Sutton AJ, Ades AE, et al. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2013; 33: 607-17.
4. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 3: heterogeneity--subgroups, meta-regression, bias, and bias-adjustment. *Med Decis Making* 2013; 33: 618-40.
5. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making* 2013; 33: 641-56.
6. Dias S, Welton NJ, Sutton AJ, et al. Evidence synthesis for decision making 5: the baseline natural history model. *Med Decis Making* 2013; 33: 657-70.
7. Dias S, Sutton AJ, Welton NJ, et al. Evidence synthesis for decision making 6: embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Med Decis Making* 2013; 33: 671-8.
8. Song F, Loke YK, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; 338: b1147.
9. Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health* 2011; 14: 417-28.
10. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health* 2011; 14: 429-37.
11. Chaimani A, Caldwell D, Li T, et al. Chapter 11: Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Toegankelijk via www.training.cochrane.org/handbook.
12. EUnetHTA. Methods guideline - D4.3.2 Direct and indirect comparisons. 2022. Toegankelijk via <https://www.eunetha.eu/wp-content/uploads/2022/08/EUnetHTA-21-Deliverable-D4.3.2-Methodological-Guideline-on-Direct-and-indirect-comparisons-V1.0.pdf>.
13. Drummond M, Sculpher M, Claxton K, et al. *Methods for the economic evaluation of health care programmes*. Fourth edition. Oxford: Oxford University Press, 2015.
14. Kellerborg K, Perry-Duxbury M, de Vries L, et al. *Practical Guidance for Including Future Costs in Economic Evaluations in The Netherlands: Introducing and Applying PAID 3.0*. *Value Health* 2020; 23: 1453-61.
15. Hoefman R, Van Exel N and Brouwer W. *iMTA Valuation of Informal Care Questionnaire (iVICQ)*. Version 1.1 (December 2011). Rotterdam: iBMG / iMTA, 2011. Toegankelijk via <https://www.imta.nl/questionnaires/ivicq/documents/>
16. Drost R, Paulus A, Ruwaard D, et al. *Handleiding intersectorale kosten en baten van (preventieve) interventies: Classificatie, identificatie en kostprijzen*. Maastricht: Maastricht University, Department of Health Services Research, 2014.
17. Pokhilenko I, Janssen LMM, Paulus ATG, et al. Development of an Instrument for the Assessment of Health-Related Multi-sectoral Resource Use in Europe: The PECUNIA RUM. *Appl Health Econ Health Policy* 2023; 21: 155-66.
18. Koopmans C, Heyma A, Hof B, et al. *Werkwijzer voor kosten-batenanalyse in het sociale domein*. Amsterdam: SEO Economisch Onderzoek, 2016.
19. Bouwmans C, Krol M, Severens H, et al. The iMTA Productivity Cost Questionnaire: A Standardized Instrument for Measuring and Valuing Health-Related Productivity Losses. *Value Health* 2015; 18: 753-8.
20. Versteegh M, Vermeulen K, Evers S, et al. Dutch Tariff for the Five-Level Version of EQ-5D. *Value Health* 2016; 19: 343-52.

21. Roudijk B, Sajjad A, Essers B, et al. A Value Set for the EQ-5D-Y-3L in the Netherlands. *Pharmacoeconomics* 2022; 40: 193-203.
22. EuroQoL. EQ-5D-Y - Available modes of administration. 2023. Geraadpleegd op 24-08-2023. Toegankelijk via <https://euroqol.org/eq-5d-instruments/eq-5d-y-available-modes-of-administration/>
23. Pennington BM. Inclusion of Carer Health-Related Quality of Life in National Institute for Health and Care Excellence Appraisals. *Value Health* 2020; 23: 1349-57.
24. Iglesias CP, Thompson A, Rogowski WH, et al. Reporting Guidelines for the Use of Expert Judgement in Model-Based Economic Evaluations. *Pharmacoeconomics* 2016; 34: 1161-72.
25. Dalkey N and Helmer O. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science* 1963; 9: 458-67.
26. European Food Safety Authority (EFSA). Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. *EFSA Journal* 2014; 12: 3734.
27. Oakley JE and O'Hagan A (2019). SHELF: The Sheffield Elicitation Framework (version 4.0). Geraadpleegd op 11-09-2023. Toegankelijk via <https://shelf.sites.sheffield.ac.uk/home>.
28. Cooke R, Mendel M and Thijs W. Calibration and information in expert resolution; a classical approach. *Automatica* 1988; 24: 87-93.
29. Bojke L, Soares M, Claxton K, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. *Health Technol Assess* 2021; 25: 1-124.
30. University of York - Centre for Health Economics. Structured expert elicitation resources (STEER). Geraadpleegd op 11-09-2023. Toegankelijk via <https://www.york.ac.uk/che/research/teehta/elicitation/steer/>.
31. Bojke L, Soares MO, Claxton K, et al. Reference Case Methods for Expert Elicitation in Health Care Decision Making. *Med Decis Making* 2022; 42: 182-93.
32. Horscroft J, Lee D, Jankovic D, et al. Structured expert elicitation for healthcare decision making: A practical guide. Datum onbekend. Toegankelijk via: <https://www.york.ac.uk/che/research/teehta/elicitation/steer/>
33. Ramsey SD, Willke RJ, Glick H, et al. Cost-effectiveness analysis alongside clinical trials II- An ISPOR Good Research Practices Task Force report. *Value Health* 2015; 18: 161-72.
34. El Alili M, van Dongen JM, Esser JL, et al. A scoping review of statistical methods for trial-based economic evaluations: The current state of play. *Health Econ* 2022; 31: 2680-99.
35. Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1. *Value Health* 2012; 15: 796-803.
36. Jin H, Robinson S, Shang W, et al. Overview and Use of Tools for Selecting Modelling Techniques in Health Economic Studies. *Pharmacoeconomics* 2021; 39: 757-70.
37. Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making* 2012; 32: 678-89.
38. Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making* 2012; 32: 722-32.
39. Werkgroep discontovoet. Rapport Werkgroep discontovoet 2020. Toegankelijk via <https://open.overheid.nl/documenten/ronl-d985cdc5-5da7-4644-be09-4c22bfa885be/pdf>.
40. Gravelle H and Smith D. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Econ* 2001; 10: 587-99.
41. Brouwer WB, Niessen LW, Postma MJ, et al. Need for differential discounting of costs and health effects in cost effectiveness analyses. *BMJ* 2005; 331: 446-8.
42. de Boer P, van Exel J, Fransen M, et al. Gezondheidseconomische aspecten van de Covidpandemie. De Bilt, RIVM rapport 2023-0318, 2023. Toegankelijk via: <https://www.rivm.nl/publicaties/gezondheidseconomische-aspecten-van-covidpandemie>
43. John J, Koerber F and Schad M. Differential discounting in the economic evaluation of healthcare programs. *Cost Eff Resour Alloc* 2019; 17: 29.
44. Mutubuki EN, El Alili M, Bosmans JE, et al. The statistical approach in trial-based economic evaluations matters: get your statistics together! *BMC Health Serv Res* 2021; 21: 475.
45. Burton A, Billingham L and Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clinical Trials* 2007; 4: 154-61.

46. Leurent B, Gomes M, Faria R, et al. Sensitivity Analysis for Not-at-Random Missing Data in Trial-Based Cost-Effectiveness Analysis: A Tutorial. *Pharmacoeconomics* 2018; 36: 889-901.
47. Phillips MR, Kaiser P, Thabane L, et al. Risk of bias: why measure it, and how? *Eye (Lond)* 2022; 36: 346-8.
48. Zorginstituut Nederland. Deliverable 3 Regie op Registers voor Dure Geneesmiddelen: Kader voor analysemethoden (D3). 2020. Toegankelijk via: <https://www.zorginstituutnederland.nl/publicaties/rapport/2020/06/24/deliverable-3-regie-op-registers-voor-dure-geneesmiddelen>
49. Faria R, Hernandez Alava M, Manca A, et al. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. 2015. Toegankelijk via <http://www.nicedsu.org.uk>.
50. Varga AN, Guevara Morel AE, Lokkerbol J, et al. Dealing with confounding in observational studies: A scoping review of methods evaluated in simulation studies with single-point exposure. *Stat Med* 2023; 42: 487-516.
51. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011. Toegankelijk via <http://www.nicedsu.org.uk>.
52. Rutherford M, Lambert P, Sweeting M, et al. NICE DSU Technical Support Document 21. Flexible Methods for Survival Analysis. 2020. Toegankelijk via <http://www.nicedsu.org.uk>.
53. Collett D. Modelling survival data in medical research. New York: Chapman and Hall/CRC, 2015.
54. Royston P and Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; 13: 33.
55. Sculpher M. Subgroups and heterogeneity in cost-effectiveness analysis. *Pharmacoeconomics* 2008; 26: 799-806.
56. Glick H, Doshi J, Sonnad S, et al. Economic Evaluation in clinical trials. Second edition. Oxford University Press, 2014.
57. Efron B and Tibshirani R. An introduction to the bootstrap. New York: CRC Press, 1994. Geraadpleegd op via.
58. Andrews D and Buchinsky M. On the number of bootstrap repetitions for BCa confidence intervals. *Econometric Theory* 2002; 18: 962-84.
59. Grimm SE, Pouwels X, Ramaekers BLT, et al. Development and Validation of the TRansparent Uncertainty ASsessmentT (TRUST) Tool for Assessing Uncertainties in Health Economic Decision Models. *Pharmacoeconomics* 2020; 38: 205-16.
60. Hahn G. Sample Sizes for Monte Carlo Simulation. *IEEE Transactions on Systems, Man, and Cybernetics* 1972; 2.
61. Degeling K, MJ IJ, Koopman M, et al. Accounting for parameter uncertainty in the definition of parametric distributions used to describe individual patient variation in health economic models. *BMC Med Res Methodol* 2017; 17: 170.
62. Briggs A, Claxton K and Sculpher M. Decision modelling for health economic evaluation. Oxford University Press, 2006.
63. Vemer P, Goossens LM and Rutten-van Molken MP. Not simply more of the same: distinguishing between patient heterogeneity and parameter uncertainty. *Med Decis Making* 2014; 34: 1048-58.
64. O'Hagan A, Stevenson M and Madan J. Monte Carlo probabilistic sensitivity analysis for patient level simulation models: efficient estimation of mean and variance using ANOVA. *Health Econ* 2007; 16: 1009-23.
65. Hatswell AJ, Bullement A, Briggs A, et al. Probabilistic Sensitivity Analysis in Cost-Effectiveness Models: Determining Model Convergence in Cohort Models. *Pharmacoeconomics* 2018; 36: 1421-6.
66. Gentle J. Generation of Random Numbers. In: Härdle W, editor. *Computational Statistics – Statistics and Computing*. New York: Springer, 2009.
67. Fenwick E, Steuten L, Knies S, et al. Value of Information Analysis for Research Decisions- An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health* 2020; 23: 139-50.

68. Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making* 2012; 32: 733-43.
69. Vemer P, Corro Ramos I, van Voorn GA, et al. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. *Pharmacoeconomics* 2016; 34: 349-61.
70. Buyukkaramikli NC, Rutten-van Molken M, Severens JL, et al. TECH-VER: A Verification Checklist to Reduce Errors in Models and Improve Their Credibility. *Pharmacoeconomics* 2019; 37: 1391-408.
71. Zorginstituut Nederland. Kosteneffectiviteit in de praktijk. 2015. Toegankelijk via: <https://www.zorginstituutnederland.nl/publicaties/rapport/2015/06/26/kosteneffectiviteit-in-de-praktijk>
72. Paulden M. Calculating and Interpreting ICERs and Net Benefit. *Pharmacoeconomics* 2020; 38: 785-807.
73. O'Mahony JF, Naber SK, Normand C, et al. Beware of Kinked Frontiers: A Systematic Review of the Choice of Comparator Strategies in Cost-Effectiveness Analyses of Human Papillomavirus Testing in Cervical Screening. *Value Health* 2015; 18: 1138-51.
74. Kunst N, Siu A, Drummond M, et al. CHEERS Value of Information (CHEERS-VOI) Reporting Standards - Explanation and Elaboration. *Value Health* 2023.
75. Novielli N, Cooper NJ and Sutton AJ. Evaluating the cost-effectiveness of diagnostic tests in combination: is it important to allow for performance dependency? *Value Health* 2013; 16: 536-41.
76. Wurcel V, Cicchetti A, Garrison L, et al. The Value of Diagnostic Information in Personalised Healthcare: A Comprehensive Concept to Facilitate Bringing This Technology into Healthcare Systems. *Public Health Genomics* 2019; 22: 8-15.
77. Lee DW, Neumann PJ and Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. *Value Health* 2010; 13: 310-4.
78. Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 2016; 19: 1-13.
79. Marsh K, M IJ, Thokala P, et al. Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 2016; 19: 125-37.
80. Lancsar E, Fiebig DG and Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. *Pharmacoeconomics* 2017; 35: 697-716.
81. Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics* 2014; 32: 883-902.
82. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics* 2019; 37: 201-26.
83. van Leeuwen KM, Bosmans JE, Jansen AP, et al. Dutch translation and cross-cultural validation of the Adult Social Care Outcomes Toolkit (ASCOT). *Health Qual Life Outcomes* 2015; 13: 56.
84. Rohrbach PJ, Dingemans AE, Essers BA, et al. The ICECAP-A instrument for capabilities: assessment of construct validity and test-retest reliability in a general Dutch population. *Qual Life Res* 2022; 31: 687-96.
85. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: Overview of the Development of a Measure of Health and Wellbeing and Key Results. *Value Health* 2022; 25: 482-91.
86. Voormolen DC, Bom JAM, de Bekker-Grob EW, et al. Development and content validation of the 10-item Well-being instrument (WiX) for use in economic evaluation studies. EsCHER Working Paper Series No. 2023004, Erasmus University Rotterdam. 2023. Toegankelijk via <https://www.eur.nl/en/research/escher/research/working-papers>.
87. Bom JAM, Voormolen DC, Brouwer WBF, et al. Construct validity, reliability and responsiveness of the 10-item Well-being instrument (WiX) for use in economic evaluation studies. EsCHER Working Paper Series No. 2023005, Erasmus University Rotterdam. 2023. Toegankelijk via <https://www.eur.nl/en/research/escher/research/working-papers>.
88. Hackert MQN, van Exel J and Brouwer WBF. Well-being of Older People (WOOP): Quantitative validation of a new outcome measure for use in economic evaluations. *Soc Sci Med* 2020; 259: 113109.

89. Rand SE, Malley JN, Netten AP, et al. Factor structure and construct validity of the Adult Social Care Outcomes Toolkit for Carers (ASCOT-Carer). *Qual Life Res* 2015; 24: 2601-14.
90. Al-Janabi H, Flynn TN and Coast J. Estimation of a preference-based carer experience scale. *Med Decis Making* 2011; 31: 458-68.
91. Brouwer WB, van Exel NJ, van Gorp B, et al. The CarerQol instrument: a new instrument to measure care-related quality of life of informal caregivers for use in economic evaluations. *Qual Life Res* 2006; 15: 1005-21.
92. van Krugten FCW, Busschbach JJV, Versteegh MM, et al. The Mental Health Quality of Life Questionnaire (MHQoL): development and first psychometric evaluation of a new measure to assess quality of life in people with mental health problems. *Qual Life Res* 2022; 31: 633-43.
93. Reckers-Droog V, Goorden M, Kaminer Y, et al. Presentation and validation of the Abbreviated Self Completion Teen-Addiction Severity Index (ASC T-ASI): A preference-based measure for use in health-economic evaluations. *PLoS One* 2020; 15: e0238858.

Bijlage I

Tabel B1: Rapportagecriteria voor *expert opinion*

Criterium	Beschrijving
Onderzoeksrationaliteit	De noodzaak voor het gebruik van expertraadpleging moet worden beschreven. Uit een systematisch literatuuronderzoek moet naar voren komen dat het bewijs niet in de literatuur beschikbaar is.
Onderzoeksvraag	Er moet een duidelijke onderzoeksvraag worden gedefinieerd.
Dataverzameling	Er moet duidelijk worden beschreven hoe de vragen zijn gesteld aan de experts (bijvoorbeeld telefonisch, <i>face-to-face</i> of door middel van een vragenlijst).
Vragen	Er moet een kopie van de gestelde vragen aangeleverd worden.
Samenstelling expertpanel	De samenstelling van een expertpanel moet worden beschreven. In de beschrijving moet duidelijk aangegeven worden welk expertisegebied ieder lid vertegenwoordigt en aan welk centrum hij/zij verbonden is.
(Conflicterende) belangen	Van iedere expert dienen mogelijke conflicterende belangen beschreven te worden middels de KNAW-belangenverklaring.
Bijeenkomst	Indien er een bijeenkomst heeft plaatsgevonden moeten duidelijke notulen worden aangeleverd waarbij eventuele verschillende standpunten van de deelnemers duidelijk naar voren komen.
Resultaten	De uitkomsten moeten duidelijk worden beschreven. De antwoorden van de afzonderlijke experts moeten gerapporteerd worden (inclusief een onderbouwing per vraag), alsmede de (niet) bereikte consensus tijdens een adviesraad.
Interpretatie	De interpretatie van de (niet) bereikte consensus moet worden beschreven.

Deze tabel is deels gebaseerd op het artikel van Iglesias et al.^[24] In het geval de Delphi methode wordt toegepast voor *expert opinion*, dienen de criteria uit het artikel van Iglesias et al.^[24] toegepast te worden in plaats van de criteria in deze tabel.

Tabel B2: Additionele rapportagecriteria voor *expert opinion* in het geval van een Delphi studie

Criterium	Beschrijving
Dataverzameling	Additioneel dient beschreven te worden hoe de resultaten van eerdere rondes zijn teruggekoppeld naar de experts en of er feedback is gegeven aan de groep en/of individuen op basis van de gegeven antwoorden.
Rondes	Het aantal geplande rondes en het aantal plaatsgevonden rondes dient gerapporteerd te worden, in combinatie met een planbeschrijving voor het overgaan naar een volgende ronde.
Resultaten	Additioneel dienen de resultaten van iedere ronde te worden beschreven

Deze criteria gelden bovenop de criteria die in **tabel B1** zijn weergegeven.

Tabel B3: Rapportagecriteria voor *expert elicitation*

Criterium	Beschrijving
Onderzoeksrationaliteit	De noodzaak voor het gebruik van expertraadpleging moet worden beschreven.

Onderzoeksvraag	Alle inputparameters waarvoor experts worden geraadpleegd moeten worden beschreven.
Type parameter	Het type inputparameter (bijvoorbeeld ratio's, kansen, verdelingen etc.) moet worden beschreven.
Samenstelling expertpanel	De samenstelling van een expertpanel moet worden beschreven. In de beschrijving moet duidelijk aangegeven worden welk expertisegebied ieder lid vertegenwoordigt en aan welk centrum hij/zij verbonden is.
(Conflicterende) belangen	Van iedere expert dienen mogelijke conflicterende belangen beschreven te worden middels de KNAW-belangenverklaring.
Methode van <i>elicitation</i>	Er moet worden beschreven hoe de gegevens worden verzameld (bijvoorbeeld door een interview of een vragenlijst).
Uitleg/training	Het gebruik van uitleg- en trainingsmateriaal voor de experts moet worden gerapporteerd en beschikbaar worden gesteld.
Vragen/opdrachten	De daadwerkelijke vragen die zijn gesteld aan de experts moeten gerapporteerd worden.
Feedback	Er moet gerapporteerd worden hoe bij iedere vraag geverifieerd is of het antwoord daadwerkelijk de gedachten van de expert reflecteert en of de expert zijn/haar antwoorden heeft herzien n.a.v. deze feedback.
Aggregatiemethode	De gebruikte aggregatiemethode moet worden beschreven.
Performance van experts	De weging van de uitkomsten van de verschillende experts en hoe dit is toegepast moet worden beschreven. Er kan gebruik worden gemaakt van differentiële weging (op basis van <i>seed variables</i>) of van gelijke weging.
Resultaten	De individuele en geaggregeerde puntenschatting(en) en verdeling(en) voor elke parameter moeten worden gepresenteerd.
Interpretatie van de resultaten	De interpretatie van de uitkomsten en hoe deze in de economische evaluatie worden gebruikt, moet worden beschreven.

Deze tabel is gebaseerd op het artikel van Iglesias et al.^[24] Voor meer informatie over de verschillende criteria wordt naar het betreffende artikel verwezen.

Colofon

Contactpersoon

Dr. H.A. Geuzinge
infoOenW@zinl.nl

Adres

Zorginstituut Nederland
Postbus 320
1110 AH Diemen

