



Zorginstituut Nederland

Beoordeling stand van de wetenschap en praktijk

| Van goede zorg verzekerd |

Beoordeling stand van de wetenschap en praktijk

Definitieve geactualiseerde versie | januari 2015

Inhoud

Samenvatting	5
Inleiding	11
1 Introductie en beschouwingen vooraf	13
1.1 Zorgverzekeringswet	13
1.1.1 Wettelijk kader	13
1.1.1.1 Samenstelling van het basispakket van de Zvw	13
1.1.1.2 Te verzekeren prestaties; open en gesloten systeem	13
1.1.2 Bedoeling van de norm 'de stand van de wetenschap en praktijk'	15
1.1.3 Legitimering vaststellen beoordelingskader 'de stand van de wetenschap en praktijk'	16
1.1.3.1 Taken Zorginstituut Nederland op gebied van pakketbeheer	16
1.1.3.2 Beoordelingspraktijk: onderwerpen die wij beoordelen	18
1.1.4 Voorwaardelijke toelating van zorg tot het pakket	19
1.2 Wet langdurige zorg	20
1.2.1 Wettelijk kader	20
1.2.2 Taken Zorginstituut Nederland op gebied van pakketbeheer	20
1.2.3 Beoordelingspraktijk: onderwerpen die wij beoordelen	21
1.3 Zorginstituut Nederland: nieuwe kwaliteitstaken	21
2 Beoordelingskader: hoofdlijnen	22
2.1 Proces van beoordeling en standpuntbepaling	22
2.2 Relatieve effectiviteit	23
2.3 Toepassen principes van evidence-based medicine	24
2.4 Wetenschap en praktijk: elementen van integrale beoordeling	25
2.5 In richtlijnen beschreven interventies	27
2.5.1 Vaste werkwijze	27
2.5.2 Kwaliteitsinstituut in de zorg: nieuwe taken	27
2.6 Verzekeringsstatus buitenland	28
3 Werkwijze en uitgangspunten	29
3.1 Zoeken en selecteren van evidence en verdere informatie	30
3.1.1 Welke interventie bij welke patiëntenpopulatie (en setting) beoordelen wij?	30
3.1.2 Waarmee vergelijken wij de te beoordelen interventie?	30
3.1.2.1 Vergelijking met de standaard- of gebruikelijke behandeling	30
3.1.2.2 Vergelijking met placebo	31
3.1.2.3 Directe en indirecte vergelijkingen	33
3.1.3 Welke uitkomsten betrekken wij in de vergelijking?	34
3.1.3.1 Gunstige en ongunstige uitkomsten	34
3.1.3.2 Uitkomstmaten nader beschouwd	35
3.1.3.3 Keuze uitkomstmaten/uitkomsten beargumenteren	37
3.1.4 Welke uitkomst van de vergelijking vinden wij voldoende?	37
3.1.4.1 Meerwaarde en gelijke waarde	37
3.1.4.2 Klinisch relevant verschil	38
3.1.4.3 Bepalen klinische relevantie	39
3.1.5 Welke follow up-duur willen wij zien?	40
3.1.6 Informatiebronnen	40

3.2 Beoordelen en graderen kwaliteit van het bewijs	41
3.2.1 Aspecten die de kwaliteit van het bewijs bepalen	41
3.2.1.1 Methodologische aspecten	41
3.2.1.2 Belang van de resultaten	41
3.2.1.3 Generaliseerbaarheid	41
3.2.1.4 Onderzoeksmethodiek (EBRO-classificatie respectievelijk GRADE)	42
3.2.2 EBRO-classificatie	42
3.2.3 'Passend bewijs'	43
3.2.4 GRADE-methodiek	44
3.2.4.1 Uitkomstmaten als vertrekpunt	44
3.2.4.2 Downgraden	45
3.2.4.3 Upgraden	45
3.2.4.4 Beoordeling kwaliteit van het bewijs	45
3.3 Vaststellen eindbeoordeling/conclusie	46
3.3.1 Argumenten voortkomend uit balans tussen voor- en nadelen van de interventie	46
3.3.2 Argumenten voortkomend uit kwaliteit van het wetenschappelijke bewijs	46
3.3.3 Argumenten voortkomend uit de 'passend bewijs benadering'	46
3.3.4 Beargumenteerde opvattingen/garanties beroepsgroepen en patiënten	46
3.3.5 Onderbouwing standpunt	48
4 Langdurige zorg	49
4.1 Inleiding	49
4.2 Kenmerken van de langdurige zorg	49
4.3 Beoordeling effectiviteit langdurige zorg in kader van pakketbeheer	50
4.4 Overzicht van activiteiten om te voorzien in toegespitst beoordelingskader	50
4.4.1 Onderzoek meetinstrumenten langdurige zorg	50
4.4.2 Project passend bewijs effectiviteit langdurige zorg	50
4.4.3 Bevorderen van totstandkoming kwaliteitstandaarden langdurige zorg	51
5 Medische tests	52
5.1 Klinisch nut	52
5.2 Van EBRO-classificatie naar GRADE	52
6 Ingangsdatum en gevolgen van standpunt over 'de stand van de wetenschap en praktijk'	54
6.1 Ingangsdatum	54
6.2 Gevolgen van een standpunt voor rechten verzekeren	54
6.2.1 Standpunt is in overeenstemming met de uitvoeringspraktijk	54
6.2.2 Standpunt is niet in overeenstemming met de uitvoeringspraktijk	55
6.3 Overige gevolgen van een standpunt	55
6.3.1 Zorgverzekeringen (polissen)	55
6.3.2 Contractering van zorg	55
6.3.3 Prestatiebeschrijving/tarieven NZa	56
Vaststelling rapport	57
Literatuurlijst	59
Bijlage	63
Colofon	72

Samenvatting

Actualisering beoordelingskader

Nederland heeft via de Zorgverzekeringswet en de Wet langdurige zorg¹ voor al zijn ingezetenen een op onderlinge solidariteit steunende sociale zorgverzekering gerealiseerd met een ruim basispakket aan zorg. De overheid, die bepaalt wat er in het basispakket zit, heeft ervoor gekozen om 'de stand van de wetenschap en praktijk' voor alle verzekerde zorg als begrenzing te laten gelden. Het komt erop neer dat alleen zorg die als effectief wordt beschouwd, onderdeel uitmaakt van het pakket van de Zorgverzekeringswet en de Wet langdurige zorg.

Ons instituut vervult in het kader van de Zorgverzekeringswet en de Wet langdurige zorg de functie van pakketbeheerder. Onderdeel van deze functie is dat wij beoordelen of zorg voldoet aan 'de stand van de wetenschap en praktijk'. Om te bevorderen dat wij bij deze beoordelingen consistentie betrachten en ook om extern onze werkwijze te verantwoorden hebben wij in 2007 een beoordelingskader vastgesteld.² Onze werkwijze is in de loop van de jaren geëvolueerd en nader ingevuld. Deze ontwikkelingen hebben wij verwerkt in een geactualiseerde versie van ons beoordelingskader. In dit rapport maken wij dit kader kenbaar.

Wij beschrijven in dit rapport uitgebreid hoe wij 'anno nu' het criterium 'de stand van de wetenschap en praktijk' beoordelen. Alle aspecten die mogelijk een rol kunnen spelen in onze beoordelingen komen daarbij aan bod. Dat wil niet zeggen dat ook alle beschreven aspecten in elke beoordeling worden betrokken. Welke aspecten een relevante rol spelen, hangt af van het specifieke dossier.

De hoofdlijnen van ons beoordelingskader laten zich als volgt samenvatten.

Relatieve effectiviteit

Voor het bepalen van 'de stand van de wetenschap en praktijk' gaan wij na of het behandelbeleid (diagnostiek, behandeling), gelet op de gunstige en de ongunstige gevolgen ervan (bijwerkingen, veiligheid), leidt tot relevante (meer)waarde voor de patiënt in vergelijking met de standaardbehandeling of gebruikelijke behandeling. Anders gezegd: vinden wij de 'netto toevoeging' van de te beoordelen interventie in vergelijking met de al bestaande zorg een gewenste, relevante toevoeging en groot genoeg, en hebben wij er voldoende vertrouwen in dat deze toevoeging ook daadwerkelijk optreedt? Het gaat ons dus om de zogenoemde relatieve effectiviteit van een interventie.³

Het komt erop neer dat wij - om een standpunt te kunnen innemen over de effectiviteit van een interventie - een vergelijking maken. Om die vergelijking te kunnen maken bepalen wij per te beoordelen interventie:

- Met welke interventie vergelijken wij de te beoordelen interventie?
- Op welke doelgroep richt de interventie zich?
- Welke uitkomsten vinden wij cruciaal/belangrijk en betrekken wij in de vergelijking?
- Vinden wij de uitkomst van de vergelijking (het relatieve effect) groot genoeg (m.a.w. is de uitkomst klinisch relevant)?

Op basis hiervan - en afhankelijk van de vraag of wij voldoende vertrouwen in de uitkomst van de vergelijking kunnen hebben - trekken wij een conclusie.

Toepassen van de principes van evidence-based medicine

Vaste werkwijze is dat wij voor het beoordelen of zorg voldoet aan 'de stand van de wetenschap en praktijk' de principes van evidence-based medicine (EBM) volgen. EBM is primair ontwikkeld als leidraad voor praktiserende artsen voor het nemen van klinische beslissingen over individuele patiënten, maar

¹ De Wet langdurige zorg is per 1 januari 2015 in de plaats gekomen van de Algemene Wet Bijzondere Ziektekosten.

² CVZ. Beoordeling stand van de wetenschap en praktijk. Diemen, 2007. Rapportnr. 254. Zie ook het rapport over de beoordeling van medische tests. CVZ. Medische tests. Beoordeling stand van de wetenschap en praktijk. Diemen, 2011. Rapportnr. 293.

³ De Engelse termen zijn: Comparative Effectiveness respectievelijk Relative Effectiveness.

deze principes vinden inmiddels bredere toepassing. Zij worden ook gevolgd bij de ontwikkeling van richtlijnen door de beroepsverenigingen van zorgverleners en bij het ontwikkelen van beleid op het terrein van volksgezondheid. In dat geval gaat het niet om beslissingen over individuele patiënten, maar om aanbevelingen respectievelijk beslissingen op populatieniveau.

EBM is ook goed bruikbaar om te beoordelen of zorg voldoet aan ‘de stand van de wetenschap en praktijk’. Ook daarvoor geldt dat het niet gaat om de vraag of de zorg bij een individuele patiënt effectief is, maar of de zorg bij een bepaald indicatiegebied effectief is. Deze vraag moet, gezien de wettelijke formulering, beantwoord worden aan de hand van ‘de stand van de wetenschap en praktijk’. De formulering duidt erop dat het niet om twee losse criteria gaat (‘de stand van de wetenschap’ en ‘de stand van de praktijk’) die ieder afzonderlijk beoordeeld moeten worden en potentieel een tegenstrijdig resultaat zouden kunnen geven. De formulering geeft aan dat het één geïntegreerde wettelijke maatstaf betreft. EBM combineert en verenigt juist de beide elementen, wetenschap en praktijk.

De kern van EBM is dat de beschikbare evidence – die uiteen kan lopen van gerandomiseerd vergelijkend onderzoek tot praktijkervaring – systematisch wordt gezocht en geselecteerd en op gestructureerde wijze wordt gewogen en gebruikt. EBM heeft als uitgangspunt dat wetenschappelijk onderzoek van hoge kwaliteit in de afweging het zwaarste weegt. Het werken volgens de EBM-principes kent een aantal vaste stappen: het zoeken en selecteren van informatie, het beoordelen van de gevonden informatie en het trekken van een conclusie.

Op de volgende punten gaan wij hierna achtereenvolgend in: integrale beoordeling elementen wetenschap en praktijk, passend bewijs en vaste stappen EBM.

Integrale beoordeling elementen wetenschap en praktijk

EBM betekent dat naast wetenschappelijke inzichten ook de in de praktijk gevormde expertise en ervaringen van zorgverleners en zorggebruikers bij de beoordeling in ogenschouw worden genomen. Dit laatste komt in onze werkwijze op verschillende manieren tot uiting. Zo maken wij bijvoorbeeld gebruik van bestaande (internationale) evidence based-richtlijnen en sluiten daar, indien mogelijk, bij aan. Verder is het zo dat wij voor het bepalen van de zogenoemde PICOT-vragen te rade gaan bij de wetenschappelijke verenigingen van professionals en patiëntenorganisaties. Vanuit hun praktische kennis en ervaring kunnen zij ons relevante informatie aanreiken over bijvoorbeeld de te kiezen uitkomsten/uitkomstmaten. Ook bij het bepalen van het eindoordeel kunnen – afhankelijk van de kwaliteit van het gevonden bewijs en onder bepaalde voorwaarden – de in de praktijk opgedane inzichten en ervaringen van professionals en patiënten een doorslaggevende rol spelen.

Het begrip ‘de stand van de wetenschap en praktijk’ heeft een normatief karakter. ‘Praktijk’ is daarom niet: ‘dat wat individuele professionals (gewoonlijk) feitelijk aan zorg leveren’. Evenmin gaat het bij ‘praktijk’ om de mening van individuele behandelaars (en individuele patiënten) over de waarde van de interventie. Wat eventueel meegenomen wordt in de afweging is dat wat de beroepsgroep als geheel maatgevend vindt respectievelijk als juiste behandelwijze beschouwt.

Passend bewijs

EBM heeft als uitgangspunt dat wetenschappelijk onderzoek van hoge kwaliteit in de afweging het zwaarste weegt. Zo geeft een goed opgezet en uitgevoerd gerandomiseerd vergelijkend onderzoek (randomised controlled clinical trial – RCT) in principe de minste kans op vertekening van het effect en derhalve de hoogste mate van zekerheid over de causale relatie tussen de interventie en het waargenomen effect. Andere onderzoeksdesigns, zoals bijvoorbeeld observationeel onderzoek, hebben als nadeel dat er meer kans is op onder- of overschatting van de effectiviteit.

Onderkend moet echter worden dat een RCT ook nadelen heeft en beperkingen kent. Zo geeft een RCT niet altijd de dagelijkse praktijk weer, vanwege het soms zeer streng toepassen van in- en exclusiecriteria. Ook is een RCT niet altijd geschikt als onderzoeksdesign (denk bijvoorbeeld aan onderzoek naar bijwerkingen op de lange termijn), niet altijd nodig en ook niet altijd haalbaar. In verband hiermee

hebben wij een kader ontwikkeld aan de hand waarvan wij op systematische wijze kunnen afwegen en beargumenteren, wat - gelet op de betreffende interventie - als 'passend bewijs' kan dienen.

Vaste stappen EBM

Het werken volgens de EBM-principes kent een aantal vaste stappen: het zoeken en selecteren van informatie, het beoordelen van de gevonden informatie en het trekken van een conclusie.

Stap 1: Zoeken en selecteren van informatie

Met behulp van de zogenoemde PICOT-vragen bepalen wij beargumenteerd welke aspecten relevant zijn om de effectiviteit van de te beoordelen interventie te bepalen. PICOT staat voor:

- Patient = de relevante patiëntenpopulatie;
- Intervention = de te beoordelen interventie;
- Comparison = de interventie waarmee wordt vergeleken (controle-interventie);
- Outcome = de relevante uitkomsten/uitkomstmaten;
- Time = minimaal vereiste follow-up periode.

Verder bepalen wij beargumenteerd:

- Het minimaal vereiste klinisch relevante verschil in uitkomst;
- Het zogenoemde 'passend bewijs profiel'.

Met behulp van de PICOT-vragen zoeken wij systematisch in bibliografische databases naar relevante literatuur. Wij gaan in beginsel uitsluitend af op gepubliceerde en peer-reviewed literatuur.

Stap 2: Beoordelen en graderen van de kwaliteit van het bewijs

De kwaliteit van de met de PICOT gevonden en geselecteerde literatuur beoordelen wij aan de hand van een aantal aspecten: methodologische aspecten, belang van de resultaten, generaliseerbaarheid en de gebruikte onderzoeksmethodiek (onderzoeksdesign). Voor de beoordeling sluiten wij - waar mogelijk - aan bij recente internationale ontwikkelingen en passen wij de zogenoemde GRADE-methode toe. GRADE staat voor 'Grading of Recommendations Assessment, Development and Evaluation'. Met behulp van deze methode kan een inschatting van de kwaliteit van het in totaal verzamelde bewijs worden gegeven, d.w.z. van de mate van vertrouwen dat het geschatte effect daadwerkelijk bestaat. Een belangrijk kenmerk van GRADE is dat per uitkomstmaat de zogenaamde 'body of evidence' wordt bepaald en beoordeeld. Daarbij kan de kwaliteit van RCT's worden afgewaardeerd, respectievelijk de kwaliteit van observationele studies worden opgewaardeerd, afhankelijk van bepaalde, omschreven beoordelingsgronden.

GRADE kent een viertal niveaus van kwaliteit van het bewijs:

- hoge kwaliteit: er is veel vertrouwen in de schatting van het effect;
- middelmatige kwaliteit: er is redelijk vertrouwen in de schatting van het effect;
- lage kwaliteit: er is beperkt vertrouwen in de schatting van het effect;
- zeer lage kwaliteit: er is zeer weinig vertrouwen in de schatting van het effect.

Stap 3: Vaststellen eindbeoordeling/conclusie.

Uiteindelijk bepalen wij als laatste stap welke conclusie wij m.b.t. de effectiviteit kunnen trekken mede op basis van de beoordeelde literatuur. Het gaat ons, zoals eerder gezegd, om de relatieve effectiviteit. Het komt erop neer dat wij het volgende afwegen: vinden wij de 'netto toevoeging' van de interventie in vergelijking met de al bestaande zorg een gewenste, relevante toevoeging en groot genoeg, en hebben wij er voldoende vertrouwen in dat deze toevoeging ook daadwerkelijk optreedt? In de afweging die wij maken om m.b.t. een concrete interventie ons eindoordeel te bepalen, betrekken wij in beginsel de volgende argumenten/aspecten:

- Argumenten voortkomend uit de balans tussen de voor- en nadelen van de interventie;
- Argumenten voortkomend uit de kwaliteit van het wetenschappelijke bewijs;
- Argumenten voortkomend uit de 'passend bewijs benadering';
- Beargumenteerde opvattingen/garanties beroepsgroepen en patiënten.

Wij dragen zorg voor een transparante onderbouwing van onze afweging en van het daarop steunende standpunt.

Proces van beoordeling en standpuntbepaling

Wij hebben de beoordeling van 'de stand van de wetenschap en praktijk' ingebed in een proces dat waarborgt dat de benodigde relevante input beschikbaar komt en dat een weloverwogen standpunt kan worden ingenomen. Zo worden in beginsel professionals via hun wetenschappelijke verenigingen en – inmiddels ook – patiëntenverenigingen van meet af aan bij een beoordeling betrokken. Ditzelfde geldt voor de zorgverzekeraars. Verder laten wij ons bij de beoordelingen adviseren door een uit onafhankelijke, externe deskundigen bestaande commissie, de Wetenschappelijke Adviesraad. Dit zorgt voor de voor de beoordeling benodigde inbreng van actuele wetenschappelijke kennis en van ervaring met de behandelpraktijk.

Specifieke onderwerpen

In het rapport gaan wij ook in op een aantal specifieke onderwerpen. Deze geven wij hier kort weer.

Beoordeling van langdurige zorg

Om verschillende redenen vindt onderzoek naar de effectiviteit van langdurige zorg nog weinig plaats. Dit bemoeilijkt de toetsing van deze zorg aan het criterium 'de stand van de wetenschap en praktijk'. Ons streven is om daar verandering in te brengen en te komen tot een op de langdurige zorg toegespitste invulling van het beoordelingskader, die rekening houdt met de specifieke kenmerken van de langdurige zorg. In het rapport zetten wij uiteen welke activiteiten wij ontplooien die daarbij behulpzaam (kunnen) zijn. Dat betreft onder meer het bruikbaar maken van de vragenlijst 'passend bewijs' voor de beoordeling van langdurige zorg. De verwachting is dat van deze activiteiten een stimulerende werking zal uitgaan en het doen van adequaat effectiviteitsonderzoek in deze sector zal bevorderen, hetgeen uiteindelijk de kwaliteit van de zorgverlening ten goede zal komen.

Beoordeling van medische tests

Op 20 januari 2011 heeft ons instituut het rapport "Medische tests, beoordeling stand van de wetenschap en praktijk" uitgebracht.² De uitgangspunten voor de beoordeling van medische tests die daarin zijn neergelegd, blijven van kracht. Bij de beoordeling van medische tests gaat het ons niet enkel om het vermogen een fraaie afbeelding of correcte diagnose te verkrijgen. Natuurlijk is van groot belang dat een test betrouwbaar is, maar een uiteindelijk positief oordeel over een test moet vooral gebaseerd zijn op aantoonbaar positieve gevolgen voor aan de gezondheid gerelateerde uitkomsten bij degenen die getest worden. Bij de evaluatie van een test beoordelen wij dan ook het hele traject, de test-plus-behandeling-strategie, op effectiviteit, ook wel omschreven als klinisch nut. Op basis van vergelijkend onderzoek tussen de gebruikelijke ('oude') en de voorgestelde ('nieuwe') test-plus-behandeling-strategie kan het klinisch nut worden bepaald. Dergelijk rechtstreeks bewijs ontbreekt echter vaak. In dat geval gaan wij door middel van de constructie van een vergelijkend analyseraam na of niet-rechtstreeks bewijs behulpzaam kan zijn voor beantwoording van de vraag naar het klinisch nut. Geleidelijk gaan wij ertoe over om bij de beoordeling van tests ook de GRADE-methode toe te passen.

Verhouding tussen pakketbeheer en kwaliteitstaken

Ons instituut heeft naast taken op het gebied van pakketbeheer ook taken op het gebied van het bevorderen van kwaliteit van zorg. Gangbare benaming voor dit onderdeel van ons takenpakket is: Kwaliteitsinstituut in de zorg. In het rapport gaan wij kort in op de verhouding tussen de beoordeling van 'de stand van de wetenschap en praktijk' van zorg in het kader van pakketbeheer en de toetsing van kwaliteitsstandaarden voor opname in het register die wij als Kwaliteitsinstituut doen. Het toetsingskader voor opname in het register bevat geen materiële, inhoudelijke normen, maar is alleen procedureel van aard. Belangrijke constatering is dan ook dat opname van bijvoorbeeld een richtlijn in het register niet per definitie betekent dat voldaan is aan de maatstaven die ons instituut aanlegt voor de beoordeling van 'de stand van de wetenschap en praktijk' in het kader van pakketbeheer. Een in het register opgenomen kwaliteitsstandaard beschrijft wat goede zorg is, maar uit opname in het register volgt niet automatisch dat de in de kwaliteitsstandaard beschreven zorg ook verzekerde basiszorg is.

Ingangsdatum en gevolgen standpunt over ‘de stand van de wetenschap en praktijk’

Een standpunt over ‘de stand van de wetenschap en praktijk’ kan – afhankelijk van de situatie - gevolgen hebben voor de rechten van verzekerden, de polisvoorwaarden, contractering van zorg en prestatiebeschrijvingen/tarieven van de NZa. Deze bespreken wij kort aan het eind van het rapport.

Inleiding

In Nederland kennen we voor alle ingezetenen een sociale zorgverzekering die uit twee onderdelen bestaat: een zorgverzekering op basis van de Zorgverzekeringswet (Zvw) voor de curatieve zorg en een verzekering op basis van de Wet langdurige zorg (Wlz)⁴ voor de langdurige zorg. Iedereen heeft gelijke toegang tot het op basis van deze wetten geregelde basispakket, ongeacht zijn/haar inkomenspositie en ook ongeacht zijn/haar leeftijd en gezondheidstoestand. Het verzekeringsstelsel steunt dus op onderlinge solidariteit. We brengen als samenleving gezamenlijk de financiële middelen op voor deze sociale zorgverzekering. Om recht te doen aan de onderlinge solidariteit en draagvlak voor een collectief gefinancierde verzekering te behouden, is het nodig dat de overheid zorgvuldig afweegt welke zorg wel en welke zorg niet uit de sociale verzekering wordt betaald.⁵

De keuze van de overheid om ‘de stand van de wetenschap en praktijk’ mede bepalend te laten zijn voor de begrenzing van het verzekerde pakket van de Zvw en de Wlz komt hieruit voort. Het komt erop neer dat alleen zorg die als effectief⁶ wordt beschouwd, onderdeel uitmaakt van het pakket van de Zvw en de Wlz.⁷

In dit rapport belichten wij dit voor onze sociale zorgverzekering geldende kernpunt. Wij schetsen allereerst de context waarin het begrip ‘de stand van de wetenschap en praktijk’ een rol speelt. Daarna beschrijven wij het beoordelingskader aan de hand waarvan kan worden vastgesteld of zorg aan genoemd criterium voldoet. Het beoordelingskader betreft een actualisering en daarmee een vervanging van het in 2007 aan de minister van VWS uitgebrachte rapport: “Beoordeling stand van de wetenschap en praktijk”.⁸ Wij gaan ook kort in op de beoordeling van ‘de stand van de wetenschap en praktijk’ van medische tests. Het betreffende hoofdstuk vormt samen met het rapport dat ons instituut in 2011 specifiek over de beoordeling van medische tests heeft uitgebracht⁹, de actuele stand van zaken wat betreft de beoordeling van medische tests.

Waarom deze update? Wij hebben vanaf 2007 tal van beoordelingen gedaan (uitmondend in een standpunt) met als gevolg dat onze werkwijze op onderdelen is geëvolueerd respectievelijk nader is ingevuld. Verder hebben wij in de afgelopen jaren een aantal onderwerpen uitgediept, zoals de invulling van het onderwerp ‘passend bewijs’ en de beoordeling van medische tests. Ook zijn er recente internationale ontwikkelingen waarbij wij ons aansluiten. Wij doelen dan op de zogenoemde GRADE-methode. Dit is een nieuwe manier om de kwaliteit van het beschikbare bewijs voor effectiviteit te beoordelen.¹⁰

Wij beschrijven in dit rapport uitgebreid hoe wij ‘anno nu’ het criterium ‘de stand van de wetenschap en praktijk’ beoordelen. Alle aspecten die mogelijk een rol kunnen spelen in onze beoordelingen komen daarbij aan bod. Dat wil niet zeggen dat ook alle beschreven aspecten in elke beoordeling worden betrokken. Welke aspecten een relevante rol spelen, hangt af van het specifieke dossier.

De opbouw van dit rapport is als volgt. Wij beginnen in hoofdstuk 1 met een introductie van het begrip ‘de stand van de wetenschap en praktijk’. Wij lichten dan onder meer het wettelijke kader toe en de legitimering voor het vaststellen van het beoordelingskader. In hoofdstuk 2 geven wij de hoofdlijnen van

4 De Wet langdurige zorg is per 1 januari 2015 in werking getreden en vervangt de AWBZ (Algemene Wet Bijzondere Ziektekosten).

5 De overheid hanteert een aantal uitgangspunten om tot een verantwoorde pakketsamenstelling te komen. Dit zijn de zogenoemde pakketprincipes, die zich als volgt laten samenvatten: is de zorg noodzakelijk, effectief (‘de stand van de wetenschap en praktijk’) en kosteneffectief. Bij de afweging of zorg onderdeel moet zijn van het verzekeringspakket wordt ook gekeken naar de uitvoerbaarheid. Met het criterium uitvoerbaarheid brengen wij in kaart welke factoren een succesvolle uitvoering van een pakketmaatregel (in- of uitstroom in/uit het pakket) kunnen belemmeren of bevorderen. Zie verder het rapport CVZ. Pakketbeheer in de praktijk deel 3. Diemen: CVZ, 2013. Beschikbaar via www.zorginstituutnederland.nl.

6 Waaronder begrepen: veilig.

7 Het vereiste van ‘de stand van de wetenschap en praktijk’ is neergelegd in respectievelijk het Besluit zorgverzekering en het Besluit langdurige zorg.

8 CVZ. Beoordeling stand van de wetenschap en praktijk. Diemen, 2007. Rapportnr. 254. Beschikbaar via www.zorginstituutnederland.nl.

9 CVZ. Medische tests. Beoordeling stand van de wetenschap en praktijk. Diemen, 2011. Rapportnr. 293. Beschikbaar via www.zorginstituutnederland.nl.

10 Wij hebben in het rapport CVZ. Pakketbeheer in de praktijk deel 3. Diemen, 2013, dat een uiteenzetting bevat van de wijze waarop het basispakket aan zorg (Zvw en AWBZ (thans Wlz)) wordt beheerd en waarin de samenhang van alle pakketprincipes wordt belicht, de update van het beoordelingskader ‘de stand van de wetenschap en praktijk’ al aangekondigd.

het beoordelingskader weer. Daarna volgen in hoofdstuk 3 onze werkwijze en de uitgangspunten die wij daarbij hanteren. De hoofdstukken 4 en 5 gaan over respectievelijk de langdurige zorg en medische tests. Hoofdstuk 6 gaat in op de ingangsdatum en de gevolgen van een standpunt over 'de stand van de wetenschap en praktijk'. Tot slot volgt op pagina 57 het besluit tot vaststelling van het rapport.

1 Introductie en beschouwingen vooraf

Voordat wij het beoordelingskader voor de toetsing aan de norm ‘de stand van de wetenschap en praktijk’ uiteenzetten, geven wij ter introductie een schets van de context waarin dit criterium een rol speelt.¹¹ Wij zoomen daarbij allereerst in op de Zvw en gaan daarna in een afzonderlijke paragraaf in op de Wlz. De Wlz is per 1 januari 2015 in de plaats gekomen van de Algemene Wet Bijzondere Ziektekosten (AWBZ).¹²

1.1 Zorgverzekeringswet

1.1.1 Wettelijk kader

1.1.1.1 Samenstelling van het basispakket van de Zvw

De overheid stelt de aard, inhoud en omvang van het basispakket van de Zvw vast. In de Zvw en in de daarop gebaseerde lagere regelgeving - het Besluit zorgverzekering (Bzv) en de Regeling zorgverzekering (Rzv) - is de omschrijving van het basispakket neergelegd. De prestaties waarop een zorgverzekering aanspraak moet bieden worden ook wel de ‘te verzekeren prestaties’ genoemd. Zorgverzekeraars zijn op grond van de Zvw verplicht om de te verzekeren prestaties in de zorgverzekering die zij op de markt brengen, in de polis op te nemen en te vertalen naar ‘verzekerde prestaties’.¹³ Deze prestaties vormen de dekking van de verzekeringsovereenkomst die de zorgverzekeraar met de individuele burger sluit.

1.1.1.2 Te verzekeren prestaties; open en gesloten systeem

De omschrijving van de ‘te verzekeren prestaties’ kent de volgende opbouw: Artikel 10 van de Zvw bevat een opsomming van de te verzekeren risico’s. Het betreft een globale typering van de prestaties waarop een zorgverzekering recht moet bieden. Het gaat om de volgende te verzekeren risico’s, te weten de behoefte aan:

- geneeskundige zorg;
- mondzorg;
- farmaceutische zorg;
- hulpmiddelenzorg;
- verpleging;
- verzorging;
- verblijf;
- vervoer.

In het Bzv en de Rzv zijn de inhoud en omvang van de in artikel 10 Zvw genoemde zorgvormen nader geregeld. Zo zijn in de artikelen 2.4 tot en met 2.15 van het Bzv de hiervoor genoemde zorgvormen verder uitgewerkt.¹⁴ De uitwerking varieert per onderdeel (per zorgvorm) en kent twee uitersten: generiek omschreven prestaties (open systeem) en specifiek omschreven prestaties (gesloten systeem).

- Generiek/open

Sommige zorgvormen heeft de regelgever in algemene termen (generiek) omschreven. Dit geldt onder meer voor geneeskundige zorg. Voor het omschrijven van die zorg heeft de regelgeving gebruik gemaakt

¹¹ Voor meer informatie over de context verwijzen wij naar het rapport CVZ. Pakketbeheer in de praktijk deel 3. Diemen, 2013. Beschikbaar via www.zorginstituutnederland.nl.

¹² De Wlz verzekert per 1 januari 2015 (instellings)zorg voor verzekerden die vanwege hun aandoening, beperking of handicap zeer kwetsbaar zijn en daarom een blijvende behoefte hebben aan permanent toezicht en zorg. Extramurale behandeling, verpleging en verzorging in de thuissituatie zijn van de AWBZ overgegaan naar de Zvw. Zorg gericht op participatie is overgegaan naar de Wet maatschappelijke ondersteuning (Wmo). Voor de jeugd is het grootste deel van de zorg overgeheveld naar de Jeugdwet.

¹³ Het betreft artikel 11 Zvw.

¹⁴ Artikel 11, derde lid, Zvw bepaalt dat bij algemene maatregel van bestuur de inhoud en omvang van de te verzekeren prestaties (die de zorgverzekeraars in de polissen moeten vertalen in verzekerde prestaties) nader worden geregeld. Deze algemene maatregel van bestuur vindt zijn uitwerking in het Bzv. Artikel 2.1, eerste lid, Bzv verwijst voor de zorg waarop de verzekerde recht heeft naar de artikelen 2.4 tot en met 2.15 van het Bzv.

van de formulering ‘plegen te bieden’. Zo is bijvoorbeeld bepaald dat ‘geneeskundige zorg’ zorg omvat zoals onder meer huisartsen, medisch specialisten, klinisch psychologen en verloskundigen die plegen te bieden (artikel 2.4, lid 1, van het Bzv). Verder geldt als nadere aanduiding (als onderdeel van de generieke omschrijving) dat de zorg moet voldoen aan ‘de stand van de wetenschap en praktijk’. Dit is geregeld in artikel 2.1, tweede lid, van het Bzv.¹⁵ Intramurale/specialistische geneesmiddelen vallen onder de te verzekeren prestatie ‘geneeskundige zorg’ en behoren derhalve tot deze categorie (generiek omschreven zorgvormen).¹⁶

Zorgvormen die (grotendeels) generiek zijn omschreven leveren een open systeem van te verzekeren prestaties op. Er vindt automatische in- en uitstroom van zorg in resp. uit het basispakket plaats. Voor medisch-specialistische zorg kennen we bijvoorbeeld een generieke omschrijving. Zorg die medisch specialisten plegen te bieden en die voldoet aan ‘de stand van de wetenschap en praktijk’ behoort tot de te verzekeren prestaties. Innovatieve zorg die (op een gegeven moment) aan die voorwaarden gaat voldoen (onder die generieke noemer valt), gaat als vanzelf tot de te verzekeren prestaties behoren. Voorafgaande toetsing en aanpassing van de regelgeving zijn daarvoor niet nodig. Zorg die op enig moment als obsoleet moet worden beschouwd, verdwijnt ‘automatisch’ uit het te verzekeren pakket. Aanpassing van de regelgeving is ook in dat geval niet vereist. De gekozen wettelijke formulering zorgt er als het ware voor dat er altijd een actueel, de laatste ontwikkelingen volgend, verzekeringspakket bestaat.

- Specifiek/gesloten

Andere zorgvormen zijn meer in detail geregeld (specifiek omschreven) en soms is zelfs sprake van een limitatieve opsomming. Dit laatste geldt voor extramurale farmaceutische zorg. Deze prestatie is op productniveau omschreven: de als te verzekeren prestaties aangewezen geneesmiddelen zijn met naam en toenaam in de bijlage bij de Rzv genoemd.

Deze - voor extramurale geneesmiddelen geldende - limitatieve positieve lijst vormt een gesloten systeem van te verzekeren prestaties. Er vindt in dat geval geen automatische in- en uitstroom plaats. De minister moet t.a.v. elk geneesmiddel vooraf beslissen of het in aanmerking komt voor toevoeging aan de lijst en daarmee voor toelating tot het basispakket. De positieve lijst van tot het basispakket behorende extramurale geneesmiddelen wordt ook wel Geneesmiddelen Vergoedingen Systeem (GVS) genoemd. De in het GVS opgenomen geneesmiddelen zijn voor zover mogelijk ingedeeld in groepen van onderling vervangbare geneesmiddelen en deze geclusterde geneesmiddelen zijn voorzien van een vergoedingslimiet.

- Twee normen in artikel 2.1, tweede lid, Bzv

Artikel 2.1, tweede lid, Bzv geeft naast ‘de stand van de wetenschap en praktijk’ nog een aanvullende norm. Er staat in het artikel: “De inhoud en omvang van de vormen van zorg of diensten worden mede bepaald door de stand van de wetenschap en praktijk en, bij ontbreken van een zodanige maatstaf, door hetgeen in het betrokken vakgebied geldt als verantwoorde en adequate zorg en diensten.” Er spelen hierbij twee punten : 1) is dit artikel van toepassing op alle in de Zvw opgenomen zorgvormen en 2) en op welke zorgvormen is de ene norm (‘de stand van de wetenschap en praktijk’) en op welke de andere norm (‘verantwoorde en adequate zorg/diensten’) van toepassing?

- Is artikel 2.1, tweede lid, Bzv van toepassing op alle Zvw-zorg?

Qua formulering en gezien de opbouw van artikel 2.1 Bzv is het tweede lid van toepassing op alle in de Zvw geregelde zorgvormen. Voor de extramurale geneesmiddelen geldt wel als bijzonderheid dat voorafgaande aan de pakkettoelating toetsing aan ‘de stand van de wetenschap en praktijk’ plaatsvindt. Een verzoek tot advisering over eventuele opname van een geneesmiddel op de limitatieve (GVS) lijst gaat altijd primair over de vraag: kan dit geneesmiddel met andere geneesmiddelen geclusterd worden? Om die vraag te kunnen beantwoorden hebben wij de beoordelingsuitkomsten ‘gelijke waarde,

¹⁵ Artikel 2.1, lid 2, Bzv luidt als volgt: De inhoud en omvang van de vormen van zorg of diensten worden mede bepaald door de stand van de wetenschap en praktijk en, bij ontbreken van een zodanige maatstaf, door hetgeen in het betrokken vakgebied geldt als verantwoorde en adequate zorg en diensten.

¹⁶ De omschrijving is: zorg zoals medisch specialisten die plegen te bieden en die voldoet aan ‘de stand van de wetenschap en praktijk’.

meerwaarde en minderwaarde' geïntroduceerd. In de praktijk komt het erop neer dat de begrippen 'gelijke waarde' of 'meerwaarde' equivalent zijn aan 'de stand van de wetenschap en praktijk'. Het in dit rapport opgenomen beoordelingskader geldt dan ook voor de beoordeling van GVS-geneesmiddelen.¹⁷

Voor de beoordelingsprocedure van extramurale geneesmiddelen, waarin daarnaast ook de onderbouwing van de doelmatigheid, de budgetimpact en de mogelijkheid tot clustering in verband met onderlinge vervangbaarheid aan bod komen, hebben VWS en Zorginstituut Nederland gezamenlijk gedetailleerde voorschriften opgesteld. Deze zijn terug te vinden in de uitgave "Procedure beoordeling extramurale geneesmiddelen" van 22 juni 2012.¹⁸ Voor wat betreft de advisering en besluitvorming over de toelating van extramurale geneesmiddelen verwijzen wij naar dat specifieke document.

- Welke van de twee normen is van toepassing?

Blijkens de Nota van toelichting bij het Bzv slaat de norm 'verantwoorde en adequate zorg en diensten' op zorgvormen die minder of geen wetenschappelijke status hebben of behoeven.¹⁹

Uit deze toelichting en ook uit de zinsnede 'bij ontbreken van een zodanige maatstaf' in artikel 2.1 Bzv volgt dat de norm 'verantwoorde en adequate zorg en diensten' alleen geldt als de andere norm ('de stand van de wetenschap en praktijk') niet van toepassing is. Het is dus niet zo dat de norm 'verantwoorde en adequate zorg en diensten' in beeld komt, indien bij een concrete beoordeling niet voldaan is aan de norm 'de stand van de wetenschap en praktijk', bijvoorbeeld vanwege zeer lage kwaliteit van het bewijs voor effectiviteit.

De norm 'de stand van de wetenschap en praktijk' is niet van toepassing, als (medisch) wetenschappelijke bewijsvoering geen enkele rol speelt respectievelijk kan spelen bij de vraag of de zorg/dienst in staat is te doen wat deze beoogt te doen. Dit geldt in ieder geval voor zittend ziekenvervoer²⁰ en voor de zogenoemde welzijnsgerelateerde hulpmiddelen. Dit zijn extramurale hulpmiddelen die de maatschappelijke participatie bevorderen, zoals bijvoorbeeld hulpmiddelen voor de communicatie²¹. Deze hulpmiddelen worden niet ingezet als onderdeel van een geneeskundige behandeling, maar zijn bedoeld om gevolgen van de beperkingen of de participatieproblemen veroorzaakt door de stoornis, te verminderen. Zij worden in de regel niet aan het lichaam gedragen. Op deze hulpmiddelen is de norm 'verantwoorde en adequate zorg en diensten' van toepassing. Ook voor deze norm geldt overigens dat op enigerlei wijze moet kunnen worden nagegaan of de zorg/dienst in staat is te doen wat deze beoogt te doen en of de veiligheid en het gebruiksgemak gewaarborgd zijn. Aan de hand van bijvoorbeeld een (gepubliceerde) praktijkevaluatie of consumentenonderzoek kan daarop een antwoord worden gegeven.²²

Tot op heden hebben wij geen andere zorg(vormen) geïdentificeerd ten aanzien waarvan ook geconcludeerd zou moeten worden dat de norm 'verantwoorde en adequate zorg en diensten' van toepassing is. Het kan zijn dat bij een toekomstige beoordeling de vraag opkomt welke van de twee normen van toepassing is. Wij zullen daar dan een gemotiveerd standpunt over innemen.

1.1.2 Bedoeling van de norm 'de stand van de wetenschap en praktijk'

Uit de wet(shistorie) en de jurisprudentie kan worden opgemaakt dat het bij 'de stand van de wetenschap en praktijk' gaat om de effectiviteit van de zorg: leidt het behandelbeleid (diagnostiek,

¹⁷ Uit het feit dat artikel 2.1, vijfde lid, van het Bzv bepaalt dat de minister bij ministeriële regeling kan regelen dat extramurale geneesmiddelen die niet voldoen aan 'de stand van de wetenschap en praktijk' toch worden toegelaten tot het basispakket, zij het tijdelijk en voorwaardelijk, kan worden afgeleid dat voor een definitieve pakkettoelating vereist is dat de zorg voldoet aan 'de stand van de wetenschap en praktijk'.

¹⁸ CVZ, Ministerie van VWS. Procedure beoordeling extramurale geneesmiddelen. Diemen, 2012. De voorschriften betreffen o.a. dossiereisen (waaronder het gebruik van vastgestelde formats), richtlijnen voor farmaco-economisch onderzoek en budgetimpact-schatting, inspraakmogelijkheid en de criteria voor beoordeling van de therapeutische waarde. Dit document is terug te vinden op www.zorginstituutnederland.nl.

¹⁹ Zie de artikelsgewijze toelichting bij artikel 2.1 van het Besluit zorgverzekering (Stb. 2005, 389).

²⁰ De Nota van Toelichting bij het Bzv noemt zittend ziekenvervoer als voorbeeld: daarvoor geldt de norm 'verantwoorde en adequate zorg/diensten' (Stb. 2005, 389).

²¹ Zie artikel 2.6, onder s, van de Rzv: uitwendige hulpmiddelen gerelateerd aan en ter compensatie van beperkingen in het spreken. Denk bijvoorbeeld aan computers, telefoons. Een insulinepomp of een pijnpomp wordt ook het lichaam gedragen, maar maakt onderdeel uit van een geneeskundige behandeling. Daarop is niet de norm 'verantwoorde en adequate zorg/diensten' van toepassing, maar de norm 'de stand van de wetenschap en praktijk'.

²² Zie CVZ. Beoordelingskader hulpmiddelen. Diemen, 2008. Rapportnr. 258.

behandeling), gelet op de gunstige en de ongunstige gevolgen (bijwerkingen, veiligheid) ervan, tot relevante (meer)waarde voor de patiënt? Het gaat dan bijvoorbeeld om genezing, een relevante mate van levensverlenging, van stijging van kwaliteit van leven, van afremming van verdere achteruitgang of - in geval van palliatieve zorg - verlichting van lijden.

Dat betaling uit het collectief (de Zvw en de Wlz) alleen gerechtvaardigd is als de zorg daadwerkelijk nuttig (effectief) is voor de gezondheid respectievelijk het welzijn van degenen die de zorg ontvangen, is een breed gedeelde opvatting en werd al genoemd in 1991 in het onder voorzitterschap van prof. Dunning opgestelde rapport “Kiezen en delen”.²³

Overigens is het zo dat als algemeen principe in de gezondheidszorg geldt dat professionals, vanuit het oogpunt van veiligheid, geen zorg (buiten onderzoeksverband) aanbieden die nog onvoldoende op effectiviteit is onderzocht. Het leidende principe is immers: “primum non nocere, in dubio abstinere”. Dus: “ten eerste: geen kwaad doen, bij twijfel: niets doen”.

1.1.3 Legitimering vaststellen beoordelingskader ‘de stand van de wetenschap en praktijk’

Waar komt de bemoeienis van ons instituut met het onderwerp uit voort? Anders gezegd: wat is de legitimering voor het vaststellen van het beoordelingskader? Deze is gelegen in de ons bij wet toebedeelde taken op het gebied van pakketbeheer. Hierna lichten wij dit toe.

1.1.3.1 Taken Zorginstituut Nederland op gebied van pakketbeheer

Het bij wet aan ons instituut toebedeelde takenpakket voor pakketbeheer bestaat – voor zover in dit kader relevant – uit de volgende onderdelen: duiding van zorg, advisering over SKGZ-geschillen en advisering over de inhoud van het te verzekeren pakket aan de minister van VWS.

- Duiding van zorg

Zoals gezegd, zijn in de Zvw, het Bzv en de Rzv de aard, inhoud en omvang van de te verzekeren prestaties omschreven en zijn de zorgverzekeraars verplicht deze in de polissen te vertalen naar verzekerde prestaties.²⁴ De prestatiebeschrijvingen in de polissen zijn – evenals de te verzekeren prestaties – grotendeels generiek omschreven en in de polissen is ook ‘de stand van de wetenschap en praktijk’ als begrenzing van de aanspraken terug te vinden. Het is primair aan de zorgverzekeraar om te beslissen of een bepaalde interventie onder de dekking van de polis valt. Dat betekent dat de verzekeraar in eerste instantie zelf nagaat of voldaan is aan het vereiste in de polis, waaronder de vraag of de zorg voldoet aan ‘de stand van de wetenschap en praktijk’.²⁵

Ons instituut is in de positie om gevraagd (als bijvoorbeeld een zorgverzekeraar er zelf niet uitkomt of twijfels heeft) en ook ongevraagd over een bepaalde interventie een standpunt in te nemen betreffende ‘de stand van de wetenschap en praktijk’. Dat berust op de wettelijke taak van ons instituut, namelijk om

²³ Commissie Keuzen in de Zorg (Commissie Dunning). Kiezen en delen. Rapport. Den Haag. 1991.

²⁴ Een verzekerde ontleent zijn/haar recht op (vergoeding van) zorg aan de verzekeringsovereenkomst die hij/zij sluit met een zorgverzekeraar. In de polisvoorwaarden staat dus omschreven waar de verzekerde recht op heeft. De officiële term daarvoor is: de verzekerde prestaties.

²⁵ Een zorgverzekering berust op een private verzekeringsovereenkomst tussen verzekerde en zorgverzekeraar. Het is de zorgverzekeraar die deze verzekering uitvoert en op basis van de polisvoorwaarden beslist over de aanspraak op zorg of vergoeding. Het is dan ook in eerste instantie aan zorgverzekeraars om te beoordelen of zorg voldoet aan het criterium ‘stand van de wetenschap en praktijk’. Dat geldt niet alleen voor een enkele, individuele casus, maar ook voor meer voorkomende gevallen. Als een zorgverzekeraar zich afvraagt of hij bepaalde zorg, die geregeld wordt gedeclareerd, eigenlijk wel zou moeten vergoeden, omdat mogelijk niet wordt voldaan aan ‘de stand van de wetenschap en praktijk’ van die zorg, zal hij zich eerst zelf een oordeel moeten vormen over ‘de stand van de wetenschap en praktijk’. Pas in tweede instantie, als de zorgverzekeraar er niet zelf uitkomt of er twijfel blijft bestaan, kan ons instituut om een duiding worden gevraagd. Het is niet zo dat de beoordeling van ‘de stand van de wetenschap en praktijk’ bij uitsluiting bij ons instituut ligt en evenmin dat wij ook alle beoordelingen moeten doen. De geschetste rolverdeling vloeit voort uit het gekozen wettelijke systeem (generiek/open systeem), hetgeen de toelichting bij het Bzv en ook een rechterlijke uitspraak bevestigen. Zie de Nota van Toelichting bij het Bzv (Stb. 2005, 389) respectievelijk Rechtbank Midden Nederland, Afdeling Civiel recht, zaaknummer/rolnummer: 334021/KG ZA 12-888 in de zaak van Yo Medical BV tegen ZN, CZ en Z-Index BV.

een eenduidige uitleg van de aard, inhoud en omvang van de te verzekeren prestaties te bevorderen. Dit is geregeld in artikel 64 van de Zvw. Deze taak wordt ook wel de duidingstaak genoemd.²⁶

Om op transparante en toetsbare wijze aan onze duidingstaak wat betreft het onderdeel ‘de stand van de wetenschap en praktijk’ te kunnen voldoen, hebben wij het in dit rapport opgenomen beoordelingskader vastgesteld. Dat kader is ook relevant voor de zorgverzekeraars die – zoals gezegd – in eerste instantie zelf beoordelen of de zorg voldoet aan ‘de stand van de wetenschap en praktijk’. De zorgverzekeraars kunnen het beoordelingskader als leidraad gebruiken voor hun beoordelingen en benutten bij het contracteren van zorg bij zorgaanbieders.²⁷

- Advisering over SKGZ-geschillen

Ons instituut heeft tot taak om de Stichting Klachten en Geschillen Zorgverzekeringen (SKGZ) van advies te dienen over geschillen tussen zorgverzekeraars en verzekerden (artikel 114, lid 3, Zvw).²⁸ Bij een geschil kan ook beoordeling van ‘de stand van de wetenschap en praktijk’ aan de orde zijn. Wij passen dan het beoordelingskader toe.

- Advisering aan de minister van VWS

Artikel 66 Zvw bepaalt dat ons instituut de minister van VWS desgevraagd adviseert over voorgenomen beleid inzake aard, inhoud en omvang van de te verzekeren prestaties. Dit artikel biedt de basis voor de instroom- en uitstroomadviezen die wij aan de minister van VWS uitbrengen.

Wij hanteren bij de vraag of zorg moet worden toegevoegd aan of uitgezonderd van de te verzekeren prestaties van de Zvw een viertal criteria (de zogenoemde pakketprincipes), te weten noodzakelijkheid, effectiviteit, kosteneffectiviteit en uitvoerbaarheid. Deze principes zijn beschreven in het rapport “Pakketbeheer in de praktijk deel 3”.²⁹

Zoals eerder gezegd, draait het bij ‘de stand van de wetenschap en praktijk’ om de effectiviteit van de zorg. Bij het pakketprincipe effectiviteit en het criterium ‘de stand van de wetenschap en praktijk’ gaat het dan ook om hetzelfde. Dit betekent dat de globale definiëring van het begrip effectiviteit in paragraaf 1.1.2 ook van toepassing is op het pakketprincipe effectiviteit en dat wij bij de toetsing aan dit pakketprincipe op gelijke wijze te werk gaan als bij het beoordelen van ‘de stand van de wetenschap en praktijk’. Het beoordelingskader geldt dus ook voor de beoordeling van het pakketprincipe effectiviteit.

Een uitgebreide toelichting over de hierboven kort aangeduide taken van ons instituut op het gebied van pakketbeheer is terug te vinden in het in het eerder genoemde rapport, “Pakketbeheer in de praktijk deel 3”.²⁹

²⁶ Een standpunt van ons instituut is geen besluit in de zin van de Algemene wet bestuursrecht (Awb) en is formeel niet bindend. Een standpunt heeft echter wel materiële juridische betekenis. Het kan gezien worden als een gezaghebbend oordeel van een deskundige (ook wel ‘richtinggevende uitspraak’ genoemd). De verzekeraar kan het standpunt van ons instituut niet zomaar naast zich neerleggen. Als hij het standpunt niet volgt, zal hij dat goed moeten kunnen motiveren. De zorgverzekeraar kan hier door de Nederlandse Zorgautoriteit (NZA) op worden aangesproken in het kader van toezicht op de rechtmatige uitvoering. Omdat het om uitleg van een wettelijke bepaling gaat, is het uiteindelijk de rechter die het laatste woord heeft. Een standpunt van ons instituut kan ter beoordeling van de rechter komen in het kader van een geschil tussen de verzekerde en de verzekeraar. Ook een zorgaanbieder, producent of een andere betrokken partij kan een standpunt ter discussie stellen in een procedure bij de civiele rechter, bijvoorbeeld op grond van een actie uit onrechtmatige daad. De rechter geeft een zwaarwegende betekenis aan een standpunt van ons instituut, mits dit voldoende is onderbouwd en zorgvuldig tot stand gekomen is.

²⁷ Bij zorgaanbieders moet duidelijkheid bestaan over de manier van toetsen. Zij moeten – en dat geldt met name als zij door een zorgverzekeraar zijn gecontracteerd om zorg te verlenen – de patiënt/verzekerde een antwoord kunnen geven op de vraag of de interventie die wordt geboden zorg betreft conform ‘de stand van de wetenschap en praktijk’ (en dus – als ook aan andere gestelde voorwaarden wordt voldaan – onder de dekking van de verzekering valt). Zie de artikelsgewijze toelichting bij artikel 2.1 van het Bzv (Stb. 2005, 389).

²⁸ Zorgverzekeraars zijn op grond van artikel 114 Zvw verplicht om erin te voorzien dat verzekerden een geschil over de uitvoering van de zorgverzekering kunnen voorleggen aan een onafhankelijke instantie. De Stichting Klachten en Geschillen Zorgverzekeringen (SKGZ) is daartoe in het leven geroepen. De SKGZ doet op verzoek van de verzekerde een bindende uitspraak in een geschil. De SKGZ is wettelijk verplicht om ons instituut om advies te vragen, indien het geschil betrekking heeft op de in de Zvw benoemde te verzekeren prestaties. Deze advisering beperkt zich – blijkens de toelichting bij het artikel – tot de vraag of de vorm van zorg die de verzekerde vergoed wil krijgen tot het te verzekeren pakket van de Zvw behoort.

²⁹ CVZ. Pakketbeheer in de praktijk deel 3. Diemen, 2013. Beschikbaar via www.zorginstituutnederland.nl.

1.1.3.2 Beoordelingspraktijk: onderwerpen die wij beoordelen

Wij passen het beoordelingskader voor de beoordeling van 'de stand van de wetenschap en praktijk' toe in het kader van de drie hiervoor besproken taken, met de nadruk op duiding en advisering aan de minister van VWS. Over welke onderwerpen buigen wij ons?

- Duidingen

Bij de duidingen gaat het vaak om beoordeling van geneeskundige zorg (en dan met name medisch specialistische zorg). Beoordeling van extramurale hulpmiddelenzorg doet zich ook geregeld voor.

Op verzoek

Beoordeling vindt plaats in reactie op verzoeken uit de praktijk, bijvoorbeeld van zorgverzekeraars.²⁵ De aanleiding voor deze verzoeken is veelal financieel van aard. De verzoeken gaan vaak over zorg die een groot beslag op de totale kosten binnen de zorg leggen, zoals nieuwe (dure) interventies of zorg waarbij het om grote volumina (veel patiënten) gaat. Dan kan het relevant zijn om na te gaan of de zorg wel als effectief beschouwd kan worden en of de zorg wel terecht ten laste van de basisverzekering wordt geleverd. Ook komt het voor dat zorgverzekeraars de uitgaven voor een bestaande, al langer toegepaste interventie zonder duidelijke redenen sterker zien stijgen dan verklaarbaar is uit de incidentiecijfers. De vraag kan dan rijzen of er sprake is van een 'sluipende' indicatieuitbreiding en zo ja of de effectiviteit van de behandeling bij de betreffende indicatie dan eigenlijk wel voldoende vaststaat. Ook zijn er soms vragen over zorg die in eerste instantie veelbelovend leek te zijn, maar uiteindelijk toch qua effectiviteit teleurstellende resultaten laat zien, bijvoorbeeld vanwege niet voorziene ernstige complicaties bij patiënten.

Risicogerichte selectie

Daarnaast selecteren wij ook zelf onderwerpen op basis van eigen observaties en op basis van signalen van andere partijen in de zorg. De benadering die wij voor de selectie van onderwerpen volgen is 'risicogericht'.³⁰ Het gaat om onderwerpen waarbij risico's zichtbaar of denkbaar zijn op het gebied van toegankelijkheid van zorg, kostenontwikkelingen en gepast gebruik van zorg. De selectie en aanpak van de onderwerpen vinden op transparante wijze en zo veel mogelijk in overleg met partijen in de zorg plaats. Overigens is het zo dat het niet in alle gevallen draait om de beoordeling van 'de stand van de wetenschap en praktijk'. De gekozen aanpak van een geselecteerd onderwerp kan per onderwerp verschillen.

De geneesmiddelen die binnen het ziekenhuis worden gebruikt (intramurale/specialistische geneesmiddelen) vallen ook onder de noemer geneeskundige zorg ('zorg zoals medisch specialisten plegen te bieden'). Nieuwe intramurale/specialistische geneesmiddelen die voldoen aan 'de stand van de wetenschap en praktijk' stromen automatisch het verzekerde pakket in. Wij hebben voor deze geneesmiddelen een specifiek, risicogericht toetsingsbeleid ontwikkeld. Het komt er in het kort op neer dat wij een beoordelingstraject starten als de fabrikant voor het intramuraal/specialistisch geneesmiddel een therapeutische meerwaarde claimt en er sprake is van een kostenbeslag van ten minste € 2,5 miljoen euro. Wij beperken ons dan niet tot een duiding (toetsing aan 'de stand van de wetenschap en praktijk'). Wij toetsen dan ook op de pakketprincipes noodzakelijkheid, kosteneffectiviteit en uitvoerbaarheid³¹ en formuleren, afhankelijk van de uitkomst van deze beoordeling, een advies aan de minister van VWS. In dit traject gaat het dus om zowel de duidingstaak als om de adviestaak van ons instituut.³²

Voor de goede orde: uit het feit dat wij een intramuraal/specialistisch geneesmiddel vanwege een lager kostenbeslag dan € 2,5 miljoen euro niet hebben beoordeeld, mag niet worden afgeleid dat derhalve vaststaat dat de betreffende zorg tot het basispakket behoort.

³⁰ CVZ. Onderwerpen Pakketbeheer 2014. Een risicogerichte selectie. Diemen, 2014. Beschikbaar via www.zorginstituutnederland.nl.

³¹ Effectiviteit is ook een pakketprincipe, maar toetsing op dit principe is al gebeurd in het kader van de duiding.

³² Het toetsingsbeleid van het Zorginstituut Nederland betreffende intramurale/specialistische geneesmiddelen is neergelegd in het rapport van 3 december 2013: CVZ. Pakketbeheer specialistische geneesmiddelen. Diemen, 2013. Het is te downloaden via www.zorginstituutnederland.nl.

Programma Zinnige Zorg

In 2013 is ons instituut gestart met een programma om het in de praktijk geleverde en ten laste van de Zvw gedeclareerde zorg systematisch door te lichten op gepast gebruik van zorg, het programma Zinnige Zorg.³³ Uit de analyses die wij doen om onderwerpen te selecteren die vanuit het oogpunt van ‘gepast gebruik van zorg’ aandacht behoeven, kan o.a. naar voren komen dat het nodig is de effectiviteit te beoordelen van een bepaalde interventie.

- Advisering aan de minister van VWS

Zoals gezegd, beslist de minister t.a.v. elk extramuraal geneesmiddel vooraf of het voor toevoeging aan het basispakket in aanmerking komt. Ons instituut dient de minister hierbij van advies. Het in dit rapport opgenomen beoordelingskader passen wij daarbij toe. Verder is de uitgave “Procedure beoordeling extramurale Geneesmiddelen” van 22 juni 2012 op deze adviseringstaak van toepassing.¹⁸

Incidenteel adviseren wij over de in- dan wel uitstroom van andere zorgvormen. Voor het beoordelen van de effectiviteit hanteren wij dan het in dit rapport opgenomen beoordelingskader.

1.1.4 Voorwaardelijke toelating van zorg tot het pakket

Uitgangspunt van de Zvw is dat alleen zorg die voldoet aan ‘de stand van de wetenschap en praktijk’ tot het verzekerde pakket behoort. M.a.w. alleen zorg die als effectief wordt beschouwd, maakt onderdeel uit van het basispakket. Sedert 1 januari 2012 heeft de minister van VWS op grond van artikel 2.1, lid 5, van het Bzv de mogelijkheid om zorg die niet aan dat vereiste voldoet, toch toe te laten tot het verzekerde pakket, onder de voorwaarde dat binnen een bepaalde periode gegevens verzameld worden over de effectiviteit van deze zorg. Aan de hand van de verzamelde gegevens kan na afloop van de periode van voorwaardelijke toelating worden vastgesteld of de zorg voldoet aan ‘de stand van de wetenschap en praktijk’ en daarmee definitief onderdeel is van het verzekerde pakket. De mogelijkheid voor de minister van VWS om zorg voorwaardelijk toe te laten tot het basispakket geldt voor geneeskundige zorg, extramurale farmaceutische zorg en - per 1 januari 2015 - extramurale hulpmiddelen.³⁴

Ons instituut adviseert de minister van VWS over onderwerpen die voor voorwaardelijke toelating tot het basispakket in aanmerking komen. Het betreft onderwerpen die partijen één keer per jaar zelf kunnen aandragen. Verder betreft het onderwerpen ten aanzien waarvan wij een negatief standpunt hebben uitgebracht, omdat naar ons inzicht niet voldaan is aan ‘de stand van de wetenschap en praktijk’. Om de geschiktheid te beoordelen van een onderwerp voor voorwaardelijke toelating moet onder meer vastgesteld worden welke ‘evidence gap’ er is en of het door partijen ingediende voorstel voor effectiviteitsonderzoek voldoende is om deze ‘gap’ weg te nemen. Het in dit rapport opgenomen toetsingskader speelt daarbij een rol. Dat geldt ook voor de situatie na afloop van de voorwaardelijke toelating. Dan zullen wij moeten beoordelen of de verzamelde gegevens over de effectiviteit de ‘evidence gap’ hebben weggenomen en dus voldoende zijn om te concluderen dat de zorg voldoet aan ‘de stand van de wetenschap en praktijk’.

Verdere informatie over de mogelijkheid van voorwaardelijke toelating en de daarvoor geldende procedure is opgenomen op de website van ons instituut.³⁵

³³ Bij de systematische analyse van zorg die uit het verzekerde basispakket wordt vergoed, kijken wij niet alleen naar de behandeling als zodanig, maar vooral ook naar de toepassing ervan. Wij voeren een programma uit dat erin voorziet om in een 5-jarige cyclus de ICD 10-gebieden te doorlopen. Per ICD10-gebied voeren we een brede systematische analyse uit om eventuele aandachtspunten daarbinnen vanuit het perspectief van goede en zinnige zorg vast te stellen. Daarna volgt nader verdiepend onderzoek van deze aandachtspunten. Verdere informatie over het Programma Zinnige Zorg is terug te vinden op de website www.zorginstituutnederland.nl.

³⁴ Op 10 juni 2014 heeft de minister van VWS de Tweede Kamer geïnformeerd over de voorgenomen wijzigingen in de regeling voor voorwaardelijke toelating van zorg. Naast de uitbreiding met extramurale hulpmiddelen wordt de regeling flexibeler wat betreft de ingangsdatum van de voorwaardelijke toelating. De voorbereidingsprocedure zal eveneens worden gewijzigd en gefaseerd gaan verlopen. Zie <http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2014/06/10/kamerbrief-over-voorwaardelijke-toelating-tot-het-basispakket.html>.

³⁵ Op de website is een speciale rubriek ingericht over het onderwerp voorwaardelijke toelating tot het pakket van de Zvw: <http://www.zorginstituutnederland.nl/pakket/werkwijze+pakketbeheer/voorwaardelijke+toelating>.

1.2 Wet langdurige zorg

1.2.1 Wettelijk kader

Verzekerden hebben ingevolge de Wlz aanspraak op langdurige zorg. Deze wet verzekert (instellings)zorg voor verzekerden die vanwege een somatische of psychogeriatrische aandoening/beperking of vanwege een verstandelijke, lichamelijke of zintuiglijke handicap een blijvende behoefte hebben aan:

- permanent toezicht ter voorkoming van escalatie of ernstig nadeel, of
- 24 uur per dag zorg in de nabijheid, omdat men zelf niet in staat is om op relevante momenten hulp in te roepen en men, om ernstig nadeel te voorkomen, door fysieke of zware regieproblemen voortdurend begeleiding/zorg nodig heeft.³⁶

Het gaat kortom om mensen die vanwege hun aandoening, beperking of handicap zeer kwetsbaar zijn en daarom een blijvende behoefte hebben aan permanent toezicht of 24 uur zorg in de nabijheid.

De omschrijving van het voor genoemde groepen verzekerde pakket aan zorg is vastgelegd in artikel 3.1.1. van de Wlz. In het Besluit langdurige zorg (Blz) zijn de aard, inhoud en omvang van de verzekerde zorg nadere geregeld.

Net als in de Zvw is voor Wlz-zorg vastgelegd dat de inhoud en omvang van de zorg mede worden bepaald door 'de stand van de wetenschap en praktijk' en, bij het ontbreken van een zodanige maatstaf, door hetgeen in het betrokken vakgebied geldt als 'verantwoorde en adequate zorg en diensten'. Dit is opgenomen in artikel 3.1.1, lid 3, van het Blz.³⁷

De Wlz is per 1 januari 2015 in de plaats gekomen van de AWBZ.³²

1.2.2 Taken Zorginstituut Nederland op gebied van pakketbeheer

De taken die ons instituut heeft met betrekking tot de Wlz zijn in de kern vergelijkbaar met de pakketbeheertaken op het gebied van de Zvw. Voor zover in dit kader relevant, zijn dat: duiding van zorg, advisering bij indicatiegeschillen tussen individuele verzekerden en het Centrum indicatiestelling zorg (CIZ), en advisering aan de minister van VWS over de pakketsamenstelling. Deze taken zijn neergelegd in respectievelijk de artikelen 5.1.1, 10.3.1, en 5.1.3 van de Wlz.

1.2.3 Beoordelingspraktijk: onderwerpen die wij beoordelen

In de duidingen en adviezen in het kader van de Wlz (voorheen de AWBZ) kan – afhankelijk van het concrete onderwerp – de beoordeling van 'de stand van de wetenschap en praktijk' een rol spelen. Het is wel zo dat wij – in vergelijking met de Zvw – minder ervaring hebben met de toetsing van langdurige zorg aan dit criterium. De redenen waarom dat zo is zijn:

- De bepaling betreffende 'de stand van de wetenschap en praktijk' is pas per 1 januari 2012 in de AWBZ, waarvoor de Wlz recentelijk in de plaats is gekomen, opgenomen.
- Gegevens over de effectiviteit van de zorg zijn beperkt beschikbaar. In de care-sector (langdurige zorg) heeft zich – in tegenstelling tot de cure-sector (Zvw) – in veel mindere mate een cultuur ontwikkeld van onderzoek naar effectiviteit. Dit hangt samen met de aard van de aandoeningen en de geleverde zorg. Cliënten hebben vaak meerdere aandoeningen en matig tot ernstige beperkingen. Bij behandeling gaat het bijvoorbeeld om complexe, multidisciplinaire zorg waarvan het doel vooral ligt in het bevorderen van kwaliteit van leven en welzijn, het voorkomen van achteruitgang en het leren omgaan met de beperkingen. Het doel van de zorg is breder en meer diffuus dan geldt voor de op genezing gerichte zorg. Voor het meten van de effectiviteit van de zorg is een doelomschrijving noodzakelijk en deze is voor de langdurige zorg omvangrijker dan voor de curatieve zorg. Er zijn vaak meerdere uitkomstmaten van belang en deze zijn moeilijk eenduidig te meten en vaak niet goed onderling vergelijkbaar. Vanwege het multidisciplinaire karakter van langdurige zorg is het bovendien lastig de uitkomst van zorg aan één specifieke interventie van een professional of zorgvorm toe te schrijven.

³⁶ Zie artikel 3.2.1 Wlz.

³⁷ Ook voor de AWBZ, die per 1 januari 2015 is omgevormd tot Wlz, gold deze norm.

- Anders dan voor de curatieve zorg is de langdurige zorg slechts beperkt ingevuld door protocollen en richtlijnen van de diverse beroepsgroepen.

Het beoordelingskader dat wij in dit rapport uiteenzetten, is in principe ook van toepassing op Wlz-zorg en de langdurige zorg die per 1 januari 2015 van de AWBZ naar de Zvw is overgeheveld. Wij streven er naar deze zorg consequenter op effectiviteit te gaan beoordelen. Om hieraan te kunnen voldoen werken wij aan een op de langdurige zorg toegespitste invulling die rekening houdt met de genoemde specifieke kenmerken en omstandigheden van deze zorg. De activiteiten die daarop zijn gericht respectievelijk daarbij behulpzaam zijn, staan beschreven in hoofdstuk 4 (langdurige zorg).

1.3 Zorginstituut Nederland: nieuwe kwaliteitstaken

Per 1 april 2014 hebben wij een aantal nieuwe taken gekregen op het gebied van het bevorderen van kwaliteit van zorg.³⁸ Gangbare benaming voor dit onderdeel van ons takenpakket is: Kwaliteitsinstituut in de zorg. Bedoeling van de instelling van dit instituut is om te komen tot een eenduidig kader dat aangeeft wat wordt verstaan onder goede zorg en dat inzichtelijk maakt of en waar die goede zorg wordt verleend.³⁹ De Memorie van Toelichting verstaat onder goede zorg: "zorg van goed niveau, die in ieder geval veilig, doeltreffend, doelmatig en cliëntgericht is, tijdig wordt verleend en is afgestemd op de reële behoefte van de cliënt en waarbij personen die de zorg verlenen, handelen in overeenstemming met de op hen rustende verantwoordelijkheid voortvloeiende uit de voor hen geldende professionele standaard".⁴⁰ De nieuwe kwaliteitstaken, zoals het opstellen van een meerjarenagenda en het toetsen van kwaliteitsstandaarden en meetinstrumenten voor opname in het register, zijn erop gericht om het leveren van goede zorg en transparantie daarover te ondersteunen.

Tussen deze nieuwe taken en de bestaande taken op het gebied van pakketbeheer bestaat samenhang: ze raken aan de publieke belangen als toegankelijkheid, kwaliteit en betaalbaarheid van de gezondheidszorg.⁴¹ Uiteraard zullen wij bij de uitvoering van de verschillende wettelijke taken attent zijn op de samenhang en zo veel mogelijk synergie betrachten. Op welke wijze wij dat doen bij de beoordeling van 'de stand van de wetenschap en praktijk' in het kader van pakketbeheer bespreken wij hierna in hoofdstuk 2.

³⁸ Bij die gelegenheid is het College voor zorgverzekeringen (CVZ) omgevormd tot Zorginstituut Nederland.

³⁹ Zie het overzichtartikel van mr. M. van Dijen over de uitbreiding van taken: Dijen M van. *Bevordering van kwaliteit in de zorg. De zorgtaken van het Zorginstituut Nederland*. PS Documenta 2013; 3: 221-36.

⁴⁰ Kamerstukken II 2011/12, 33243, nr. 3, pag.3.

⁴¹ Kamerstukken II 2012/13, 33243, nr. 8. Het is overigens wel de bedoeling, zo blijkt uit de parlementaire behandeling van het wetsontwerp, dat de scheiding tussen de kwaliteitstaken en de pakketbeheertaken gewaarborgd is. Zie o.a. Kamerstukken II 2011/12, 33243, nr. 3 pag. 24.

2 Beoordelingskader: hoofdlijnen

Om onderdeel te zijn van het pakket van de Zvw en de Wlz moet de zorg voldoen aan ‘de stand van de wetenschap en praktijk’, d.w.z. als effectief kunnen worden beschouwd. In dit hoofdstuk geven wij de hoofdlijnen van ons beoordelingskader weer. Allereerst gaan wij globaal in op het proces. Daarna beschrijven wij wat wij bedoelen met het begrip effectiviteit en welke principes wij toepassen om de effectiviteit van een interventie te beoordelen. Daarbij gaan wij in op de twee elementen (wetenschap en praktijk) die in de beoordeling een rol spelen. Verder zetten wij uiteen welke benadering wij volgen indien de te beoordelen interventie in een richtlijn⁴² is beschreven. Tot slot volgt een paragraaf over de verzekeringsstatus van zorg in het buitenland.

In hoofdstuk 3 gaan wij vervolgens meer in detail in op de werkwijze en de daarbij gehanteerde uitgangspunten.

De hoofdstukken 2 en 3 gaan vooral op voor de Zvw⁴³, maar zijn mutatis mutandis ook van toepassing op de Wlz-zorg en de langdurige zorg die per 1 januari 2015 van de AWBZ is overgeheveld naar de Zvw. Wij gaan in een apart hoofdstuk (hoofdstuk 4) nader in op specifieke, voor de langdurige zorg geldende aspecten. Hoofdstuk 5 zoomt in op de beoordeling van medische tests.

2.1 Proces van beoordeling en standpuntbepaling

Wij hebben de beoordeling van de effectiviteit van interventies ingebed in een proces, dat waarborgt dat de benodigde relevante input beschikbaar komt en dat een weloverwogen standpunt kan worden ingenomen. Zo worden in beginsel professionals via hun wetenschappelijke verenigingen en – inmiddels ook – patiëntenverenigingen door middel van raadpleging en consultatie van meet af aan bij een beoordeling betrokken.⁴⁴ Ditzelfde geldt voor de zorgverzekeraars.

Om ons te verzekeren van inbreng van actuele wetenschappelijke kennis en van ervaring met de behandelpraktijk, heeft ons instituut een Wetenschappelijke Adviesraad (WAR) in het leven geroepen. Deze is multidisciplinair samengesteld en bestaat uit externe, onafhankelijke leden met deskundigheid en ervaring op het terrein van assessmentvraagstukken in de zorg.⁴⁵ De WAR adviseert de Raad van Bestuur van ons instituut op basis van de kwaliteit van het wetenschappelijke bewijs en van de overige overwegingen/argumenten die naar zijn inzicht een rol in de assessment spelen. De Raad van Bestuur weegt alle relevante informatie en formuleert op basis daarvan een standpunt over ‘de stand van de wetenschap en praktijk’.

Het oordeel over ‘de stand van de wetenschap en praktijk’ beschouwen wij als een afronding van de zogenaamde beoordelingsfase (ook wel assessment genoemd). Dit om een onderscheid te maken ten opzichte van de zogenaamde maatschappelijke waarderingsfase (ook wel appraisal genoemd). In de beoordelingsfase vindt oordeelsvorming plaats op basis van de beschikbare informatie over een interventie. In de waarderingsfase vindt een maatschappelijke toetsing plaats, waarin – naast het oordeel over ‘de stand van de wetenschap en praktijk’ (effectiviteit) – ook andere maatschappelijke argumenten een rol spelen. Waar in de beoordelingsfase de vraag centraal staat of de interventie voldoet aan het

42 Wij hebben het in dit rapport over richtlijnen, omdat voor de beoordeling van ‘de stand van de wetenschap en praktijk’ vooral dat soort documenten van belang is. Het overkoepelende begrip is ‘kwaliteitsstandaarden’. Dat is de verzamelnaam die wij in het kader van onze kwaliteitstaak gebruiken voor landelijk opgestelde richtlijnen, zorgstandaarden en zorgmodules.

43 Voor extramurale geneesmiddelen gelden wel enige bijzonderheden. Daarvan hebben wij in paragraaf 1.1.1.2 melding gemaakt en onder meer verwezen naar de beoordelingsprocedure. Deze is terug te vinden in de uitgave: Procedure beoordeling extramurale Geneesmiddelen. Diemen, Ministerie van Volksgezondheid, Welzijn en Sport en CVZ, Ministerie van VWS. Procedure beoordeling extramurale geneesmiddelen. Diemen, 2012.

Voor de hulpmiddelenzorg is voor de toetsing aan de pakketprincipes, waaronder de effectiviteit, een apart toetsingskader opgesteld: CVZ. Beoordelingskader hulpmiddelenzorg. Diemen, 2008. Rapportnr. 258. Wat betreft de vraag of een hulpmiddel als effectief kan worden beschouwd (voldoet aan ‘de stand van de wetenschap en praktijk’) wordt in het toetsingskader hulpmiddelenzorg verwezen naar het rapport over beoordeling ‘de stand van de wetenschap en praktijk’. Het beoordelingskader hulpmiddelenzorg is te vinden op www.zorginstituutnederland.nl.

44 Zo nodig – afhankelijk van de te beoordelen interventie – zullen wij ook (koepels) van geneesmiddelen- en hulpmiddelenfabrikanten bij de beoordeling betrekken.

45 Het gaat o.a. om epidemiologen, clinici, farmaco-economen. Zie ook: CVZ. Reglement Wetenschappelijke Adviesraad. Diemen, 2013. Beschikbaar via www.zorginstituutnederland.nl.

wettelijke criterium ‘de stand van de wetenschap en praktijk’, staat in de waarderingsfase de vraag centraal of de interventie thuishoort in het verzekerde pakket. In beide fasen wordt ons instituut geadviseerd door respectievelijk de eerder genoemde WAR en de Adviescommissie Pakket (ACP).⁴⁶

Wij realiseren ons dat er aan het beoordelen van ‘de stand van de wetenschap en praktijk’ ook maatschappelijke aspecten kunnen zijn verbonden. Denk hierbij bijvoorbeeld aan de keuze van de comparator, de gewenste uitkomstmaten en het gewenste effectverschil. Het kan daarom gewenst zijn om de ACP hierover te raadplegen.

2.2 Relatieve effectiviteit

Zoals eerder gezegd, draait het bij ‘de stand van de wetenschap en praktijk’ om de effectiviteit van de zorg, hetgeen betekent: leidt het behandelbeleid (diagnostiek, behandeling), gelet op de gunstige en de ongunstige gevolgen (bijwerkingen, veiligheid) ervan, tot relevante (meer)waarde voor de patiënt? Meer concreet gaat het ons om de relatieve effectiviteit van een interventie: in welke mate draagt de (nieuwe) interventie bij aan het met de interventie beoogde doel in vergelijking met datgene wat in de praktijk al aan zorg voor de betreffende aandoening wordt geboden? Kortom: wat is de ‘netto toevoeging’ van de nieuwe interventie in vergelijking met de al bestaande zorg?⁴⁷ Dit uitgangspunt sluit aan bij de internationale praktijk en wordt aangeduid met de term Comparative Effectiveness danwel Relative Effectiveness.⁴⁸

Het komt erop neer dat wij - om een standpunt te kunnen innemen over de effectiviteit van een interventie - een vergelijking maken. Om die vergelijking te kunnen maken bepalen wij per te beoordelen interventie het volgende:

- Met welke interventie vergelijken wij de te beoordelen interventie?
- Op welke doelgroep richt de interventie zich?
- Welke uitkomsten vinden wij cruciaal/belangrijk en betrekken wij in de vergelijking?
- Vinden wij de uitkomst van de vergelijking (het relatieve effect) groot genoeg (m.a.w. is de uitkomst klinisch relevant)?

Op basis hiervan - en afhankelijk van de vraag of wij voldoende vertrouwen in de uitkomst van de vergelijking kunnen hebben - trekken wij onze conclusie.

Voor de goede orde: het gaat ons dus niet uitsluitend om de werkzaamheid van een interventie. Voor een positief oordeel op de vraag (voldoet de interventie wel of niet aan ‘de stand van de wetenschap en praktijk’) moet (ook) in voldoende mate zijn aangetoond dat de interventie in de praktijk een klinisch relevant effect heeft op de gezondheid en/of het welbevinden van de patiënt. Met werkzaamheid (de Engelse term is: efficacy) wordt bedoeld op het vermogen om in een ideale en gecontroleerde setting een bepaald effect te bereiken, bijvoorbeeld in een laboratoriumsetting. Voor onderzoek naar de werkzaamheid van een interventie kan placebo-gecontroleerd onderzoek nuttige informatie opleveren. De effectiviteit van een interventie heeft betrekking op het bereiken van een klinisch relevant doel of doelen in de dagelijkse praktijk. We spreken bijvoorbeeld van werkzaamheid als de interventie leidt tot een goede uitslag van de sputumkweek na een kuur bij bronchitis, maar we spreken over effectiviteit wanneer het gaat om (de snelheid van) het daadwerkelijk klinisch herstel van de patiënt. Effectiviteit (effectiveness) omvat dus meer dan werkzaamheid (efficacy).⁴⁹

⁴⁶ De ACP wordt door de minister van VWS benoemd en bestaat uit deskundigen met maatschappelijke kennis en ervaring.

⁴⁷ In paragraaf 3.1.2 lichten wij toe waarmee wij de te beoordelen interventie vergelijken.

⁴⁸ Zie o.a. de volgende EUnetHTA documenten:

- EUnetHTA. JA WP5: Relative effectiveness assessment (REA) of pharmaceuticals, July 2011.

- EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

⁴⁹ Dit heeft consequenties voor de keus van de uitkomstmaten in een uit te voeren onderzoek: studies naar de werkzaamheid van een interventie gebruiken meestal andere uitkomstmaten (bijvoorbeeld bepaalde laboratoriumparameters) dan studies naar de effectiviteit van een interventie. In die studies draait het vooral om klinische uitkomstmaten met veelal een langere tijdshorizon, zoals morbiditeit en kwaliteit van leven. In paragraaf 3.1.3 gaan wij in op de keuze van de uitkomstmaten.

2.3 Toepassen principes van evidence-based medicine

Vaste werkwijze van ons instituut is om - ter beantwoording van de vraag of een interventie als effectief is te beschouwen - de principes van evidence-based medicine (EBM) te volgen. EBM komt voort uit de klinische praktijk en dient ertoe om tegenwicht te bieden aan geneeskunde die uitsluitend steunt op persoonlijke (subjectieve), op ervaring berustende inzichten van de behandelend arts. EBM is inmiddels een internationaal aanvaarde denkwijze om klinisch handelen op te baseren en kan als volgt worden gedefinieerd:

“Evidence based medicine is het zorgvuldig, expliciet en oordeelkundig gebruik van het huidige beste bewijsmateriaal om beslissingen te nemen voor individuele patiënten. De praktijk van evidence-based medicine impliceert het integreren van individuele klinische expertise met het beste externe bewijsmateriaal dat vanuit systematisch onderzoek beschikbaar is.”⁵⁰

EBM is primair ontwikkeld als leidraad voor praktiserende artsen voor het nemen van klinische beslissingen over individuele patiënten. De principes van EBM vinden echter bredere toepassing en worden ook gevolgd bij de ontwikkeling van richtlijnen door de beroepsverenigingen van zorgverleners. EBM is voor de richtlijnontwikkeling al lange tijd de nationale en internationale standaard.⁵¹ Ook bij het ontwikkelen van beleid op het terrein van volksgezondheid worden de principes van EBM gevolgd.⁵² In die gevallen gaat het niet om beslissingen over individuele patiënten, maar om aanbevelingen/beslissingen op populatieniveau.

EBM is ook goed bruikbaar om te beoordelen of zorg voldoet aan ‘de stand van de wetenschap en praktijk’. Ook daarvoor geldt dat het niet gaat om de vraag of de zorg bij een individuele patiënt effectief is, maar of de zorg bij een bepaald indicatiegebied effectief is. Deze vraag moet, gezien de formulering in het Bzv, beantwoord worden aan de hand van ‘de stand van de wetenschap en praktijk’. De formulering duidt erop dat het niet om twee losse criteria gaat (‘de stand van de wetenschap’ en ‘de stand van de praktijk’) die ieder afzonderlijk beoordeeld moeten worden en potentieel een tegenstrijdig resultaat zouden kunnen geven. De formulering geeft aan dat het één geïntegreerde wettelijke maatstaf betreft.⁵³ EBM combineert en verenigt juist de beide elementen, wetenschap en praktijk. Zij propageert immers de integratie van inzichten die voortkomen uit wetenschappelijk onderzoek met in de praktijk gevormde expertise en ervaring van zorgverleners en zorggebruikers.⁵⁴

EBM wordt vaak geassocieerd met gecontroleerde klinische onderzoeken, zoals (dubbelblinde) randomized controlled trials (RCT's) en met meta-analyses van zulke RCT's. Het is belangrijk om te benadrukken dat het toepassen van de principes van EBM niet wil zeggen dat er alleen een positief standpunt over de effectiviteit van een interventie kan worden ingenomen als er sprake is van uit dergelijk onderzoek blijken positieve resultaten over het effect van de interventie. ‘Harde evidence’ zoals uit RCT's is niet per definitie vereist; ook evidence uit andersoortige studies, zoals observationeel onderzoek, kan tot een positieve waardering en tot de kwalificatie ‘evidence-based’ leiden.⁵⁵ Uiteindelijk

50 Grondlegger van de EBM-benadering is Sackett (en zijn collega's). Zie Sackett DL, Rosenberg WMC, Gray JAM, et al. Evidence based medicine: what it is and what it isn't. BMJ 1996; 312:71-2. Het citaat is ontleend aan: Scholten RJPM, Offringa M, Assendelft WJJ van. Inleiding in evidence-based medicine. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

51 Zie paragraaf 6.1 van Scholten RJPM, Offringa M, Assendelft WJJ van. Inleiding in evidence-based medicine. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

52 Een voorbeeld is: het besluit van de overheid om een pilot te starten voor de screening op darmkanker via de ontlasting.

53 Een recente rechterlijke uitspraak bevestigt dit. Zie: Vonnis Rechtbank Amsterdam van 19 november 2014, gepubliceerd onder nummer ECLI:NL:R-BAMS:2014:7727.

54 Bij het opstellen van behandelrichtlijnen gaat het ook om de integratie van wetenschap en praktijk. Een richtlijn is een document met aanbevelingen, gericht op het verbeteren van de kwaliteit van zorg, berustend op systematische samenvattingen van wetenschappelijk onderzoek en afwegingen van de voor- en nadelen van de verschillende zorgopties, aangevuld met expertise en ervaringen van zorgprofessionals en zorggebruikers. Het gaat daarbij om de integratie van de volgende complementaire bronnen van kennis: wetenschappelijke kennis, de expertise van zorgverleners en de ervaringen en voorkeuren van patiënten. Zie p. 3 en p. 42 van: Everdingen JJE van, Dreesens DHH, Burgers JS. Handboek evidence-based richtlijnontwikkeling. 2e druk. Houten: Bohn Stafleu van Loghum, 2014.

55 In de Engelstalige literatuur wordt met opzet gesproken van ‘evidence’ en niet van ‘proof’. Evidence is strikt genomen niet hetzelfde als bewijs: het is een aanwijzing die zo sterk kan zijn dat nauwelijks twijfel hoeft te bestaan over de juistheid, of zo zwak dat zij nauwelijks overtuigt. Ontleend aan paragraaf 1.1.1 van Scholten RJPM, Offringa M, Assendelft WJJ van. Inleiding in evidence-based medicine. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

gaat het om de 'kwaliteit van de evidence', zijnde een maat voor de geloofwaardigheid van de resultaten.⁵⁶ EBM wil evenmin zeggen dat alleen aandacht wordt geschonken aan 'harde' uitkomstmaten, zoals mortaliteit. Ook 'zachte(re)' eindpunten, zoals kwaliteit van leven, worden – afhankelijk van de te beoordelen interventie – betrokken bij de beoordeling.

De kern van EBM is dat de beschikbare evidence – die uiteen kan lopen van gerandomiseerd vergelijkend onderzoek tot praktijkervaring – systematisch is gezocht en geselecteerd en op gestructureerde wijze is gewogen en gebruikt. Als uitgangspunt geldt verder dat wetenschappelijk onderzoek van hoge kwaliteit het zwaarste weegt. Het werken volgens de EBM-principes kent een aantal vaste stappen: het zoeken en selecteren van informatie, het beoordelen van de gevonden informatie en het trekken van een conclusie.

Voor een uitgebreide beschrijving van de principes van EBM verwijzen wij naar de volgende publicaties: "Inleiding in evidence-based medicine"⁵⁷ en "Handboek evidence-based richtlijn ontwikkeling"⁵⁸.

2.4 Wetenschap en praktijk: elementen van integrale beoordeling

Zoals gezegd, betekent EBM dat naast wetenschappelijke inzichten ook de in de praktijk gevormde expertise en ervaringen van zorgverleners en zorggebruikers bij de beoordeling in ogenschouw worden genomen. In onze werkwijze komt dit laatste op verschillende manieren tot uiting:

- Wij maken gebruik van bestaande (internationale) EBM-richtlijnen en sluiten ons daar, indien mogelijk⁵⁹, bij aan. Startpunt van een beoordeling is dan ook altijd dat wij nagaan of de te beoordelen interventie in een richtlijn is opgenomen.
- Indien wij zelf een literatuurreview (laten) opstellen over het te beoordelen onderwerp⁶⁰, aan de hand waarvan wij de kwaliteit van het wetenschappelijke bewijs van een interventie bepalen, zullen wij op een aantal punten beargumenteerde keuzes moeten maken. Wij gaan hiervoor te rade bij de wetenschappelijke verenigingen van de professionals en de patiëntenorganisaties. Vanuit hun praktische kennis en ervaring kunnen zij ons relevante informatie aanreiken. Het gaat dan om de keuze van de comparator en de keuze van de uitkomstmaten. Verder gaat het om: wat is het minimaal vereiste klinisch relevante verschil in uitkomst en wat is de minimaal vereiste follow-up periode?
- Onderdeel van de beoordeling is om een zogenaamd 'passend bewijsprofiel' op te stellen. Dat betreft een inschatting van het soort onderzoek dat – gegeven de interventie en de indicatie – praktisch gezien wenselijk en haalbaar is, oftewel 'passend' is. Voor het opstellen van dit profiel gebruiken wij de door wetenschappelijke verenigingen van professionals en patiëntenorganisaties verschaft informatie.
- Inhoudelijke en praktische kennis over de interventie winnen wij in bij de betreffende Nederlandse wetenschappelijke verenigingen. Deze zijn vanuit de specifieke deskundigheid en de ervaring in de praktijk in staat om eventueel ontbrekende relevante informatie of literatuur aan te vullen of om bij technische interpretatie van gegevens behulpzaam te zijn.
- Wij bespreken onze voorlopige conclusies met de wetenschappelijke verenigingen van de professionals en de patiëntenorganisaties of leggen deze aan hen voor. Dat biedt hun de gelegenheid om commentaar te geven en om eventuele in de praktijk opgebouwde expertise en ervaring in te brengen. Afhankelijk van de concrete casus overleggen wij eventueel ook over nadere onderbouwing en/of garanties van de kant van de professionals en/of patiënten.

De beoordeling van de kwaliteit van het gevonden wetenschappelijk bewijs, leidt niet automatisch, als in een beslisboom, tot een positief of negatief standpunt over de vraag of de interventie voldoet aan 'de stand van de wetenschap en praktijk'. Naast argumenten voortkomend uit de kwaliteit van het wetenschappelijk bewijs, kunnen ook argumenten van professionals en patiënten een rol spelen. Dit doet recht aan de behoefte aan nuance in het pakketbeheer, maar stelt anderzijds hoge eisen aan onze werkwijze. Het vergt

⁵⁶ Zie paragraaf 1.1.1 van Scholten RJPM, Offringa M, Assendelft WJJ van. *Inleiding in evidence-based medicine*. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

⁵⁷ Scholten RJPM, Offringa M, Assendelft WJJ van. *Inleiding in evidence-based medicine*. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

⁵⁸ Everdingen JJE van, Dreesens DHH, Burgers JS. *Handboek evidence-based richtlijnontwikkeling*. 2e druk. Houten: Bohn Stafleu van Loghum, 2014.

⁵⁹ Het is dan wel van belang de kwaliteit, de eventuele gedateerdheid en de onafhankelijkheid van de richtlijn vast te stellen. Zie verder paragraaf 2.5.

⁶⁰ Dit doen wij in ieder geval als een (bruikbare) richtlijn ontbreekt.

overleg met de relevante partijen in de zorg en een goede onderbouwing van onze afwegingen en daarop gestoelde conclusie.

De weging van bevindingen en argumenten en het formuleren van een conclusie zijn uiteindelijk voorbehouden aan ons instituut. De weging zal van geval tot geval moeten worden gemaakt. Het gaat er in de kern om dat wij – alles afwegende – er in voldoende mate van overtuigd zijn dat de interventie bij patiënten met de betreffende indicatie(s) effectief is, d.w.z. in voldoende mate gezondheidswinst oplevert. Over de weging kan in grote lijnen het volgende worden opgemerkt:

- Als er wetenschappelijk bewijs is van hoge kwaliteit op cruciale of belangrijke uitkomsten zal dat bewijs in beginsel doorslaggevend zijn. De geloofwaardigheid van de gevonden resultaten is dan immers heel groot. Dit gaat op voor zowel positieve als negatieve resultaten: als er bewijs van hoge kwaliteit is voor een positief effect van de interventie dan leidt dit in beginsel tot een positief standpunt. Is er bewijs van hoge kwaliteit voor een negatief effect van de interventie, dan leidt dit in principe tot een negatief standpunt.
- Bewijs van zeer lage kwaliteit op cruciale of belangrijke uitkomstmaten betekent dat de geloofwaardigheid van de gevonden resultaten – of die nu positief of negatief zijn – uiterst gering is. Dit vormt een belemmering om te concluderen of de interventie nu wel of niet effectief is. Dit gebrek aan bewijs van voldoende kwaliteit is in beginsel van doorslaggevende betekenis en zal leiden tot een negatief standpunt. Een positieve waardering van de interventie bij professionals of patiënten – mocht daar sprake van zijn – zal in de regel de grote onzekerheid over de gevonden effecten niet kunnen compenseren.
- Als er wetenschappelijk bewijs is van middelmatige⁶¹ of lage kwaliteit voor een positief effect op cruciale dan wel belangrijke uitkomsten, dan kunnen in de praktijk opgedane inzichten en ervaringen eventueel van betekenis zijn en bijdragen aan een positief standpunt. Er zal dan wel aan een aantal voorwaarden moeten zijn voldaan, waaronder in ieder geval:
 - Er zijn overtuigende redenen om aan te nemen dat verder onderzoek dat de kwaliteit van het bewijs naar de effectiviteit van de interventie zou kunnen verhogen, zeer waarschijnlijk niet zal (kunnen) plaatsvinden.⁶²
 - Professionals hebben (bij voorkeur met steun van de patiënten) deugdelijk en overtuigend onderbouwd waarom zij, ondanks de middelmatige of lage kwaliteit van het bewijs, de interventie toch bij (een) bepaalde indicatie(s) als een effectieve behandelmethode beschouwen.
 - Deze opvatting moet stoelen op brede consensus binnen en tussen de relevante beroepsgroep(en) en in een publicatie van de wetenschappelijke vereniging(en) zijn geformaliseerd, bijvoorbeeld in een richtlijn of consensusdocument.⁶³

Voor de goede orde: hetgeen in deze derde bullit staat, gaat alleen op als er sprake is van een positief effect. Duidt het bewijs van middelmatige of lage kwaliteit op een negatief effect, dan is dat onvoldoende om een positief standpunt over ‘de stand van de wetenschap en praktijk’ in te nemen.

Dit vereiste (brede consensus binnen/tussen de relevante beroepsgroep(en)) volgt uit het normatieve karakter van het begrip ‘de stand van de wetenschap en praktijk’. ‘Praktijk’ is niet: ‘dat wat individuele professionals (gewoonlijk) feitelijk aan zorg leveren’. Evenmin gaat het bij ‘praktijk’ om de mening van individuele behandelaars (en individuele patiënten) over de waarde van de interventie.⁶⁴ Wat

⁶¹ Wij kiezen voor de term ‘middelmatig’ als vertaling van ‘moderate’. In paragraaf 3.2.4 over de GRADE-systematiek komt dat verder aan bod.

⁶² Zie verder paragraaf 3.2.3 die betrekking heeft op ‘passend bewijs’ en ingaat op de vraag wanneer gerandomiseerde vergelijkend onderzoek (RCT) niet wenselijk of niet haalbaar is.

⁶³ Wij noemen in deze opsomming niet de situatie dat er geen enkel bewijs is. Bij de beoordelingen die wij doen komt dit vrijwel niet voor; er is altijd wel enige mate van bewijs voor effectiviteit. Mocht dit bij uitzondering anders liggen en er geen enkel bewijs voor effectiviteit zijn, dan is dat uiteraard zwaarwegend en dat zal dan leiden tot een negatief standpunt over ‘de stand van de wetenschap en praktijk’. De handelwijze in de praktijk kan daarin geen verandering brengen.

⁶⁴ Een recente rechterlijke uitspraak bevestigt dit. De Rechtbank Amsterdam vindt het in de betreffende kwestie terecht dat ons instituut verklaringen van individuele urologen, inhoudende dat volgens hun observaties de behandeling in de praktijk (in een aantal gevallen) helpt, niet heeft meegewogen bij de beoordeling van ‘de stand van de wetenschap en praktijk’, nu er sprake is van wetenschappelijk bewijs van het hoogste niveau. Verder overweegt de rechtbank in het betreffende vonnis: ‘Als al lagere ‘evidence’ wordt meegenomen, gaat het Zorginstituut bovendien slechts uit van gepubliceerde expert opinions, (...) en de overgelegde verklaringen kwalificeren niet als zodanig.’ Zie: Vonnis Rechtbank Amsterdam van 19 november 2014, gepubliceerd onder nummer ECLI:NL:RBAMS:2014:7727.

eventueel meegenomen wordt in de afweging is dat wat de beroepsgroep als geheel maatgevend vindt respectievelijk als juiste behandelwijze beschouwt. Om die reden richten wij ons voor het overleg en de consultatie tot de wetenschappelijke verenigingen van de zorgaanbieders en tot de patiëntenorganisaties en niet tot individuele zorgaanbieders en individuele patiënten.

In paragraaf 3.3 werken wij de wijze van standpuntbepaling verder uit.

2.5 In richtlijnen beschreven interventies

2.5.1 Vaste werkwijze

Richtlijnen spelen een belangrijke rol in het (para)medisch handelen en zijn ook van groot belang voor het pakketbeheer.⁶⁵ Vaste werkwijze is dat wij een pakketstandpunt over 'de stand van de wetenschap en praktijk', indien mogelijk, laten aansluiten bij de conclusies en bij de aanbeveling(en) over de interventie in de richtlijn(en). Of dat mogelijk is hangt af van een aantal door ons te beoordelen aspecten. Zo zal de richtlijn actueel moeten zijn en volgens de regelen der kunst (met toepassing van de principes van EBM) tot stand moeten zijn gekomen. Verder gaan wij na – indien mogelijk aan de hand van de systematische literatuurreview die verricht is voor het opstellen van de richtlijn – of de gekozen methode van verzameling/selectie/weging van data en het afleiden van conclusies aansluit bij onze werkwijze om 'de stand van wetenschap en praktijk' te bepalen.⁶⁵ Is dat het geval en betreft het een sterke, d.w.z. op krachtig bewijs steunende en eenduidige (positieve/negatieve) aanbeveling betreffende de interventiewaar het ons om gaat, dan zullen wij in beginsel ons pakketstandpunt daarbij laten aansluiten. Richtlijnen kunnen echter ook positieve aanbevelingen voor de toepassing van interventies bevatten die op (zeer) weinig bewijs steunen. Het zal dan in de regel gaan om zwakke – in voorzichtige bewoordingen gestelde – aanbevelingen.⁶⁶ Het is niet zonder meer mogelijk om een pakketstandpunt, dat per definitie maar twee uitkomsten kan hebben (de zorg voldoet wel of voldoet niet aan 'de stand van de wetenschap en praktijk'), te laten aansluiten bij dergelijke zwakke positieve aanbevelingen. In dat geval zal een nadere afweging volgen. Wij verwijzen hiervoor naar de paragrafen 2.4 en 3.3.

Er zijn ook omstandigheden waarin wij terughoudend zijn met het (direct) innemen van een standpunt over 'de stand van de wetenschap en praktijk'. Als bijvoorbeeld bekend is dat de beroepsgroep(en) en andere relevante partijen bezig zijn om een richtlijn op te stellen of aan te passen, waarin ook aanbevelingen zullen komen te staan over de interventie waarover vanuit pakketbeheer vragen zijn gerezen, dan kan het zijn dat wij dat proces afwachten. Daarna gaan wij dan na hoe de richtlijn zich verhoudt tot het verzekerde pakket (d.w.z. of wij kunnen aansluiten bij de in de richtlijn gegeven (onderbouwing van de) aanbeveling). Ook komt het voor dat een literatuurreview die wij hebben (laten) opstellen voor de beoordeling van een interventie in het kader van pakketbeheer, door professionals gebruikt wordt voor het maken van (een update/aanvulling van) een richtlijn. Onze standpunten en de daaraan ten grondslagliggende beoordelingen zijn openbaar en integraal beschikbaar via onze website.

2.5.2 Kwaliteitsinstituut in de zorg: nieuwe taken

Sedert 1 april 2014 heeft Zorginstituut Nederland als opvolger van het CVZ een aantal nieuwe wettelijke taken gericht op het bevorderen van kwaliteit van zorg. Eén van de taken is om een openbaar register bij te houden waarin op voordracht van organisaties van cliënten, zorgaanbieders en zorgverzekeraars gezamenlijk een kwaliteitsstandaard (zoals bijvoorbeeld een richtlijn) wordt opgenomen. Ons instituut heeft een toetsingskader ontwikkeld waarin beschreven staat waar een kwaliteitsstandaard aan moet voldoen om in het register opgenomen te kunnen worden.⁶⁷ Eén van de toetscriteria is dat de kwaliteitsstandaard een adequate beschrijving van de inhoud van de zorg bevat. In de toelichting bij dit onderdeel van het toetsingskader staat dat het onder meer gaat om een toelichting op de bewijsvoering.

65 Naast de kwaliteit is ook de onafhankelijkheid van de richtlijn relevant. Een hulpmiddel om e.e.a. te beoordelen is het zogenoemde AGREE II-Instrument: AGREE Next Steps Consortium. AGREE II-Instrument: Appraisal Instrument for Guidelines, Research and Evaluation. Update 2013. Beschikbaar via www.agreetrust.org. Dit is een internationale standaard voor het beoordelen van de kwaliteit en de onafhankelijkheid van een richtlijn. Zie voor verdere uitleg: Scholten RJPM, Offringa M, Assendelft WJJ van. *Inleiding in evidence-based medicine*. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

66 De formulering kan bijvoorbeeld zijn: te overwegen is of een optie is.

67 CVZ. Beleidsregel Toetsingskader kwaliteitsstandaarden en meetinstrumenten, Staatscourant 2014, nr. 9953.

Het is als volgt beschreven:

“De kwaliteitsstandaard bevat een toelichting op de bewijsvoering. Deze toelichting beschrijft de kracht en beperkingen van de gevonden bewijsvoering voor de zorg die beschreven wordt in de kwaliteitstandaard en benoemt de onzekere punten. Bij voorkeur is de bewijsvoering gebaseerd op wetenschappelijke literatuur die ten grondslag ligt aan de beschrijving van zorg in de kwaliteitsstandaard. De presentatie van de methode van literatuursamenvatting in de kwaliteitsstandaard is transparant: er wordt gerapporteerd hoe de literatuur is gezocht, geïnccludeerd en beoordeeld. Bij het doen van uitspraken en aanbevelingen wordt het niveau van bewijs aangegeven in een graderingssysteem (bijvoorbeeld GRADE). Indien geen wetenschappelijke literatuur over de behandeling voorhanden is, dient te worden toegelicht op welke wijze de aanbevelingen in de kwaliteitsstandaard tot stand zijn gekomen.”

Bij dit onderdeel van het toetsingkader toetsen wij niet inhoudelijk of de bewijsvoering – de (wetenschappelijke) onderbouwing – afdoende is om de conclusie en de aanbevelingen in de zorgstandaard te kunnen dragen. Het toetsingskader voor opname in het register bevat namelijk geen materiële, inhoudelijke normen. De toetsingscriteria zijn louter procedureel van aard. Opname van bijvoorbeeld een richtlijn in het register betekent dan ook niet per definitie dat voldaan is aan de maatstaven die ons instituut aanlegt voor de beoordeling van ‘de stand van de wetenschap en praktijk’ in het kader van pakketbeheer.

2.6 Verzekeringsstatus buitenland

De vraag of een interventie onder de dekking van de Zvw/Wlz valt zal altijd op basis van de Nederlandse regelgeving moeten worden vastgesteld. Omdat ook voor sommige buitenlandse (stelsels van sociale) ziektekostenverzekeringen geldt dat ‘de stand van de wetenschap en praktijk’ een factor is aan de hand waarvan de inhoud en omvang van het pakket (mede) wordt bepaald, kan het van belang zijn om na te gaan welke overwegingen een rol hebben gespeeld bij de pakketbeslissing in het specifieke geval in het betreffende land. Om die reden gaan wij, indien mogelijk, na wat de verzekeringsstatus in het buitenland is van de te beoordelen interventie, welke onderbouwing daaraan ten grondslag ligt en of die relevant is voor onze beoordeling.

Voor de goede orde: het gegeven dat de te beoordelen interventie volgens de regelgeving van het land waar de interventie in kwestie is verleend door de ziektekostenverzekeraars aldaar wordt vergoed, is op zich zelf onvoldoende om tot het standpunt te komen dat de zorg voldoet aan ‘de stand van de wetenschap en praktijk’.

In paragraaf 3.1.6 geven wij aan welke bronnen wij raadplegen om informatie te verkrijgen over de verzekeringsstatus van een interventie in het buitenland

3 Werkwijze en uitgangspunten

In dit hoofdstuk beschrijven wij hoe wij bij de beoordeling van een interventie op effectiviteit te werk gaan. Wij volgen daarbij de vaste EBM-stappen:

- Zoeken en selecteren van evidence en verdere informatie;
- Beoordelen en graderen van de kwaliteit van het bewijs;
- Vaststellen eindbeoordeling/conclusie.

Voor de overzichtelijkheid geven wij in het volgende kader een overzicht van de items die per beoordelingsstap aan bod komen. In de daarop volgende paragrafen lichten wij die items verder toe.

Overzicht van items per beoordelingsstap

Zoeken en selecteren van evidence en verdere informatie:

- Welke interventie bij welke patiëntenpopulatie (en setting) beoordelen wij?
- Waarmee vergelijken wij de te beoordelen interventie?
 - Vergelijking met de standaard- of gebruikelijke behandeling
 - Vergelijking met placebo
 - Directe en indirecte vergelijkingen
- Welke uitkomsten betrekken wij in de vergelijking?
 - Gunstige en ongunstige uitkomsten
 - Uitkomsten nader beschouwd
 - Keuze uitkomstmaten/uitkomsten beargumenteren
- Welke uitkomst van de vergelijking vinden wij voldoende?
 - Meerwaarde en gelijke waarde
 - Klinisch relevant verschil
 - Bepalen klinische relevantie
- Welke follow up-duur willen wij zien?
- Informatiebronnen

Beoordelen en graderen van de kwaliteit van het bewijs:

- Aspecten die de kwaliteit van het bewijs bepalen
 - Methodologische aspecten
 - Belang van de resultaten
 - Generaliseerbaarheid
 - Onderzoeksmethodiek (EBRO-classificatie respectievelijk GRADE)
- EBRO-classificatie
- 'Passend bewijs'
- GRADE-methodiek
 - Uitkomstmaten als vertrekpunt
 - Downgraden
 - Upgraden
 - Beoordelen kwaliteit van het bewijs

Vaststellen eindbeoordeling/conclusie:

- Argumenten voorkomend uit balans tussen voor- en nadelen van de interventie
- Argumenten voortkomend uit kwaliteit van het wetenschappelijk bewijs
- Argumenten voortkomend uit 'passend bewijs benadering'
- Beargumenteerde opvattingen/garanties beroepsgroepen en patiënten
- Onderbouwing standpunt

3.1 Zoeken en selecteren van evidence en verdere informatie

Zoals gezegd, gaat het ons om de relatieve effectiviteit van een interventie: in welke mate draagt de (nieuwe) interventie bij aan het met de interventie beoogde doel in vergelijking met datgene wat in de praktijk al aan medische zorg voor de betreffende aandoening wordt geboden? Een gebruikelijk hulpmiddel om te zoeken naar informatie die relevant is voor een dergelijke beoordeling, is om te werken aan de hand van de zogenoemde PICO-vragen. Deze vragen bewerkstelligen een precieze omschrijving/definiëring van de:

- Patient = de relevante patiëntenpopulatie, waarbij ook de setting van belang kan zijn (bijvoorbeeld: huisartsenpraktijk versus medisch specialistische praktijk);
- Intervention = de te beoordelen interventie;
- Comparison = de interventie waarmee wordt vergeleken (controle-interventie);
- Outcome = de relevante uitkomsten.

In dit rapport gebruiken wij voor de relevante uitkomsten zowel de term uitkomsten als de term uitkomstmaten.

Van belang is ook om informatie te zoeken over:

- Het minimaal vereiste klinisch relevante verschil in uitkomst;
- De minimaal vereiste follow-up periode. Dit aspect - 'time' - wordt ook wel aan de zoekopdracht toegevoegd: PICOT.

De verschillende onderdelen (de PICO(T) en de twee andere punten) bespreken wij in de volgende paragrafen. In de zoek- en selectiefase gaan wij ook na wat voor de specifieke interventie het zogenoemde 'passend bewijs profiel' is. In paragraaf 3.2.3 lichten wij toe wat wij daarmee bedoelen. In de laatste paragraaf van dit hoofdstuk geven wij aan welke bronnen wij gebruiken om informatie te zoeken.

3.1.1 Welke interventie bij welke patiëntenpopulatie (en setting) beoordelen wij?

De definiëring van de interventie (de I van de PICO) en de patiëntenpopulatie (de P van de PICO) is het vertrekpunt voor een beoordeling. Deze wordt vaak in eerste instantie bepaald door de vraag die wij ons zelf stellen, of door de vraag die belanghebbende partijen (zoals bijvoorbeeld VWS, zorgverzekeraars) aan ons voorleggen. Verdere definiëring gebeurt aan de hand van het oriënterend literatuuronderzoek dat altijd voorafgaat aan de beoordeling zelf (het inlezen).

3.1.2 Waarmee vergelijken wij de te beoordelen interventie?

3.1.2.1 Vergelijking met de standaard- of gebruikelijke behandeling

De te toetsen interventie beoordelen wij in principe in vergelijking met de op het moment van toetsing bestaande, meest toegepaste en beste interventie. Deze werkwijze sluit aan bij de internationale praktijk.⁶⁸

De 'meest toegepaste' interventie is:

- de standaardbehandeling. Dit is de interventie die in de behandelpraktijk wordt gezien als behandeling van eerste keuze, de 'gouden standaard';
- de gebruikelijke behandeling ('usual care'), indien geen standaardbehandeling kan worden vastgesteld.

⁶⁸ Zie de EUnetHTA documenten:

- EUnetHTA. JA WP5: Relative effectiveness assessment (REA) of pharmaceuticals, July 2011.

- EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

In het EUnetHTA document over het kiezen van de comparator is een overzicht opgenomen van hoe een groot aantal Europese landen, Australië, Canada, Nieuw Zeeland en de VWS hiermee omgaan. De meeste landen houden als comparator aan 'de op dit moment in de praktijk toegepaste behandeling'.

Actuele (inter)nationale richtlijnen en standaarden kunnen als bron dienen om vast te stellen welke interventie als comparator gekozen moet worden. Soms ontbreken die echter of het kan zijn dat zij geen uitsluitsel geven. Het is bijvoorbeeld lang niet altijd zo dat er maar één standaardbehandeling is voor een aandoening. Soms zijn meerdere interventies mogelijk, die onderling goed vergelijkbaar zijn qua uitkomsten.⁶⁹ Wat de 'usual care' betreft, deze kan van land tot land of zelfs - binnen één land - van regio tot regio verschillen. Verder kan 'usual care' er ook uit bestaan dat geen behandeling wordt geboden, maar dat geadviseerd wordt om (voorlopig) af te wachten of rust te houden. De 'usual care' kan ook bestaan uit verlichting van klachten, zoals bijvoorbeeld aan de orde is bij palliatieve zorg. Ook is het mogelijk dat tot dan toe geen alternatief beschikbaar is. In deze gevallen moet de te beoordelen (nieuwe) interventie vergeleken worden met 'niets doen' of 'best ondersteunende zorg'.

De 'beste' comparator is bij voorkeur de interventie waarvan de effectiviteit (in voldoende mate) vaststaat. Soms kan het echter schorten aan bewijs voor de effectiviteit van de 'meest toegepaste' interventie. Bewijs van goede kwaliteit voor effectiviteit ontbreekt immers geregeld.

Het kiezen van een goede comparator is soms dus lastig, maar er zal toch altijd een keuze gemaakt moeten worden. Het ligt in de rede in dat geval een gemotiveerde, pragmatische benadering te volgen.⁷⁰ Het is van belang om de beargumenteerde keuze af te stemmen met de relevante wetenschappelijke verenigingen van beroepsgroepen en patiëntenorganisaties.

Basis voor de beoordeling is de vergelijking met de standaard- of gebruikelijke behandeling. Het gaat er immers, zoals eerder gezegd, om wat de (nieuwe) interventie 'netto' toevoegt aan datgene wat in de praktijk al aan medische zorg voor de betreffende aandoening wordt geboden. Vooral bij studies over medicijnen vindt echter geregeld een andere vergelijking plaats, namelijk de vergelijking met het toedienen van een placebo (placebo-gecontroleerd onderzoek). Een dergelijk onderzoek kan bij de beoordeling een toegevoegde waarde hebben.

3.1.2.2 Vergelijking met placebo

Bij placebo-gecontroleerd medicijnonderzoek wordt de pil met de werkzame stof vergeleken met een pil zonder werkzame stof. Met dit onderzoek kan de werkzaamheid van de actieve stof sec worden vastgesteld en worden onderscheiden van een eventueel effect van het gebruiken van een pil op zich. Bijvoorbeeld: een placebo kan ook een positief effect hebben op een depressie en qua effect op één lijn te stellen zijn met een antidepressivum. De aandacht, het idee er zelf iets aan te kunnen doen door pillen te slikken, zou op zich – zonder dat er een werkzame stof nodig is – ook al kunnen helpen. Soms wordt in de chirurgie een 'sham' operatie gedaan (bijvoorbeeld een knie-arthroscopie): de patiënt krijgt wel een verdoving en een incisie, maar er vindt geen daadwerkelijke, actieve ingreep plaats.

⁶⁹ Een voorbeeld hiervan is de lumbale HNP (hernia). Deze aandoening kan veel pijn in het been geven en kan soms zelfs invaliderend zijn. In het merendeel van de gevallen gaan deze klachten vanzelf weer over na enkele maanden, zonder specifieke behandeling. Met een operatie gaan de klachten ook over. Afhankelijk van de voorkeur van de patiënt kan gekozen worden voor directe operatie of enkele maanden afwachten. Iemand die een drukke baan of een druk gezin heeft, kan ervoor kiezen om toch meteen een operatie te ondergaan, daarbij de risico's van een operatie op de koop toe nemend. De richtlijn HNP beveelt aan om afwachting te zijn met operatie, tenzij er sprake is van ernstige/onhoudbare pijn of sensorische uitval. Dit is echter geen wet van meden en perzen. En wat is onhoudbaar: iemand met kleine kinderen thuis zal eerder over onhoudbaar spreken dan iemand zonder veel verplichtingen.

⁷⁰ In de literatuur wordt het probleem dat bewijs van goede kwaliteit voor effectiviteit soms ontbreekt ook onderkend. Er wordt geadviseerd in die gevallen op pragmatische wijze te handelen. Zie EUnetHTA document: EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

De plaats van placebo-gecontroleerd onderzoek is bepaald in de Verklaring van Helsinki en gaat uit van: ‘geen placebo-gecontroleerd onderzoek, tenzij’.⁷¹ In de volgende situaties kan het gebruik van placebo-gecontroleerd onderzoek worden overwogen:⁷²

- als er geen effectieve interventie bestaat;
- als in de studie een add-on interventie wordt onderzocht (nieuwe interventie toegevoegd aan standaardzorg versus placebo toegevoegd aan standaardzorg);
- als er geen, of minimale, nadelige gevolgen zijn van het onthouden/uitstellen van de gebruikelijke behandeling (denk bijvoorbeeld aan een interventie tegen verkoudheid);
- als er zwaarwegende redenen zijn om een placebo-controle te gebruiken, én er geen grote kans op schade bestaat bij deelnemers aan de studie (bij hoge kans op placebo-effect, bijvoorbeeld pijnstillers, acupunctuur, of bij een chronische aandoening met een sterk fluctuerend verloop).

Het kan in sommige gevallen toegestaan en ook zinvol zijn om (aanvullend) placebo-gecontroleerd onderzoek te doen, bijvoorbeeld als de effectiviteit van de bestaande behandeling (waarmee eigenlijk vergeleken moet worden) onvoldoende vaststaat. Het zou dan kunnen zijn dat het effect van de interventie in werkelijkheid een placebo-effect is. Een voorbeeld hiervan is de ‘sham’ arthroscopie bij knie-arthrose of degeneratief meniscusletsel (zie het kader).

Arthroscopie versus sham: geen verschil in effect

Moseley et al⁷³ onderzochten het effect van aan arthroscopie met débridement of lavage ten opzichte van een sham-procedure bij patiënten met slijtage van de knie. De sham procedure bestond uit het maken van uitsluitend de huid-incisies. De arthroscoop werd niet in het kniegewricht geïntroduceerd. De resultaten in de groepen (na de echte scopie en na de sham-procedure) waren niet verschillend van elkaar over een totale follow-up periode van twee jaar. De uitkomstmaten waren zowel subjectief (door de patiënt gerapporteerde pijn- en functiescores) als objectief (tijdsduur van een bepaalde loopafstand en van traplopen). Deze studie liet zien dat een algemeen geaccepteerde en op grote schaal uitgevoerde behandeling geen voordelen biedt boven een placebo-procedure. Recent nog werd een dergelijke placebo-gecontroleerde studie ook uitgevoerd om te kijken naar de effectiviteit van een arthroskopische partiële meniscectomie. Ook in deze studie werd geen verschil gezien in effectiviteit na de werkelijke ingreep en na de sham-procedure.⁷⁴

Ook uit een ander, recent voorbeeld blijkt de toegevoegde waarde van een placebo-gecontroleerde studie (zie het kader).

Rofecoxib met placebo: rofecoxib geeft verhoogde kans op infarct

Het gaat hier om rofecoxib (Vioxx), een pijnstiller, o.a. toegepast bij gewrichtsklachten. Vergeleken met naproxen leidde rofecoxib tot meer kans op hartinfarcten. Het was echter niet duidelijk of dit verschil gevolg was van een beschermend effect van naproxen, of van een werkelijk verhoogde kans door rofecoxib. Pas uit een latere, placebogecontroleerde studie bleek dat het werkelijk ging om een verhoogde kans op infarct door de rofecoxib. Eerst na afronding van deze studie werd rofecoxib uit de handel genomen.⁷⁵

Het komt erop neer dat als uitgangspunt geldt dat wij altijd de effectiviteit van een (nieuwe) interventie vergelijken met de bestaande, meest toegepaste en beste interventie (relative/ comparative effectiveness). De keus van de comparator dient goed beargumenteerd te zijn. In sommige gevallen kan placebo-gecontroleerd onderzoek extra (nuttige) informatie opleveren. Dit zal van geval tot geval

⁷¹ World Medical Association. Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. 2013. Beschikbaar via www.wma.net.

⁷² Millum J, Grady C. The ethics of placebo-controlled trials: methodological justifications. *Contemp Clin Trials* 2013; 36:510-4.

⁷³ Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81-8.

⁷⁴ Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med* 2013;369:2515-24.

⁷⁵ Voorbeeld ontleend aan Stang A, Hense HW, Jöckel KH, et al Is it always unethical to use a placebo in a clinical trial? *PLoS Med* 2005;2:177-80.

beoordeeld moeten worden. Als er placebo-gecontroleerd onderzoek beschikbaar is, betrekken wij dit bij onze beoordeling.

3.1.2.3 Directe en indirecte vergelijkingen

Idealiter zou je de te beoordelen interventie en de comparator voor één aandoening direct met elkaar willen vergelijken in een gerandomiseerde gecontroleerde studie (RCT). Op deze manier kan het meest betrouwbare bewijs worden geleverd voor de waarde van een bepaalde interventie ten opzichte van de controle-interventie.⁷⁶ Deze ideale situatie doet zich lang niet altijd voor. De vraag is of er ook gebruik gemaakt kan worden van zogenoemde indirecte vergelijkingen. Dat is de situatie waarin resultaten van studie-armen uit verschillende onderzoeken met elkaar worden vergeleken. Bijvoorbeeld: interventie A is in RCT X vergeleken met interventie B. Interventie C is in RCT Y vergeleken met interventie B. Door middel van een indirecte vergelijking kan men ook interventie A met interventie C vergelijken.

Directe vergelijking in RCT X: A versus B

Directe vergelijking in RCT Y: C versus B

Indirecte vergelijking: A versus C.

Met behulp van indirecte vergelijkingen kan het tekort aan directe vergelijkingen worden gecompenseerd, of kan het bewijs afkomstig uit directe vergelijkingen worden aangevuld/versterkt. Echter, dit kan niet zomaar: als studies niet goed vergelijkbaar zijn is het risico van vertekening en daarmee het risico dat de conclusies minder geloofwaardig zijn, zeker aanwezig. De vergelijkbaarheid betreft verschillende aspecten, zoals klinische aspecten (waaronder patiëntkenmerken, interventie, studiecontext, follow-up, uitkomstmaten) en methodologische aspecten (waaronder studie-opzet, data-analyse).

Ter toelichting hiervan geven wij enkele voorbeelden (zie het kader).

Indirecte vergelijking niet zonder meer mogelijk

Als in RCT X middel A wordt onderzocht in patiënten met hypertensie, bestaande uit 20- tot 50-jarigen zonder co-morbiditeit, en in RCT Y middel B in patiënten met dezelfde mate van hypertensie, maar bestaande uit 60-plussers met overgewicht, dan zijn de groepen niet vergelijkbaar en is een indirecte vergelijking (A versus B) niet betrouwbaar. Als de uitkomstmaat in RCT X is gemeten d.m.v. een éénmalige meting van de bloeddruk en in RCT Y d.m.v. een 24-uurs bloeddrukmeting, is deze evenmin vergelijkbaar en is een indirecte vergelijking dus niet betrouwbaar. Ook de dosering van een interventie kan verschillen, of de toepassing, zoals de mate van ervaring van een chirurg. Als de studieresultaten op verschillende manieren zijn geanalyseerd kan evenmin zonder meer indirect vergeleken worden.

Voor het uitvoeren van indirecte vergelijkingen is een aantal methodieken gangbaar. In grote lijnen gaat het om het volgende:

- De meest simpele manier om een indirecte vergelijking uit te voeren is door het vergelijken van de resultaten uit twee verschillende studies, uit twee verschillende meta-analyses of van studie-armen uit verschillende onderzoeken met elkaar. Als niet duidelijk is waarin de groepen, de behandeling en de analyse van elkaar verschillen, zal een dergelijke ruwe indirecte vergelijking tot veel bias leiden en niet goed bruikbaar zijn.
- Een verbetering is een indirecte vergelijking van studie-armen waarbij de controle-arm hetzelfde is. Voorbeeld: A versus C en B versus C gebruiken we om indirect A met B te vergelijken. Dit kan als gecheckt is of de studies vergelijkbaar zijn qua klinische en qua methodologische aspecten (zie boven).
- Een geavanceerde methodiek is om directe en indirecte vergelijkingen met elkaar te combineren in één analyse (mixed treatment comparisons, MTC; netwerk meta-analyse). Uitgangspunt is dat alle

⁷⁶ De controle-interventie hoeft in een directe vergelijking overigens geen actieve interventie te zijn, het kan ook gaan om een placebo-behandeling of om geen behandeling. Bij de beoordeling van geneesmiddelen wordt overigens de term 'head-to-head comparison' gebruikt voor vergelijking met de actieve interventie.

beschikbare behandelingen voor één indicatie in één grote analyse worden gestopt (een netwerk meta-analyse) en dat dan al deze behandelingen met elkaar worden vergeleken: direct en/of indirect. Daaruit volgt dan een rangorde van de verschillende behandelingen qua effectiviteit. De kwaliteit van dergelijke analyses, en daarmee de betrouwbaarheid, neemt toe naarmate de vergelijkbaarheid van de studies toeneemt.

EUnetHTA⁷⁷ heeft onlangs een guideline gepubliceerd over directe en indirecte vergelijkingen waarin de verschillende technieken kritisch worden besproken aan de hand van een aantal reviews. In de guideline staat een aantal aanbevelingen betreffende de bruikbaarheid van indirecte vergelijkingen.⁷⁸ Dat geeft een indruk van waar op gelet moet worden bij het beoordelen van indirecte vergelijkingen. Daarnaast is onlangs een mogelijke aanpak voorgesteld over hoe bij de toepassing van GRADE de kwaliteit van het bewijs uit netwerk meta-analyses gewaardeerd zou kunnen worden.⁷⁹

Wat betekent een en ander voor onze toetsing van interventies aan 'de stand van de wetenschap en praktijk'? Wij kunnen voor onze beoordelingen gebruik maken van gepubliceerde indirecte vergelijkingen, mits de auteurs de kans op bias en de betrouwbaarheid van de resultaten goed hebben ingeschat. De genoemde aanbevelingen in de EUnetHTA-guideline en wellicht ook de GRADE-aanpak kunnen behulpzaam zijn om ons over dit laatste een oordeel te vormen.

3.1.3 Welke uitkomsten betrekken wij in de vergelijking?

Bij de beoordeling van de effectiviteit gaat het ons om de mate waarin de (nieuwe) interventie bijdraagt (de 'netto toevoeging') in vergelijking met de standaard of gebruikelijke behandeling voor de betreffende aandoening. Om dit te bepalen kijken wij naar de gunstige en naar de ongunstige gevolgen van de te vergelijken interventies voor de patiënt. Het verschil in gunstige en ongunstige gevolgen wegen wij ten opzichte van elkaar.

3.1.3.1 Gunstige en ongunstige uitkomsten

Gunstige effecten zijn de positieve (beoogde) gevolgen van een interventie. Deze worden bij voorkeur uitgedrukt in klinisch relevante uitkomsten. Dat zijn uitkomsten die de mate van morbiditeit, mortaliteit of kwaliteit van leven van een behandeling weerspiegelen. Geregeld komt het voor dat klinische uitkomsten als morbiditeit en mortaliteit zich slechts zelden zullen gaan voordoen in het verloop van de follow-up. Studies met dit soort uitkomstmaten zijn daarom vanwege het grote aantal te includeren patiënten en de benodigde lange follow up niet altijd haalbaar. In plaats daarvan worden in klinische onderzoeken resultaten weergegeven op intermediaire uitkomstmaten (ook wel genoemd: surrogaat-uitkomstmaten). Het kan onder bepaalde voorwaarden gerechtvaardigd zijn om bij een beoordeling af te gaan op intermediaire uitkomstmaten, zoals bijvoorbeeld een laboratoriumbepaling of een lichamelijk kenmerk. Hierna gaan wij daar verder op in.

Een ongunstige uitkomst is een gevolg dat niet beoogd is, maar dat wel kan optreden bij de behandeling van een patiënt. De meeste ongunstige uitkomsten zijn bijwerkingen, maar bijvoorbeeld ook effecten als het optreden van resistentievorming van bacteriën door toepassing van antibiotica worden beschouwd als ongunstige effecten. De nadruk ligt uiteraard met name op de ernstige en de vaak voorkomende bijwerkingen van een behandeling. Onder een ernstige bijwerking wordt verstaan een bijwerking die dodelijk is, levensgevaar oplevert, invaliditeit of arbeidsongeschiktheid veroorzaakt, of tot opname in een ziekenhuis of verlenging daarvan, leidt. Altijd zal bij de beoordeling of een interventie voldoet aan 'de stand van de wetenschap en praktijk' informatie over bijwerkingen en complicaties (onder de noemer veiligheid) worden betrokken.

⁷⁷ EUnetHTA staat voor European Network for Health Technology Assessment. Het is het Europese netwerk waar alle Europese HTA-organisaties aan deelnemen.

⁷⁸ EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

⁷⁹ Puhan MA, Schünemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630.

Bepaalde kenmerken of aspecten van een behandeling kunnen veroorzaken dat patiënten de voorgestelde behandeling niet (trouw) opvolgen. Gebrek aan ‘uptake’ en ‘compliance’ ofwel therapietrouw kan het verloop en het uiteindelijke effect van de behandeling ernstig negatief beïnvloeden. Er kan dan reden zijn om dat in de afweging mee te nemen. Dit soort aspecten kunnen door het meten van patiëntpreferenties zichtbaar worden. Hoe op wetenschappelijk verantwoorde wijze onderzoek naar patiëntpreferenties plaats kan vinden (dat ook bruikbaar is voor pakketbeheer), is echter nog onvoldoende uitgekristalliseerd.

3.1.3.2 Uitkomstmaten nader beschouwd

Het bepalen van de waarde van een interventie gebeurt aan de hand van meetbare uitkomsten. Voorbeelden van te meten uitkomsten zijn mortaliteit, kwaliteit van leven, functie van een gewricht, patiënttevredenheid, bloeddruk aan de bovenarm gemeten, cholesterolgehalte in het bloed, tumorgrootte op een CT-scan. Uit deze serie voorbeelden blijkt dat de verscheidenheid in uitkomstmaten groot is, en dat ze variëren in robuustheid.⁸⁰ Verder is het zo dat er diverse termen gebruikt worden die soms hetzelfde betekenen. Denk aan de termen ‘harde/zachte’ uitkomstmaten, surrogaat-uitkomstmaten, intermediaire uitkomstmaten, door de patiënt gerapporteerde uitkomstmaten, samengestelde uitkomstmaten. Hierna brengen wij daar enige ordening in, daarbij gebruik makend van recent gepubliceerde EUnetHTA-documenten over dit onderwerp.⁸¹ Uiteraard speelt de keuze voor uitkomstmaten ook een rol in de GRADE-beoordeling (zie verderop). Wij besteden hier op deze plaats extra aandacht aan vanwege het belang van het gebruik van de juiste uitkomsten.

- Klinische uitkomstmaten

Klinische uitkomsten zijn – zoals eerder gezegd – uitkomsten die de mate van morbiditeit, mortaliteit of kwaliteit van leven van een behandeling weerspiegelen. Deze uitkomsten geven het beste aan hoe het staat met de gezondheidstoestand van een patiënt: hoe de patiënt zich voelt, hoe hij functioneert, of hij in leven blijft. Voorbeelden van klinische uitkomsten zijn de SF-36 vragenlijst (gevalideerde kwaliteit van leven score), de Oswestry disability index (een gevalideerde score-maat voor pijn en functie bij patiënten met chronische lage rugpijn), het wel of niet optreden van een hartinfarct⁸² (waarvan de definitie moet zijn gespecificeerd) en mortaliteit.

- PROM's

PROM's zijn door patiënten gerapporteerde uitkomsten, gebruikt als uitkomstmaat. Daaronder vallen uitkomstmaten als pijnscores, bijvoorbeeld op een ‘visual analogue score’ (VAS), en kwaliteit van leven. Dergelijke uitkomstmaten worden ook wel als ‘subjectief’ aangeduid. Zij worden door de patiënt in kaart gebracht zonder tussenkomst van de zorgverlener. Zij vallen onder de klinische uitkomstmaten, omdat zij immers veranderingen in de gezondheidstoestand van de patiënt weergeven.

Kwaliteit van leven-vragenlijsten kunnen generiek zijn of ziektespecifiek.⁸³ Het voordeel van een generieke vragenlijst, zoals de SF-36 en de EQ-5D, is dat vergelijking tussen verschillende aandoeningen mogelijk is. In alle gevallen dient een vragenlijst gevalideerd te zijn.⁸⁴

- Intermediaire of surrogaat-uitkomstmaten

Met name als het gaat om ‘harde’ uitkomstmaten, zoals het optreden van hartinfarct of mortaliteit, is het vaak moeilijk en kostbaar om een studie op te zetten die robuust genoeg is om verschillen te

⁸⁰ De mate van robuustheid van een uitkomstmaat bepaalt mede de kwaliteit van de evidence.

⁸¹ EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

⁸² Ook bij een klinische uitkomst, zoals bijvoorbeeld een hartinfarct, moet in het studieprotocol worden gedefinieerd wanneer er sprake is van het event. Een hartinfarct kan op verschillende manieren gedefinieerd worden, en een onnauwkeurige beschrijving introduceert onzekerheid in de geloofwaardigheid van de data en de conclusies van de studie.

⁸³ EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

⁸⁴ Interne validiteit, reproduceerbaarheid, gevoeligheid.

kunnen aantonen.⁸⁵ In onderzoeken wordt dan vaak gebruik gemaakt van intermediaire of surrogaat-uitkomstmaten. Van dergelijke uitkomstmaten moet wel vaststaan dat zij verband houden met veranderingen in de klinische uitkomst.⁸⁶ Bijvoorbeeld: het gebruik van bloeddruk of cholesterol als intermediaire uitkomstmaat heeft alleen zin als vaststaat dat een bepaalde mate van bloeddrukdaling of cholesteroldaling leidt tot een verandering in het optreden van cardiovasculaire events (zoals hartinfarct, CVA). Zo'n verband is niet altijd even sterk. Gebruik maken van intermediaire uitkomstmaten heeft dus vaak een afname van de robuustheid van de evidence tot gevolg.

Van een intermediaire of surrogaat-uitkomstmaat mag worden verwacht dat de klinische (meer) waarde van een interventie er goed mee kan worden ingeschat, gebaseerd op pathofysiologische, epidemiologische en klinisch-therapeutische gegevens. Idealiter moet van een intermediaire of surrogaat-uitkomstmaat zijn vastgesteld dat hij op betrouwbare wijze samenhangt met het klinische effect van een interventie. Daarbij moet in ieder geval aandacht zijn voor de analytische accuratesse van de test waarmee de intermediaire of surrogaat-uitkomstmaat wordt gemeten. Denk hierbij bijvoorbeeld aan een laboratoriumbepaling. Plausibiliteit (beschikbaarheid van aanwijzingen voor een causaal verband tussen intermediair/surrogaat- en harde uitkomstmaat) en lineair verband tussen intermediair/surrogaat- en harde uitkomstmaat zijn belangrijke kenmerken, die het vertrouwen in de intermediaire of surrogaat-uitkomstmaat vergroten.

Met resultaten van studies die gebruik maken van uitkomstmaten die niet zijn gevalideerd, of waarvan het verband met de klinische uitkomst onvoldoende is, kan niet zomaar geëxtrapoleerd worden naar de klinische uitkomst. Dit leidt tot meer (grote) onzekerheid over de geloofwaardigheid van de uitkomst.

Ter illustratie volgt een casus (zie het kader).

Ongeschikte surrogaat-uitkomstmaat

Osteoporose brengt een verhoogd risico op fractures met zich mee. Vooral bij ouderen komen fractures (van een ruggenwervel, van de heup of de pols) vaak voor en hebben ze een aanzienlijke ziektelast tot gevolg. Over de voorspellende waarde van de botmineraaldichtheid ('DEXA'), gemeten in wervels en heupkop, voor het optreden van fractures is in het verleden veel discussie geweest. Deze is relatief slecht. Dit hangt samen met de grote overlap in de mate van botdichtheid van personen met en zonder fractures, en omdat het risico van fractures multifactorieel is bepaald. De botdichtheidsmeting kan wel, in combinatie met andere risicofactoren voor osteoporotische fractuur, worden gebruikt om vast te stellen of er een indicatie bestaat voor medicamenteuze behandeling. De botmineraaldichtheid is in het verleden gebruikt als surrogaat-uitkomstmaat voor het optreden van fractures. Veranderingen in de botmineraaldichtheid bleken echter niet te correleren met veranderingen in het fractuurrisico. De DEXA is dus ongeschikt als surrogaat-uitkomstmaat voor de medicamenteuze behandeling van osteoporose.⁸⁷

- Samengestelde uitkomstmaten

Soms wordt een samengestelde uitkomstmaat gebruikt om het effect van een interventie te schatten. Samengestelde uitkomstmaten kunnen bijvoorbeeld gebruikt worden als er sprake is van twee uitkomsten die duidelijk klinisch met elkaar verbonden zijn. Als een bepaalde uitkomst zelden voorkomt (bijvoorbeeld 'mortaliteit'), kan deze worden gecombineerd met een andere uitkomst (bijvoorbeeld 'infarct'), en kunnen toch in een zo opgezette studie zonder dat de studie al te groot, kostbaar of langdurig (en daarmee onuitvoerbaar) wordt, relevante verschillen binnen een redelijke termijn worden aangetoond. Ook hier geldt dat de keuze voor de samenstelling van de uitkomstmaten goed

⁸⁵ Zo'n studie vereist vaak grote patiëntenaantallen en lange follow-up.

⁸⁶ EUnetha. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013. Beschikbaar via www.eunetha.eu/eunetha-guidelines.

⁸⁷ Ongleend aan: NHG-Standaard Osteoporose. Huisarts Wet 2005;48:559-70 en EMA. Guideline on the evaluation of medicinal products in the treatment of primary osteoporosis. London, 2006.

onderbouwd moet zijn. Bovendien moet – uiteraard – van iedere afzonderlijke uitkomstmaat duidelijk zijn wat de validiteit is. De onderdelen van de samengestelde uitkomstmaat moeten redelijk homogeen zijn wat aard en relevantie betreft. Als zij erg verschillend zijn kan het lastig zijn om de klinische relevantie van een gevonden verschil in effect te interpreteren.

Bijvoorbeeld: een samengestelde uitkomstmaat bestaat uit het optreden van een hartinfarct en het door de patiënt registreren van pijnklachten. Een statistisch significant verschil in effect kan teruggevoerd worden op het optreden van hartinfarcten, op de pijnscore of op beide. Het gevolg kan dan zijn dat het belang van de interventie wordt overschat, omdat een groot verschil in effect op een minder belangrijk onderdeel van de uitkomstmaat domineert.

3.1.3.3 Keuze uitkomstmaten/uitkomsten beargumenteren

Per te beoordelen interventie bepalen wij beargumenteerd welke uitkomsten klinisch relevant zijn om te kunnen beoordelen of de interventie als effectief kan worden beschouwd. De keuze voor een bepaalde uitkomst hangt af van de te onderzoeken patiëntenpopulatie, van de aandoening en van het doel van de behandeling. Intermediaire of surrogaat-uitkomstmaten betrekken wij bij voorkeur alleen in de beoordeling als deze in de artikelen/literatuur gevalideerd en onderbouwd zijn. Verder is van belang, zoals eerder opgemerkt, dat het ons primair gaat om de klinische effectiviteit van de interventie: is de interventie in de praktijk ook echt belangrijk voor de patiënt. Anders gezegd, heeft de interventie voor de patiënt een duidelijke (meer)waarde, in de zin dat zijn gezondheid en overleving verbetert. Het feit dat het om 'effectiveness' gaat en niet alleen om 'efficacy' (werkzaamheid) bepaalt ook de keuze van de uitkomstmaten.⁸⁸

Welke uitkomstmaten/uitkomsten relevant zijn in een concrete beoordeling stemmen wij af met de relevant wetenschappelijke verenigingen van professionals en patiëntenorganisaties.

3.1.4 Welke uitkomst van de vergelijking vinden wij voldoende?

Wij vergelijken aan de hand van de relevante uitkomstmaten/uitkomsten de te toetsen interventie met de standaardbehandeling of de gebruikelijke behandeling. Wanneer kunnen wij concluderen dat de interventie in kwestie als effectief is te beschouwen en dus is aan te merken als zorg is conform 'de stand van de wetenschap en praktijk'? Twee aspecten komen hierna aan bod: 1) is meerwaarde vereist of is gelijke waarde voldoende? en 2) is het gevonden verschil klinisch relevant?

3.1.4.1 Meerwaarde en gelijke waarde

Om de voorkomen dat het verzekerde pakket een verslechtering ondergaat is het uitgangspunt dat de uitkomst van de beoordeling positief moet zijn (d.w.z. een gelijke of grotere effectiviteit). Een interventie die qua effectiviteit inferieur is ten opzichte van de standaardbehandeling of gebruikelijke behandeling (dus in vergelijking minder waarde heeft), kan in beginsel niet worden beschouwd als zorg conform 'de stand van de wetenschap en praktijk'.

Verder kan als leidraad worden aangehouden:

- Komt er een nieuwe interventie beschikbaar die ingezet kan worden als alternatief voor/in plaats van een bepaalde standaardbehandeling of een bepaalde gebruikelijke behandeling, dan voldoet die interventie aan 'de stand van de wetenschap en praktijk' als zij qua effectiviteit in ieder geval gelijk is aan de standaardbehandeling/gebruikelijke behandeling. Gelijke waarde is in dit geval voldoende.
- Is het zo dat een nieuwe interventie bestaat uit de standaardbehandeling of de gebruikelijke behandeling én een toevoeging daarop, dan moet de nieuwe interventie qua effectiviteit een meerwaarde hebben ten opzichte van de standaardbehandeling of gebruikelijke behandeling wil er sprake zijn van zorg conform 'de stand van de wetenschap en praktijk'. Anders gezegd: de toevoeging moet daadwerkelijk relevante meerwaarde teweeg brengen in vergelijking met de standaardbehandeling/gebruikelijke behandeling. Een voorbeeld van een situatie waarbij sprake is van een aan een standaardbehandeling toegevoegd onderdeel is: hyperthermie in combinatie met radiotherapie versus radiotherapie sec.

⁸⁸ Het verschil tussen 'efficacy' en 'effectiveness' komt in paragraaf 2.2 aan bod.

Wij moeten wel onderkennen dat gelijke waarde en meerwaarde niet altijd makkelijk te scheiden zijn. Vaak zijn studies waarin de effectiviteit van een behandeling wordt onderzocht niet ontworpen om gelijkwaardigheid aan te tonen, maar gericht op het onderzoeken van meerwaarde.⁸⁹ Het hangt dan van de keuze van de uitkomstmaat af of je spreekt van gelijke waarde of van meerwaarde (welke uitkomstmaat is cruciaal of belangrijk, en gebruiken wij om onze conclusie op te baseren).

3.1.4.2 Klinisch relevant verschil

Als een gevonden verschil in effect tussen twee interventies statistisch gezien significant is, betekent dit niet dat het verschil ook klinisch relevant is. Waarom dit onderscheid moet worden gemaakt en hoe wij dit in onze beoordelingen benaderen, lichten wij hierna toe.⁹⁰

Met statistische significantie van een gevonden verschil in effect wordt bedoeld dat de kans dat het verschil 'toevallig' is en niet toegeschreven kan worden aan de interventie, verwaarloosbaar is. Meestal wordt ' $p < 0.05$ ' aangehouden als grens, waarmee bedoeld wordt dat de kans dat het verschil op toeval berust, kleiner dan 5% is. Daarmee aangevend dat wij er met 95% zekerheid vanuit kunnen gaan dat het gevonden verschil, bijvoorbeeld als resultaat van een goed uitgevoerde RCT, op de onderzochte interventie berust.

Belangrijk is echter ook de vraag of een gevonden statistisch significant verschil ook een klinisch relevant verschil is. Een verschil kan heel significant zijn maar niet klinisch relevant. Alleen bij klinisch relevante verschillen tussen twee interventies is het aangewezen om een wijziging in het behandelbeleid aan te brengen, omdat slechts dan een voor de patiënt wezenlijke verbetering is te verwachten. Hiervoor is het nodig om te bepalen wat de minimale klinische relevantie van een effectverschil is. Alleen verschillen die gelijk aan of groter dan deze MCID (minimal clinical important difference) zijn, zouden consequenties moeten hebben voor de praktijk. Vaak ontbreekt in wetenschappelijke studies een (gevalideerde) schatting van de MCID of wordt het onderscheid tussen significantie en klinische relevantie niet expliciet gemaakt. Daardoor wordt het belang van een statistisch significant verschil nogal eens overschat. Het volgende voorbeeld illustreert dit (zie kader).

Vershil in effect wel statistisch significant, maar niet klinisch relevant

Van Tulder et al⁹¹ gingen na of statistisch significante resultaten van trials op het gebied van lage rugpijn klinisch relevant waren. Uit een systematisch literatuuronderzoek naar de effectiviteit van oefentherapie bij lage rugklachten bleek dat van de 43 onderzochte studies er 18 positieve (statistisch significante) resultaten op de uitkomst pijn rapporteerden, maar dat het bij slechts vier daarvan ook om klinisch relevante resultaten ging. Klinische relevantie was gedefinieerd als een verschil van minstens 20% op de pijnscore en van minstens 10% op de functiescore. De auteurs concluderen dat de conclusies uit de studies die het effect van oefentherapie op pijn onderzochten voornamelijk op statistische significantie waren gebaseerd en veel minder op klinische relevantie en dus te positief waren over de resultaten. Zij pleiten voor meer aandacht voor de klinische relevantie van effectverschillen, zowel bij het ontwerpen van een studie als bij de interpretatie van de resultaten.

⁸⁹ Dit gaat met name op voor studies waarin de effectiviteit van medisch specialistische zorg wordt onderzocht. Voor het opzetten van dergelijke studies wordt uitgegaan van de nulhypothese dat beide behandelingen identiek zullen zijn. Vervolgens wordt berekend wat nodig is om die hypothese te verwerpen. Als het effectverschil dat men met behulp van de studie zichtbaar wil laten worden, uiteindelijk niet wordt aangetoond ('een niet significant verschil'), kan hooguit worden geconcludeerd dat – uitgaande van de studieresultaten – van meerwaarde niet is gebleken. Dat geen meerwaarde is aangetoond, kan ook het gevolg zijn van het feit dat de vereiste groepsgrootte niet is gehaald, dat er veel uitval was of dat er sprake was van andere methodologische tekortkomingen.

⁹⁰ In de medisch wetenschappelijke literatuur wordt in toenemende mate aandacht gevraagd voor de relevante van het onderscheid tussen statistische significantie en klinische relevantie. Zie bijvoorbeeld: Tulder van, M, Malmivaara A, Hayden J, Koes B. Statistical significance versus clinical importance. *Spine* 2007;32:1785-90.

⁹¹ Tulder van, M, Malmivaara A, Hayden J, Koes B. Statistical significance versus clinical importance. *Spine* 2007;32:1785-90.

Voor een aantal uitkomstmaten bij grote ziektebeelden zijn MCIDs vastgesteld, bijvoorbeeld voor pijnscores, veranderingen in kwaliteit van levenscore, aantal mmol/l cholesterolverschil in het plasma, functiescores bij rugklachten.

Op grond van een vastgestelde MCID kan bij het plannen van een studie de vereiste studiegrootte worden berekend. Als de studie vervolgens goed wordt uitgevoerd moet een duidelijke conclusie (namelijk: gevonden verschil is wel of niet klinisch relevant) over de klinische relevantie van de interventie kunnen volgen. Als bij het opzetten van de studie geen acht wordt geslagen op de MCID, kan het zich voordoen dat een effectverschil weliswaar statistisch significant is, maar dat uiteindelijk over de klinische relevantie – of het er echt toe doet – geen uitspraak kan worden gedaan. De studie is, voor zover het consequenties voor het behandelbeleid betreft, dan zinloos geweest en – als experimentele zorg is onderzocht – laakbaar, omdat patiënten zijn blootgesteld aan experimentele zorg in een onderzoek dat niet tot degelijke conclusies kon leiden.

De MCID van een effectverschil in een bepaalde uitkomstmaat dient door de onderzoekers per aandoening te worden vastgesteld. Een voorbeeld ter illustratie (zie het kader).

Validering klinische relevantie effectverschil

De 6-minuten looptest meet de afstand die afgelegd is in zes minuten tijd. De test wordt gestandaardiseerd uitgevoerd en is oorspronkelijk ontworpen als een geïntegreerde maat voor de hart- long- en spierfunctie van patiënten met een longziekte, maar wordt ook bij andere aandoeningen gebruikt. De uitkomst en veranderingen daarin kunnen uiteraard afhankelijk zijn van een interventie, maar ook van de onderliggende aandoening, van aanwezige comorbiditeit of van andere aspecten zoals bijvoorbeeld de leeftijd van de patiëntenpopulatie, lichaamsgewicht, etc. De minimal clinical important difference (MCID) van de 6-minuten looptest werd bepaald op 17.8 meter in kwetsbare Aziatische ouderen⁹², op 33 meter in patiënten met pulmonale hypertensie⁹³, en 24-45 meter in patiënten met longfibrose⁹⁴. In COPD patiënten werd een verschil van 45 meter beschouwd als klinisch relevant en 30 meter het verschil waarbij patiënten zich 'een klein beetje beter gingen voelen'. In studies bij patiënten met late-onset ziekte van Pompe is deze uitkomstmaat ook gebruikt, daarbij gebruik makend van de MCID voor COPD patiënten, welke ook was gebruikt in het Myozyme assessment report (de EPAR). Het ontbreken van goede uitkomstmaten waarvan de klinische relevantie is vastgesteld bij de late-onset ziekte van Pompe⁹⁵, droeg bij aan de discussie over de kosteneffectiviteit van behandeling met alglucosidase alfa.

3.1.4.3 Bepalen klinische relevantie

Het is dus niet alleen van belang om te bepalen ten aanzien van welke uitkomstmaat of -maten zich meerwaarde moet voordoen, maar ook welke mate van meerwaarde er toe doet (en als klinisch relevant moet worden beschouwd). Als het bijvoorbeeld bij een bepaalde interventie draait om het realiseren van levensverlenging, zal moeten worden beredeneerd hoeveel weken/maanden levensverlenging (ten opzichte van de standaardbehandeling/gebruikelijke behandeling) bereikt moet zijn om te kunnen spreken van een in medisch opzicht wezenlijke verbetering voor de patiënt die tot wijziging van het behandelbeleid leidt. Bij voorkeur gaan wij af op klinisch relevante effectverschillen die in de literatuur zijn vastgesteld. Bij de beoordeling van de literatuur zijn wij daar attent op. Indien er geen of onduidelijke gegevens zijn, kan het raadplegen van de relevante wetenschappelijke beroepsgroepen en patiëntenorganisaties over de vereiste mate van meerwaarde (de klinische relevantie van een effectverschil) meer duidelijkheid geven.

92 Kwok BC, Pua YH, Mamun K, et al. The minimal clinically important difference of six-minute walk in Asian older adults. *BMC Geriatrics* 2013;13:23.

93 Mathai SC, Puthan MA, Lam D, et al. The minimal important difference in the 6-minute walk test for patients with pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2012;186:428-33.

94 Du Bois RM, Weycker D, Albera C, et al. Six-minute walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. *Am J Respir Crit Care Med* 2011;183:1231-7.

95 Lachmann R, Schoser B. The clinical relevance of outcomes used in late-onset Pompe disease: can we do better? *Orphanet J Rare Dis* 2013;8:160.

3.1.5 Welke follow up-duur willen wij zien?

Voor de beoordeling van de effectiviteit is een follow up periode van belang waarin zich het gros van de verwachte gunstige en ongunstige effecten zullen voordoen. Per te beoordelen interventie bepalen wij beargumenteerd bij welke periode dat het geval is. Daarnaast spreekt voor zich dat de gegevens, verzameld tijdens de follow-up duur, zo compleet mogelijk moeten zijn. Wij geven een voorbeeld ter illustratie (zie het kader).

MoM-heupprothese: welke follow-up duur is lang genoeg?

De Metal-on-metal (MoM) resurfacing heupprothese is een aantal jaren gepropageerd als alternatief voor de total-hip replacement (THR), met name voor jongere patiënten met slijtage van de heup.⁹⁶ In 2002 en in 2005 heeft ons instituut de MoM resurfacing heupprothese beoordeeld en geconcludeerd dat deze interventie niet voldoet aan 'de stand van de wetenschap en praktijk', o.a. vanwege een op die momenten nog te korte follow-up duur en onzekerheid over eventuele schadelijke effecten van blootstelling aan metaaldeeltjes. In 2007 vond herbeoordeling plaats. Op dat moment luidde ons oordeel wel positief, aangezien er voor $\pm$ 5000 patiënten een follow-up duur van 5 jaar of langer bekend was en er geen aanwijzingen waren voor mogelijke carcinogene effecten van circulerende metaaldeeltjes. Verder was ook de beroepsgroep van mening dat de MoM resurfacing heupprothese zich afdoende had bewezen. Wij hebben er destijds wel op aangedrongen bij de beroepsgroep om een goede dataregistratie op te zetten, met name voor het bijhouden van de resultaten op lange termijn.

In 2011 ontstond internationaal ophef toen bleek dat het gebruik van bepaalde types MoM-heupprothesen tot een aantal problemen leidde, waaronder lokale reactie van weefsel rondom de heup, een hoog revisiepercentage en hoge bloedspiegels van chroom en kobalt. Klachten en complicaties waren wisselend per prothese en per kliniek. Eén bepaald type prothese werd in de VS van de markt gehaald. De Nederlandsche Orthopaedie Vereniging (NOV) adviseerde daarop haar leden om, tot er meer gegevens beschikbaar zouden zijn, zeer terughoudend te zijn met het plaatsen van MoM prothesen. Er kwam vervolgens veel kritiek op ons instituut, dat immers vier jaar daarvoor had geoordeeld dat MoM resurfacing heupprothesen voldoende evidence based waren om vergoeding uit de basisverzekering te rechtvaardigen. De vraag is of – achteraf gezien – een negatief oordeel, vanwege de voor prothesen relatief beperkte middellange follow-up duur van vijf jaar, passender was geweest.

3.1.6 Informatiebronnen

Een beoordeling start altijd met een oriënterend literatuuronderzoek. Voor dit 'inlezen' raadplegen wij leerboeken en databases met klinische informatie, zoals UptoDate en Dynamed. Verder gaan wij altijd na of de te beoordelen interventie in een richtlijn is opgenomen. Relevante richtlijnen zoeken wij via Diliguide, de Kwaliteitskoepel Medisch Specialisten, Dynamed, GIN⁹⁷, NGC⁹⁸, AWMF⁹⁹ en via de sites van de Nederlandse wetenschappelijke verenigingen.

Aan de hand van de PICO(T) zoeken wij vervolgens naar literatuur die relevant is voor de beoordeling. Meerdere bibliografische databases staan ons ter beschikking om op systematische wijze individuele onderzoeksartikelen en systematische reviews¹⁰⁰ te traceren, zoals Medline via Pubmed, Cochrane

⁹⁶ De levensduur van de THR is sterk afhankelijk van de activiteit van de patiënt: hoe actiever, hoe korter de levensduur. Dit betekent dat er in de regel zo lang mogelijk gewacht wordt met de plaatsing van een THR, en dat na verloop van tijd de heupprothese vervangen moet worden. Met name voor jongere patiënten met slijtage van de heup zou de MoM-prothese voordelen bieden boven een conservatief, afwachtend beleid, o.a. vanwege betere succespercentages dan bij ouderen, en omdat vervanging op latere leeftijd met een THR beter mogelijk is.

⁹⁷ Guidelines International Network.

⁹⁸ National Guideline Clearinghouse: betreft een Amerikaanse databank.

⁹⁹ Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften.

¹⁰⁰ In een systematische review wordt de beschikbare informatie uit de oorspronkelijke onderzoeken op een transparante, valide en reproduceerbare wijze gezocht, geselecteerd, geëxtraheerd, beoordeeld en samengevoegd. Zie paragraaf 3.2.2 van Scholten RJPM, Offringa M, Assendelft WJ van. Inleiding in evidence-based medicine. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

library¹⁰¹, NICE¹⁰² en - al naar gelang het onderwerp - CINAHL¹⁰³, PsycInfo¹⁰⁴ en de HTA-database. Wij gaan in beginsel uitsluitend af op gepubliceerde en peer-reviewed literatuur.

De buitenlandse verzekeringsstatus van een interventie gaan wij na via de Amerikaanse grote verzekeraars zoals AETNA, CIGNA, Blue Cross Shield en via “zusterorganisaties” zoals GBA, HAS en KCE¹⁰⁵.

Aanvullende informatie zoeken wij via de HTA-organisaties, zoals LBI en KCE¹⁰⁶, het EUNetHTA netwerk⁷⁷ en via eerder genoemde UptoDate en Dynamed.

3.2 Beoordelen en graderen kwaliteit van het bewijs

De volgende stap is om de met behulp van de PICO(T) gevonden en geselecteerde informatie te beoordelen: wat is de kwaliteit van het daaruit te destilleren bewijs? Deze beoordeling kent een aantal aspecten.

3.2.1 Aspecten die de kwaliteit van het bewijs bepalen

3.2.1.1 Methodologische aspecten

Bij de beoordeling op methodologische aspecten wordt nagegaan of er tekortkomingen zijn in de opzet of uitvoering van het onderzoek, waardoor de resultaten vertekend en daarmee minder geloofwaardig kunnen zijn. De inschatting van de kans op vertekening noemen we ook wel de ‘risk of bias’. Het gaat hier om de interne validiteit.

Aspecten die bij het beoordelen van de risk of bias worden bekeken zijn:

- randomisatie;
- geblindeerde toewijzing van de interventie;
- blinding van patiënten en behandelaars;
- blinding van effectbeoordelaars;
- vergelijking van de groepen aan het begin van het onderzoek;
- volledigheid van follow up;
- analyse van patiënten in de groep waarin ze waren gerandomiseerd;
- gelijke behandeling van de groepen, afgezien van de interventie;
- selectieve publicatie en sponsoring.¹⁰⁷

3.2.1.2 Belang van de resultaten

De uitkomsten moeten uiteraard van belang zijn voor de patiënt: gaat het om klinisch relevante uitkomstmaten en is er een klinisch relevant verschil zichtbaar? Het gaat dan om de beoordeling van de relevantie en de grootte van het in de studies gevonden effect en hoe betrouwbaar dat is gemeten (de precisie van de effectschatting)¹⁰⁸. Ook de geloofwaardigheid van het effect is van belang. Zie hierna onderdeel 3.2.1.4 en paragraaf 3.2.4 over de GRADE-methodiek.

3.2.1.3 Generaliseerbaarheid

Het gaat hier om de vraag of de resultaten van de gevonden studies ook van toepassing zijn op andere dan de in het onderzoek betrokken patiënten. De in de studies geïnccludeerde patiënten kunnen bijvoorbeeld in een specifieke leeftijdsgroep vallen of uit alleen mannen of alleen vrouwen bestaan. Dit wordt ook wel externe validiteit genoemd.

¹⁰¹ Betreft een set databases waaronder de door de Cochrane Collaboration onderhouden database of Cochrane reviews.

¹⁰² National Institute for Clinical Excellence.

¹⁰³ Database met verpleegkundige en paramedische literatuur.

¹⁰⁴ Database voor psychologie en gedragswetenschappen.

¹⁰⁵ Respectievelijk: Der Gemeinsame Bundesausschuss in Duitsland, Haute Autorité de Santé in Frankrijk en Federaal Kenniscentrum voor de Gezondheidszorg in België.

¹⁰⁶ Respectievelijk: Ludwig Boltzmann Institut in Oostenrijk en Der Gemeinsame Bundesausschuss in Duitsland.

¹⁰⁷ Zie pag. 58 van Scholten RJPM, Offringa M, Assendelft WJJ van. Inleiding in evidence-based medicine. 4e druk. Houten: Bohn Stafleu van Loghum, 2014.

¹⁰⁸ Dit wordt uitgedrukt met behulp van een 95%-betrouwbaarheidsinterval.

3.2.1.4 Onderzoeksmethodiek (EBRO-classificatie respectievelijk GRADE)

Ook de gebruikte onderzoeksmethodiek ('design') bepaalt de kwaliteit van het bewijs. Tot nog toe hebben wij bij onze beoordelingen voor dit onderdeel de EBRO-classificatie gevolgd. Geleidelijk aan hebben wij een andere methodiek, de GRADE-methodiek, voor onze beoordelingen geïntroduceerd.

De volgende paragraaf gaat in op de EBRO-classificatie en de 'passend bewijs-benadering' die wij in aanvulling daarop hebben ontwikkeld. Daarna komt in een aparte paragraaf de GRADE-methodiek aan bod. De 'passend bewijs-benadering' vormt ook een aanvulling op GRADE.

3.2.2 EBRO-classificatie

De EBRO-classificatie is een classificatie om de mate van bewijskracht van de evidence ('de sterkte' van het bewijs) in kaart te brengen. Uitgaande van het onderzoeksdesign, worden 'levels of evidence' toegekend.¹⁰⁹ Daarmee ontstaat een hiërarchie in evidence (bewijskracht), waarvoor in principe geldt dat sterkere evidence zwakkere evidence verdringt.

De EBRO-classificatie luidt als volgt (hieronder alleen weergegeven voor therapeutische interventies):

- A1: systematische review van tenminste twee onafhankelijk van elkaar uitgevoerde onderzoeken van A2-niveau;
- A2: gerandomiseerd dubbelblind vergelijkend klinisch onderzoek van goede kwaliteit en voldoende omvang (RCT);
- B : vergelijkend onderzoek, maar niet met alle kenmerken van A2;
- C : niet-vergelijkend onderzoek;
- D : mening van deskundigen.

Ter toelichting hierbij het volgende. Een goed opgezet en uitgevoerd gerandomiseerd vergelijkend onderzoek (randomised controlled clinical trial – RCT) geeft in principe de minste kans op vertekening van het effect en geeft derhalve de hoogste mate van zekerheid over de causale relatie tussen de interventie en het waargenomen effect. Anders gezegd: een goede RCT geeft in het algemeen de grootst mogelijke mate van zekerheid dat het gevonden effect ook inderdaad 'waar' is. Door de te onderzoeken interventie ad random (bijvoorbeeld door loting) te verdelen over de onderzoekspopulatie, worden in principe twee groepen gecreëerd die uitsluitend verschillen wat betreft de interventie, maar die verder identiek zijn aan elkaar. Een gevonden verschil in effectiviteit kan dan in principe uitsluitend worden toegeschreven aan de interventie en niet aan andere, bekende (bijvoorbeeld de leeftijd) of onbekende vertekende factoren.

Dat betekent dat dit studiedesign (RCT), indien goed opgezet en uitgevoerd, in het algemeen de effectiviteit van een interventie het beste kan aantonen. Andere onderzoeksdesigns hebben als nadeel dat er meer kans is op confounding, i.e. de mogelijkheid dat andere (onbekende) factoren een rol spelen die kunnen leiden tot onder- of overschatting van het verschil in effect van de interventie t.o.v. de behandeling waarmee vergeleken wordt.

Onderkend moet echter worden dat een RCT ook nadelen heeft en beperkingen kent. Een RCT geeft niet altijd de dagelijkse praktijk weer, vanwege het soms zeer streng toepassen van in- en exclusiecriteria, en vanwege de 'extra' aandacht die een patiënt krijgt. Een RCT is erg duur om uit te voeren, en patiënten willen niet altijd dat het lot hun behandeling bepaalt. Ook is een RCT niet geschikt om (zelden voorkomende) bijwerkingen of problemen op de lange termijn te ontdekken.

Verder is het zo dat er situaties denkbaar zijn waarin een RCT niet nodig is en/of een RCT niet haalbaar is. Uiteraard moeten wij daar bij onze beoordelingen rekening mee houden en dat in onze overwegingen betrekken. Met dat doel hebben wij een kader ontwikkeld aan de hand waarvan wij kunnen afwegen en beargumenteren, wat - gelet op de concrete casus - als 'passend bewijs' kan dienen.

¹⁰⁹ EBRO staat voor Evidence Based Richtlijn Ontwikkeling. Zie Kwaliteitsinstituut voor de Gezondheidszorg CBO. Evidence-based Richtlijnontwikkeling. Handleiding voor werkgroepleden. Utrecht, 2007. Zie verder de Kennisbank voor het maken en herzien van richtlijnen (HARING) op www.zorginstituutnederland.nl.

3.2.3 'Passend bewijs'

Een RCT is niet in alle gevallen nodig en ook niet altijd haalbaar. In de literatuur zijn daar verschillende publicaties over verschenen. Zo hebben Glasziou et al.¹¹⁰ een lijst opgesteld van situaties waarin het denkbaar is dat een RCT niet nodig is:

- als uit observationeel onderzoek is gebleken dat het effect van de interventie zeer groot is, zo groot dat onwaarschijnlijk is dat dit door confounding factors wordt veroorzaakt;
- als er een duidelijk pathofysiologisch werkingsmechanisme ten grondslag ligt aan het effect (plausibiliteit)¹¹¹ ;
- als er een duidelijke relatie in de tijd is tussen interventie en effect;
- als het effect consistent is in meerdere studies;
- als er een duidelijke dosis-respons relatie is.

Dergelijke situaties vergroten namelijk het vertrouwen dat er een oorzakelijk verband is tussen de interventie en het waargenomen effect. De belangrijkste hiervan zijn de relatie in de tijd of dosis, de grootte van het effect en de plausibiliteit.¹¹²

Er zijn ook situaties denkbaar waarin een RCT niet haalbaar is, zoals:

- er is sprake van een zeldzame ziekte¹¹³;
- er is sprake van een ziektebeeld met een slechte prognose, waarvoor geen alternatieve behandeling bestaat;
- patiënten/behandelaars zijn niet bereid om (zich) te (laten) randomiseren.¹¹⁴

Ook de Raad voor de Volksgezondheid en Zorg (RVZ) heeft in een publicatie uit 2007 – “Passend bewijs” geheten - aandacht gevraagd voor de ‘overwaarding’ van RCT’s binnen het zorgdomein en pleitte ervoor om het soort bewijsmateriaal dat nodig (passend) is te laten afhangen van de vorm van zorg waar het om draait.¹¹⁵

Deze publicaties, maar ook de ervaring dat onze positieve beoordelingen kunnen stoelen op bewijs uit andere studies dan RCT’s, zijn voor ons aanleiding geweest om de ‘passend bewijs benadering’ te ontwikkelen.¹¹⁶ Daarmee kunnen wij voorafgaand aan het verzamelen van literatuur op systematische wijze een inschatting maken van de onderzoekskenmerken die - gegeven de interventie en de indicatie - wenselijk en haalbaar zijn, oftewel ‘passend’ zijn.

Aan de hand van een vragenlijst (zie de bijlage) kan worden achterhaald welke onderzoekskenmerken wenselijk respectievelijk noodzakelijk zijn voor het vaststellen van de effectiviteit van de te beoordelen interventie en indicatie. Zo wordt bijvoorbeeld gevraagd naar het natuurlijk beloop van de aandoening. Als het gaat om een aandoening met spontane schommelingen in het natuurlijk beloop, dan is onderzoek met een, eventueel historische, controlegroep als een van de onderzoekskenmerken zeer wenselijk. Zonder controlegroep zou namelijk onduidelijk zijn of eventuele positieve uitkomsten toe te schrijven zijn aan de interventie zelf of aan het natuurlijk beloop, of aan beide aspecten.

¹¹⁰ Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise. *BMJ*. 2007;334:349-51. Zie ook: Schünemann H, Hill S, Guyatt G, et al. The GRADE approach and Bradford Hill's criteria for causation. *J Epidemiol Community Health* 2011;65:392-5.

¹¹¹ Wij gaan nog nader uitwerken op welke wijze wij mechanistische argumenten betrekken bij de beoordelingen.

¹¹² Ook de NICE heeft in zijn guidelines aangegeven wanneer het een RCT niet nodig acht: namelijk als voldaan is aan alle van de volgende criteria: 1) er is een slechte uitkomst te verwachten als de patiënt niet behandeld wordt; 2) de behandeling geeft een dramatische verbetering, die zo groot is dat het onwaarschijnlijk is dat dit door bias komt; 3) de bijwerkingen van de interventie zijn acceptabel; 4) er is geen alternatieve behandeling; 5) er is een overtuigende pathofysiologische basis. Dit is ontleend aan: NICE. Developing NICE guidelines: the manual. 2014. Beschikbaar via <http://www.nice.org.uk/article/pmg20/resources/non-guidance-developing-nice-guidelines-the-manual-pdf>.

¹¹³ Ook als er sprake is van een zeldzame ziekte kun je onderzoek doen, bijvoorbeeld een n = 1 trial.

¹¹⁴ Zie paragraaf 5.d.1 van CVZ. Beoordeling stand van de wetenschap en praktijk. Diemen, 2007. Rapportnr. 254. Beschikbaar via www.zorginstituutnederland.nl.

¹¹⁵ Raad voor de Volksgezondheid en Zorg. Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid. Den Haag: Centrum voor ethiek en gezondheid, 2007. Signalering ethiek en gezondheid 2007/4.

¹¹⁶ Ons instituut heeft de vragenlijst om te bepalen welk bewijs passend is voor het vaststellen van de effectiviteit van een medisch specialistische behandeling, ontwikkeld samen met het Rotterdamse Institute for Medical Technology Assessment (IMTA).

Verder bevat de betreffende vragenlijst vragen die helpen te achterhalen welke onderzoekskenmerken haalbaar zijn. Bij uitkomsten in de verre toekomst zonder mogelijke intermediaire uitkomsten is randomisatie bijvoorbeeld vaak niet mogelijk.¹¹⁷

De vragenlijst is met name ontwikkeld voor de beoordeling van medisch specialistische zorg, maar in potentie breder toepasbaar. Voor de langdurige zorg gaan wij op dit moment na of en zo ja welke specifieke aanpassingen nodig zijn om de ‘passend bewijs benadering’ ook te kunnen gebruiken voor dat zorgdomein.

Bedoeling is om in al in de zoekfase (na het opstellen van de PICO) de inschatting – het passend bewijsprofiel – met behulp van de vragenlijst te maken. Op basis van gevonden literatuur of na consultatie van de wetenschappelijke verenigingen van beroepsgroepen en patiëntenorganisaties kan het profiel eventueel worden bijgesteld.

3.2.4 GRADE-methodiek

Zoals gezegd is er een nieuwe methode ontwikkeld om de kwaliteit van het wetenschappelijk bewijs systematisch te beoordelen en te graderen, de zogenoemde GRADE-methode. Wij zijn ertoe overgegaan om GRADE geleidelijk te introduceren als onderdeel van onze werkwijze bij de beoordeling en gradering van het gevonden bewijs.¹¹⁸ Wij moeten nog verdere ervaring opdoen met deze methodiek. Zo zal nog moeten blijken of deze wel voor alle beoordelingen die wij doen, bruikbaar is.

GRADE staat voor ‘Grading of Recommendations Assessment, Development and Evaluation’. De GRADE-methode verenigt de verschillende aspecten die de kwaliteit van het bewijs bepalen. Met behulp van deze methode kan een inschatting van de kwaliteit van het in totaal verzamelde bewijs worden gegeven, d.w.z. van de mate van vertrouwen dat het geschatte effect daadwerkelijk bestaat. Een belangrijk kenmerk van GRADE is dat per uitkomstmaat de zogenaamde ‘body of evidence’ wordt bepaald en beoordeeld. Daarbij kan de kwaliteit van RCT’s worden afgewaardeerd, resp. de kwaliteit van observationele studies worden opgewaardeerd, afhankelijk van bepaalde, omschreven beoordelingsgronden. GRADE bevordert – meer dan de EBRO-methodiek – een systematische, integrale en transparante beoordeling van de literatuur.¹¹⁹

Welke kenmerken heeft de GRADE-methodiek?

3.2.4.1 Uitkomstmaten als vertrekpunt

Allereerst is het noodzakelijk om vast te stellen welke patiëntrelevante uitkomstmaten van belang zijn voor het uiteindelijke oordeel. Bij voorkeur gaan wij uit van gevalideerde uitkomstmaten. Voor de GRADE-beoordeling worden de uitkomstmaten onderverdeeld in drie categorieën, afhankelijk van hun belang voor de uiteindelijke besluitvorming: cruciale, belangrijke en niet-belangrijke uitkomstmaten.

Uitgaande van de gevonden studies wordt een evidence synthese (een systematische review) gemaakt en wordt – indien mogelijk – per uitkomstmaat een geschat (gepooled) effect gegenereerd. Indien er een bestaande, recente systematische review van goede kwaliteit is, gebruiken wij die. Is dat niet het geval dan maken wij zelf een review of besteden het opstellen van een review uit.

¹¹⁷ Heymans J, Kleijnen S, Verstijnen I, Passend bewijs bij het bepalen van effectiviteit. *Ned Tijdschr Geneeskd* 2013;157:A5479. De Engelse term voor de vragenlijst is Feasible-Information-Trajectory (FIT).

¹¹⁸ De indeling in ‘levels of evidence’ volgens de EBRO-classificatie gebruiken wij in principe niet meer.

¹¹⁹ Bij GRADE worden alle aspecten die de kwaliteit van het relevante bewijs bepalen integraal meegenomen. De EBRO-methodiek is wezenlijk anders. Het is een gradering van individuele studies die vooral wordt bepaald door het studiedesign. De kwaliteit van het bewijs wordt echter niet alleen daardoor bepaald, maar is ook afhankelijk van andere aspecten, zoals onnauwkeurigheid en inconsistentie van resultaten. Bij GRADE worden al die aspecten meegenomen. Volgens de GRADE-methode moeten de cruciale en belangrijke uitkomstmaten expliciet worden benoemd. Vervolgens wordt per cruciale en per belangrijke uitkomstmaat op basis van alle studies die over die uitkomstmaten rapporteren, een totaal oordeel over de geloofwaardigheid van de betreffende uitkomstmaat gegeven (de zogenoemde ‘body of evidence’ per uitkomstmaat). Bij de EBRO-methode is daarvan geen sprake. In de volgende publicatie worden de verschillen inzichtelijk gemaakt aan de hand van een voorbeeld uit een richtlijn, waarbij de beoordeling plaatsvindt zowel volgens de oude graderingsmethode als volgens de GRADE-methode: Boluyt N, Rottier BL, Langendam MW. Richtlijnen worden transparanter met GRADE-methode. *Ned Tijdschr Geneeskd*. 2012;156:A4379.

3.2.4.2 Downgraden

Als in de body of evidence een RCT (of meerdere RCT's) beschikbaar is (zijn), is het vertrekpunt bij GRADE dat de kwaliteit van het bewijs hoog is. Er kunnen echter verschillende factoren een rol spelen, waardoor de kwaliteit van de 'body of evidence' naar beneden bijgesteld moet worden. Deze factoren zijn als volgt gegroepeerd:

- beperkingen in opzet en uitvoering van de studie, bijvoorbeeld door randomisatiefouten of door uitval in de follow-up;
- inconsistentie van de resultaten van de verschillende studies, bijvoorbeeld tegengestelde effecten;
- indirect bewijs, bijvoorbeeld als er gebruik wordt gemaakt van intermediaire of surrogaat-uitkomstmaten of als er geen head-to-head vergelijking van de experimentele en standaard-interventie beschikbaar is;
- imprecisie/onnauwkeurigheid van het geschatte effect, bijvoorbeeld ruime betrouwbaarheidsintervallen doordat klinische events weinig voorkomen, of doordat de klinische relevantie van een verschil in effect niet goed is gedefinieerd;
- publicatiebias, bijvoorbeeld als aannemelijk is dat studies met negatieve resultaten niet zijn aangeboden voor publicatie.

3.2.4.3 Upgraden

De kwaliteit van de 'body of evidence' kan ook worden verhoogd. Resultaten uit een observationele studie betekenen bijvoorbeeld in de GRADE-methodiek dat de kwaliteit van het bewijs in principe laag is, omdat er relatief veel vertekenende factoren zijn die het vertrouwen in het gevonden effect kunnen beïnvloeden. De kwaliteit van het bewijs kan echter toenemen in de volgende gevallen:

- als het effect groot is. Een groot effect kan het vertrouwen erin doen toenemen omdat het dan minder waarschijnlijk is dat het effect wordt veroorzaakt door (onbekende) vertekenende factoren;
- als er een relatie is tussen dosis en respons. De aanwezigheid van een dosis-respons relatie doet het vertrouwen toenemen, omdat dit een causaal verband aannemelijk maakt;
- als beredeneerd kan worden dat correctie van vertekenende factoren tot een sterke toename van het effect zou leiden.

Aan de hand van bovenstaande factoren kan dus de kwaliteit van het totale bewijs per uitkomstmaat naar boven of naar beneden worden bijgesteld. Dit 'up- of downgraden' dient goed beargumenteerd te worden en in een evidence-tabel te worden samengevat.

3.2.4.4 Beoordeling kwaliteit van het bewijs

GRADE heeft de volgende niveaus van kwaliteit:

- hoge kwaliteit: er is veel vertrouwen in de schatting van het effect (en verder onderzoek zal zeer waarschijnlijk het vertrouwen in het gevonden effect niet veranderen);
- middelmatige kwaliteit¹²⁰: er is redelijk vertrouwen in de schatting van het effect (en verder onderzoek zal waarschijnlijk invloed hebben op het vertrouwen in het gevonden effect, en kan mogelijk de schatting van het effect veranderen);
- lage kwaliteit: er is beperkt vertrouwen in de schatting van het effect (en verder onderzoek zal zeer waarschijnlijk een belangrijke invloed hebben op het vertrouwen in het gevonden effect en zal waarschijnlijk de schatting van het effect veranderen);
- zeer lage kwaliteit: er is zeer weinig vertrouwen in de schatting van het effect (de schatting van het effect is zeer onzeker).

Bij GRADE vormen RCT's het startpunt voor het bepalen van de effectiviteit van therapeutische interventies. De passend bewijs benadering, zoals wij die hiervoor in paragraaf 3.2.3 uiteen hebben gezet, blijft dan ook 'onder GRADE' relevant.

¹²⁰ De Engelse term is 'moderate quality of evidence'. De vertaling die wij gebruiken is: 'middelmatige kwaliteit van bewijs', met de toevoeging dat er een 'redelijk' vertrouwen is in de schatting van het effect. De vertaling "middelmatige kwaliteit" wordt gebruikt in het artikel van Ali U, Boormeester M. Beoordelen van bewijs voor nieuwe behandelingen. Ned Tijdschr Geneesk. 2014;158:A7281. Zie verder paragraaf 3.2.4 over de GRADE-systematiek.

3.3 Vaststellen eindbeoordeling/conclusie

Uiteindelijk bepalen wij als laatste stap welke conclusie wij m.b.t. de effectiviteit kunnen trekken mede op basis van de beoordeelde literatuur. Het gaat ons, zoals eerder gezegd, om de relatieve effectiviteit. Het komt erop neer dat wij het volgende afwegen: vinden wij de ‘netto toevoeging’ van de interventie in vergelijking met de al bestaande zorg een gewenste, relevante toevoeging en voldoende/groot genoeg, en hebben wij er voldoende vertrouwen in dat deze toevoeging ook daadwerkelijk optreedt? In de afweging die wij maken om m.b.t. een concrete interventie ons eindoordeel te bepalen, betrekken wij in beginsel de volgende argumenten.

3.3.1 Argumenten voortkomend uit balans tussen voor- en nadelen van de interventie

Het draait dan om de vraag of de gunstige effecten voldoende opwegen tegen eventuele ongunstige effecten. In dat verband nemen wij ook in ogenschouw of de uitkomstmaten die voor de beoordeling cruciaal dan wel belangrijk zijn een klinisch relevant verschil laten zien tussen interventie en comparator.

3.3.2 Argumenten voortkomend uit kwaliteit van het wetenschappelijke bewijs

In onderdeel 3.2 hebben wij uiteengezet op welke wijze wij tot een oordeel komen over de kwaliteit van het bewijs. De GRADE-methodiek kent een viertal uitkomsten. De benadering die wij – daarvan uitgaande – voor de eindbeoordeling/conclusie kiezen, is op hoofdlijnen in de regel als volgt:

- Als er wetenschappelijk bewijs is van hoge kwaliteit op cruciale of belangrijke uitkomstmaten zal dat bewijs in beginsel doorslaggevend zijn. De geloofwaardigheid van de gevonden resultaten is dan immers heel groot. Dit gaat op voor zowel positieve als negatieve resultaten: als er bewijs van hoge kwaliteit is voor een positief effect van de interventie dan leidt dit in beginsel tot een positief standpunt. Is er bewijs van hoge kwaliteit voor een negatief effect van de interventie, dan leidt dit in principe tot een negatief standpunt.
- Bewijs van zeer lage kwaliteit op cruciale of belangrijke uitkomstmaten betekent dat de geloofwaardigheid van de gevonden resultaten – of die nu positief of negatief zijn – uiterst gering is. Dit vormt een belemmering om te concluderen of de interventie nu wel of niet effectief is. Dit gebrek aan bewijs van voldoende kwaliteit is in beginsel van doorslaggevende betekenis en zal leiden tot een negatief standpunt. Een positieve waardering van de interventie bij professionals of patiënten – mocht daar sprake van zijn – zal in de regel de grote onzekerheid over de gevonden effecten niet kunnen compenseren.
- Is het bewijs van middelmatige of lage kwaliteit, dan betekent dit dat het vertrouwen dat wij kunnen hebben in het gevonden effect in dat geval redelijk respectievelijk beperkt is. Wij zullen dan in de regel een nadere afweging maken. Op de argumenten die wij dan in de afweging betrekken gaan wij hierna in de paragrafen 3.3.3 en 3.3.4 in.

Voor de goede orde: hetgeen in deze derde bullit staat, gaat alleen op als er sprake is van een positief effect. Duidt het bewijs van middelmatige of lage kwaliteit op een negatief effect, dan is dat in de regel onvoldoende om een positief standpunt over ‘de stand van de wetenschap en praktijk’ in te nemen.

3.3.3 Argumenten voortkomend uit de ‘passend bewijs benadering’

Zoals eerder gezegd, vormen bij de gradering volgens GRADE de RCT's het startpunt voor het bepalen van de effectiviteit van therapeutische interventies. Het kan zijn dat uit de GRADE beoordeling komt dat verder onderzoek (zeer waarschijnlijk) invloed kan/zal hebben op het vertrouwen in het gevonden effect, maar dat het tegelijkertijd – gelet op de uitkomst van de beantwoording van de passend bewijs vragenlijst – niet mogelijk is verder onderzoek in de vorm van een RCT te vergen.

3.3.4 Beargumenteerde opvattingen/garanties beroepsgroepen en patiënten

Bij de beoordeling van de effectiviteit (‘de stand van de wetenschap en praktijk’) van een interventie volgens de principes van EBM gaat het niet alleen om de wetenschappelijke kant, maar ook om de in praktijk gevormde expertise en ervaringen van zorgverleners en zorggebruikers. Met de ‘passend bewijs benadering’ wordt al voor een deel recht gedaan aan het meewegen van deze laatste aspecten. De passend bewijs benadering houdt immers rekening met wat wenselijk respectievelijk noodzakelijk is en wat – praktisch gezien – haalbaar is. Daarbij betrekken wij onder meer de kennis van pathofysiologische mechanismen en de klinische ervaring.

Daarnaast nemen wij ook de beargumenteerde opvatting van de beroepsgroepen en patiënten over de ‘waarde’ van de interventie in ogenschouw. Als in een concreet geval het bewijs voor een positief effect van middelmatige of van lage kwaliteit is, vindt er in de regel een nadere afweging plaats. Er is in dat geval enige mate van bewijs voor effectiviteit (bijvoorbeeld voor een specifieke patiëntengroep), maar het bewijs is niet van hoge kwaliteit. Een beargumenteerde en op consensus berustende opvatting van de betrokken beroepsgroep(en) over de waarde van de interventie kan er – in aanvulling op de aanwezige ‘body of evidence’ – mogelijk toe leiden dat wij e.e.a. toch voldoende vinden om – bij de benoemde indicatie(s) – te kunnen spreken van een effectieve interventie. Ditzelfde kan eventueel opgaan in geval de aanwezige ‘body of evidence’ wordt aangevuld met garanties van de beroepsgroep(en) die een gepast gebruik van de betreffende zorg bevorderen. Er zal wel aan een aantal voorwaarden moeten zijn voldaan, waaronder in ieder geval:

- Er zijn overtuigende redenen om aan te nemen dat verder onderzoek dat de kwaliteit van het bewijs naar effectiviteit van de interventie zou kunnen verhogen zeer waarschijnlijk niet zal (kunnen) plaatsvinden⁶²;
- De wetenschappelijke verenigingen van professionals hebben (bij voorkeur met steun van de patiëntenverenigingen) deugdelijk en overtuigend onderbouwd waarom zij, ondanks de middelmatige of lage kwaliteit van het bewijs, de interventie toch bij (een) bepaalde indicatie(s) als een effectieve behandelmethode beschouwen;
- Deze opvatting moet steunen op brede consensus binnen/tussen de relevante verenigingen van professionals (en patiënten) en in een publicatie (bijvoorbeeld in een richtlijn of consensusdocument) zijn geformaliseerd.¹²¹

Om wat voor soort situaties kan het gaan? Wij noemen enkele voorbeelden:

- In een recente, actuele richtlijn wordt de positieve aanbeveling voor de inzet van de interventie – ondanks middelmatig/laag wetenschappelijk bewijs – naar ons inzicht overtuigend beargumenteerd;
- De beroepsgroepen gaan ertoe over om nadere gegevens te verzamelen. Met name als het gaat om onzekerheid over de effecten op lange termijn of over de toepassing in de dagelijkse praktijk, kan het starten van een dataregistratie zinvol zijn. Een goede dataregistratie kan deze gegevens na enkele jaren opleveren. Het is dan wel nodig dat wij over het tijdstip van evaluatie van de data en onze eventuele herbeoordeling met de beroepsgroepen afspraken maken;
- De beroepsgroepen garanderen dat het aanbod van de zorg door middel van kwaliteitsafspraken wordt ingekaderd, bijvoorbeeld door te zorgen voor een goed onderbouwd indicatieprotocol en eventuele andere kwaliteitseisen voor zorgverlener of instelling. Ook dan zullen tussen beroepsgroepen en ons instituut afspraken moeten worden gemaakt over monitoring en evaluatie hiervan.

¹²¹ Zie de toevoeging in paragraaf 2.4: Dit vereiste (brede consensus binnen/tussen de relevante beroepsgroepen) volgt uit het normatieve karakter van het begrip ‘de stand van de wetenschap en praktijk’. ‘Praktijk’ is niet: ‘dat wat individuele professionals (gewoonlijk) feitelijk aan zorg leveren’. Evenmin gaat het bij ‘praktijk’ om de mening van individuele artsen (en individuele patiënten) over de waarde van de interventie. Wat meegenomen moet worden in de afweging is dat wat de beroepsgroep als geheel maatgevend vindt respectievelijk als juiste behandelwijze beschouwt. Om die reden richten wij ons voor het overleg/de consultatie tot de wetenschappelijke verenigingen van de zorgaanbieders en tot de patiëntenorganisaties en niet tot individuele zorgaanbieders en individuele patiënten.

Nieuwe behandeling (TAVI): aanvullende kwaliteitsafspraken maakten positief standpunt mogelijk

Ernstige vernauwing van de aortaklep, aortaklepstenose (AS), is een aandoening die vooral op hogere leeftijd optreedt en een slechte prognose kent indien deze niet wordt behandeld. In 2011 hebben wij beoordeeld of een nieuwe interventie, de transcatheter aortaklepvervangende (transcatheter aortic valve implantation, TAVI), voldoet aan ‘de stand van de wetenschap en praktijk’. TAVI, een minimaal invasieve ingreep, is ontwikkeld voor patiënten voor wie de standaardbehandeling niet mogelijk is. De standaardbehandeling is het operatief vervangen van de aortaklep. Ongeveer 30% van de patiënten is echter vanwege leeftijd en comorbiditeit niet operabel of heeft een verhoogd operatierisico.

Een groot aantal cohortstudies en een RCT van goede kwaliteit lieten zien dat TAVI leidt tot een significante levensverlenging (in vergelijking met conservatieve behandeling) bij patiënten met ernstige AS en een onaanvaardbaar hoog operatierisico (inoperabele patiënten). Voor patiënten met een weliswaar verhoogd, maar aanvaardbaar operatierisico, bleek uit (slechts) één RCT van goede kwaliteit dat TAVI niet minder effectief is dan chirurgische klepvervangende wat betreft mortaliteit en hartfunctie, maar dat dit ten koste gaat van een verhoogd risico op CVA. Voor deze patiëntengroep heeft TAVI dus geen voordelen boven operatief ingrijpen. De beschikbare studies gaven geen uitsluitsel over de vraag welke patiënten wel en welke niet operabel zijn. Om tot een positief oordeel te kunnen komen over TAVI was het naar onze mening daarom nodig dat een onderbouwd indicatieprotocol beschikbaar zou komen en dat er waarborgen zouden komen om te bewerkstelligen dat ook in de praktijk zorgvuldige indicatiestelling plaatsvindt.

De wetenschappelijke verenigingen van cardiologen en cardiochirurgen (NVVC en NVT), die ook zelf het belang van zorgvuldige indicatiestelling onderkennen, hebben een gezamenlijk indicatieprotocol opgesteld. Aan de hand daarvan kan een multidisciplinair team per patiënt de risico's (een verhoogd operatierisico dan wel inoperabiliteit) tevoren zo goed mogelijk inschatten. Verder hebben de wetenschappelijke verenigingen in hun richtlijnen voor het uitvoeren van TAVI volumenormen en kwaliteitseisen t.a.v. de behandelaars opgenomen.

Deze aanvullende kwaliteitseisen, die steunen op brede consensus binnen de relevante beroepsgroepen, gevoegd bij de beschikbare wetenschappelijke inzichten, waren voor ons voldoende om te concluderen dat TAVI voor patiënten met een ernstige AS die niet operabel zijn, zorg is conform ‘de stand van de wetenschap en praktijk’.¹²²

3.3.5 Onderbouwing standpunt

Zoals gezegd, betrekken wij in de afweging de in de vorige paragrafen genoemde argumenten. Wij dragen zorg voor een transparante onderbouwing van onze afweging en van het daarop gestoelde standpunt (de zorg voldoet wel respectievelijk niet aan ‘de stand van de wetenschap en praktijk’).

¹²² Ten tijde van dit standpunt maakten wij nog geen gebruik van de GRADE-methode. Het standpunt is van 26 oktober 2011 en is terug te vinden op: www.zorginstituutnederland.nl.

4 Langdurige zorg

4.1 Inleiding

Per 2015 is de langdurige zorg hervormd. De AWBZ is versmald tot de Wlz. Deze wet verzekert (instellings) zorg voor verzekerden die vanwege hun aandoening, beperking of handicap zeer kwetsbaar zijn en daarom een blijvende behoefte hebben aan permanent toezicht of 24 uur zorg.¹² Voor de langdurige zorg geldt ook de voorwaarde dat deze dient te voldoen aan 'de stand van de wetenschap en praktijk'.¹²³ In de praktijk bevindt de langdurige zorg zich in een andere fase van wetenschappelijk onderzoek naar de effectiviteit. In de volgende paragraaf gaan wij in op de achterliggende redenen daarvan. Daarna geven wij aan waarom er meer aandacht moeten komen voor de beoordeling van de effectiviteit van langdurige zorg in het kader van pakketbeheer. Wij schetsen vervolgens welke activiteiten wij ontplooiën om te bevorderen dat toetsing op effectiviteit consequenter kan gebeuren.

4.2 Kenmerken van de langdurige zorg

Wetenschappelijk onderzoek naar de effectiviteit van langdurige zorg vindt nog weinig plaats. De RVZ geeft in zijn signalering over 'passend bewijs'¹¹⁵ aan welke oorzaken in het algemeen ten grondslag liggen aan het ontbreken van effectiviteitsonderzoek. De belangrijkste, meest waarschijnlijke oorzaak is volgens de RVZ het gebrek aan financiële middelen. Met het opzetten van experimenten is veel geld gemoeid en dat geld is vaak niet beschikbaar. Ook het ontbreken van een adequate infrastructuur speelt een belangrijke rol. In een klinische, academische setting heeft men veelal de nodige onderzoeksexpertise en beschikt men over voldoende faciliteiten, terwijl daar in veel andere zorgsettings, zoals de langdurige zorg, geen sprake van is. Als gevolg daarvan heeft zich buiten de academische setting in veel mindere mate een onderzoekscultuur ontwikkeld.

Daarnaast zijn er volgens de RVZ methodologische aspecten die maken dat onderzoek naar de effectiviteit van langdurige zorg lastig is. Zo kan de patiëntengroep erg complex zijn met grote onderlinge verschillen, bijvoorbeeld hoogbejaarden met multimorbiditeit. Ook de geleverde zorg is veelal complex van aard en bovendien multidisciplinair; de zorg varieert vaak per cliënt en is veelal afhankelijk van context, setting en individuele doelen. Voor langdurige zorg geldt dat deze - in vergelijking met curatieve zorg - minder gericht is op genezing en het verhelpen van een aandoening. Doel van de zorg is vooral om de gevolgen van de beperkingen op het dagelijks leven van de cliënt te verminderen. De langdurige zorg heeft daarom meerdere (subjectieve) uitkomstmaten, waaronder kwaliteit van leven, het welbevinden van cliënten, het voorkomen van achteruitgang, het leren omgaan met de gevolgen van de beperkingen en cliënttevredenheid. Er zijn veel verschillende vragenlijsten ontwikkeld om dit soort uitkomsten te meten bij cliënten. Behalve voor cliënttevredenheid (de CQ-indexen) is in de langdurige zorg nog weinig standaardisatie van uitkomsten van zorg en van de manieren om dit te meten.

Wij hebben in 2009 een verkennend onderzoek laten uitvoeren naar de effectmaat kwaliteit van leven in de care. Uit het onderzoek komt naar voren dat in de care bij verschillende domeinen kwaliteit van leven aan de orde kan zijn: psychisch of emotioneel welzijn; sociaal en maatschappelijk welzijn; zingeving en spiritueel welzijn; lichamelijk welzijn; autonomie, controle en regie over het eigen leven; financieel welzijn en administratieve druk. Uit het onderzoek komt verder naar voren dat er veel verschillende 'kwaliteit van leven-instrumenten' voor de langdurige zorg bestaan, maar dat heel weinig van deze vragenlijsten alle genoemde domeinen omvatten. Richt een kwaliteit van leven vragenlijst zich op een beperkt aantal van de genoemde domeinen, dan wordt slechts een deel van de uitkomsten van de zorg gemeten. De effectmeting is dan niet compleet.¹²⁴

¹²³ Als de norm 'de stand van de wetenschap en praktijk' niet van toepassing is, geldt de norm 'verantwoorde en adequate zorg en diensten'. Zie voor verdere uitleg paragraaf 1.1.1.2.

¹²⁴ EMGO+, Maastricht University. Verkenning effectmaat voor de care sector. Rapportage in opdracht van het College voor Zorgverzekeringen (CVZ). Amsterdam, oktober 2009.

Tot slot speelt een rol dat de langdurige zorg, anders dan bij de curatieve zorg, slechts beperkt wordt omschreven in kwaliteitsstandaarden van de diverse beroepsgroepen. Binnen de langdurige zorg bestaat met het opstellen van deze standaarden in beperkte mate ervaring. Bovendien wordt de zorg meestal multidisciplinair uitgevoerd, hetgeen het opstellen van kwaliteitsstandaarden bemoeilijkt.

4.3 Beoordeling effectiviteit langdurige zorg in kader van pakketbeheer

Terugkerend naar het onderwerp in dit rapport: ons instituut heeft tot op heden weinig ervaring met de toetsing van langdurige zorg aan 'de stand van de wetenschap en praktijk'. Gegevens over effectiviteit van langdurige zorg zijn geregeld niet voorhanden, zoals uit de vorige paragraaf naar voren komt. Verder speelt mee dat de eis dat de zorg moet voldoen aan 'de stand van de wetenschap en praktijk' pas per 1 januari 2012 in de AWBZ – thans de Wlz – is geïntroduceerd. In beperkte mate hebben wij dan ook tot op heden het criterium toegepast.

Omdat de langdurige zorg uit collectieve middelen wordt gefinancierd, is het vanuit maatschappelijk oogpunt wenselijk om deze middelen zo goed mogelijk in te zetten. Een toets of de zorg effectief is (te beschouwen), dient daar onderdeel van uit te maken. De urgentie om dat consequenter dan tot op heden te doen stijgt resp. is gestegen vanwege twee ontwikkelingen: de geplande hervorming van de langdurige zorg en de toenemende druk in de afgelopen jaren op de budgetten voor deze zorg. Wij streven er daarom naar, zoals in paragraaf 1.2.3 al is aangestipt, om langdurige zorg consequenter te kunnen toetsen aan 'de stand van de wetenschap en praktijk'.

4.4 Overzicht van activiteiten om te voorzien in toegespitst beoordelingskader

Om ons streven naar een consequentere toetsing te kunnen waarmaken, is het nodig om te komen tot een op de langdurige zorg toegespitste invulling van het beoordelingskader. De invulling moet rekening houden met de specifieke kenmerken en omstandigheden van de langdurige zorg. Hierna zetten wij kort uiteen welke activiteiten wij ontplooiën die hierbij behulpzaam (kunnen) zijn. De verwachting is dat van deze activiteiten een stimulerende werking zal uitgaan en het doen van adequaat effectiviteitsonderzoek in deze sector zal bevorderen, hetgeen uiteindelijk de kwaliteit van de zorgverlening ten goede zal komen.

4.4.1 Onderzoek meetinstrumenten langdurige zorg

Het onderzoek naar een effectmaat in de care (kwaliteit van leven) en beschikbare instrumenten in de langdurige zorg om de kwaliteit van leven te meten, zoals hiervoor genoemd, is een verkenning geweest op basis van literatuur en gesprekken met wetenschappers. In een vervolgtraject dat ons instituut heeft ingezet, vindt een inventarisatie plaats van de instrumenten (kwantitatief en kwalitatief) die in de praktijk worden gebruikt voor het meten van kwaliteit van leven van cliënten in de langdurige zorg. Wij willen er zicht op krijgen met welke instrumenten/vragenlijsten in de praktijk van de zorg gewerkt wordt, wat de ervaring is met deze instrumenten en waarvoor ze gebruikt worden. Bedoeling hiervan is om op termijn - in samenspraak met de sectoren in de langdurige zorg - te kunnen komen tot (enige) standaardisatie wat betreft de manier van effectmeting in de langdurige zorg. Hierbij speelt natuurlijk ook de methodologische kwaliteit van de meetinstrumenten een rol.

4.4.2 Project passend bewijs effectiviteit langdurige zorg

In paragraaf 3.2.3. hebben we de 'passend bewijs benadering' toegelicht. Het betreft een methodiek die in eerste instantie ontwikkeld is voor de medisch specialistische zorg, maar die zoals aangegeven, in potentie breder toepasbaar is. Wij onderzoeken momenteel welke mogelijkheden er zijn om de reeds ontwikkelde vragenlijst 'passend bewijs' bruikbaar te maken voor de langdurige zorg. Hiervoor inventariseren wij de PICO(T) in bestaand onderzoek voor de verschillende zorgvormen in de langdurige zorg en onderzoeken daarbij welke onderzoeksdesigns gebruikt worden. Bedoeling is om het volgende helder geformuleerd te krijgen: welke doelen (O) worden nagestreefd met welke langdurige zorg (I) en – vervolgens – met welk onderzoek kan de (oorzakelijke) relatie tussen beide het best aangetoond worden? Als is vastgesteld wat passende vormen van onderzoek zijn, kan daaruit vervolgens worden afgeleid wat 'passend bewijs' is. Zoals eerder aangegeven, zijn er in de langdurige zorg vele redenen waarom de effectiviteit van zorg lastig te onderzoeken is. Wij houden hiermee rekening en besteden daarom naast gegevens uit kwantitatief onderzoek (zoals gerandomiseerde klinische trials en observationeel onderzoek) ook aandacht aan bewijsvoering door middel van kwalitatief onderzoek.

4.4.3 Bevorderen van totstandkoming kwaliteitstandaarden langdurige zorg

Kwaliteitsbevordering binnen de langdurige zorg is één van de speerpunten van ons kwaliteitsinstituut. Het programma 'Good Practices in de Langdurige zorg' is daartoe in het leven geroepen.¹²⁵ Bedoeling hiervan is om de professionaliteit van de sector te stimuleren en de inzet van goede zorg en de beschrijving daarvan in kwaliteitsstandaarden te bevorderen. Zo maken wij via het "Erkenningstraject good practices" voorbeelden van goede zorg uit de praktijk bekend. Later, na verdere ontwikkeling (zoals bijvoorbeeld onderzoek naar effectiviteit), kunnen deze 'goede interventies' de status van (geregistreerde) kwaliteitstandaard krijgen. Zo komt de kwaliteitscyclus voor de langdurige zorg op gang. Verder is nog van belang de Leidraad Kwaliteitsstandaarden die ons instituut onlangs heeft vastgesteld.¹²⁶ Deze leidraad geeft instructies voor de ontwikkeling en implementatie van kwaliteitsstandaarden. Hij is niet alleen van toepassing op de curatieve zorg die via de eerste lijn en het ziekenhuis wordt verleend, maar geldt tevens voor de preventieve zorg en de langdurige zorg. Ook deze leidraad kan eraan bijdragen dat - meer dan nu nog het geval is - langdurige zorg in kwaliteitsstandaarden wordt beschreven.

¹²⁵ Dit programma bestaat uit drie onderdelen: 1) Erkenningstraject good practices, 2) Onderzoek naar kwaliteit van leven en kwaliteit van zorg en 3) Stimuleren kwaliteitscyclus. Verdere informatie is te vinden op onze website: www.zorginstituutnederland.nl.

¹²⁶ De Leidraad is voorbereid door de Advies- en expertgroep Kwaliteitsstandaarden (AQUA) van de Adviescommissie Kwaliteit van ons instituut en te vinden op onze website: www.zorginstituutnederland.nl.

5 Medische tests

5.1 Klinisch nut

Op 20 januari 2011 heeft ons instituut het rapport “Medische tests, beoordeling stand van de wetenschap en praktijk” uitgebracht.⁹ De uitgangspunten voor de beoordeling van medische tests die daarin zijn neergelegd, blijven bij het verschijnen van het voorliggende rapport over de actualisering van de beoordeling van ‘de stand van de wetenschap en praktijk’ van kracht. Bij de beoordeling van medische tests gaat het wat ons instituut betreft om de effecten van het toepassen van tests op de gezondheid van degenen die de test ondergaan. Wij zijn van mening dat een test niet enkel beoordeeld moet worden op zijn vermogen een fraaie afbeelding of correcte diagnose te verkrijgen. Natuurlijk is wel van groot belang dat een test betrouwbaar is, maar een uiteindelijk positief oordeel over een test moet vooral gebaseerd zijn op aantoonbaar positieve gevolgen voor aan de gezondheid gerelateerde uitkomsten bij degenen die getest worden. Bij de evaluatie van een test beoordelen wij dan ook het hele traject, de test-plus-behandeling-strategie, op effectiviteit, ook wel omschreven als klinisch nut. Op basis van vergelijkend onderzoek tussen de gebruikelijke (‘oude’) en de voorgestelde (‘nieuwe’) test-plus-behandeling-strategie kan het klinisch nut worden bepaald. Dergelijk rechtstreeks bewijs ontbreekt echter vaak. In dat geval gaan wij door middel van de constructie van een vergelijkend analyseraam na of niet-rechtstreeks bewijs behulpzaam kan zijn voor beantwoording van de vraag naar het klinisch nut.¹²⁷

Een voorbeeld van een beoordeling van een medische test geven wij in het volgende kader.

Medische test PCA3 bij diagnostische interventies voor prostaatkanker: klinisch nut staat niet vast

Prostaatkanker is een vorm van kanker die geregeld voorkomt bij mannen. De aandoening brengt vaak geen gezondheidsproblemen met zich mee en behoeft in die gevallen dan ook geen behandeling. In 2011 heeft ons instituut beoordeeld of het toepassen van de gen-expressietest PCA3 voldoet aan ‘de stand van de wetenschap en praktijk’. Concreet draaide het om de vraag of genoemde test effectief is (klinisch nut heeft) als triage-test voor het nemen van een besluit over het al dan niet verrichten van een tweede serie prostaatbipten bij patiënten met verdenking op prostaatkanker en een negatieve eerste serie bipten.

Onze constatering was dat rechtstreeks bewijs, d.w.z. rechtstreeks vergelijkend onderzoek tussen de gebruikelijke en de voorgestelde test-plus-behandeling-strategie, ontbrak. Bij de analyse van het niet-rechtstreekse bewijs bleek de diagnostische accuratesse van de PCA3-test in vergelijking met de tweede serie bipten ernstig beperkt. De PCA3-test bleek dus niet een goede triage-test, omdat bij toepassing ervan veel mannen met prostaatkanker gemist zouden worden. Gegevens over gezondheidsgelateerde effecten van de test-plus-behandeling-strategie ontbraken geheel.

Op basis hiervan concludeerden wij dat het toepassen van de PCA3 test-plus-behandeling-strategie bij patiënten met een eerste serie prostaatbipten zonder kanker en een klinische verdenking op prostaatkanker niet als effectief kan worden geschouwd. Dit leidde tot het standpunt dat PCA3-test niet voldoet aan het criterium ‘de stand van de wetenschap en praktijk’.¹²⁸

5.2 Van EBRO-classificatie naar GRADE

Destijds hebben wij in het rapport uit 2011 over de beoordeling van medische tests aangegeven bij de beoordeling van de literatuur over het klinisch nut van medische tests de op dat moment gebruikelijke

¹²⁷ In wetenschappelijk onderzoek is helaas nog in beperkte mate aandacht voor medische tests. Zo blijkt bijvoorbeeld bij een systematische zoektocht in de literatuur dat een minderheid van de studies naar accuratesse (gepubliceerd in belangrijke tijdschriften) wordt opgenomen in het trialregister en dat slechts 1 op de 22 van deze studies de primaire uitkomstmaten voorafgaand aan voltooiing van de studie registreert. Zie: Korevaar DA, Bossuyt PMM, Hooft L. Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports. *BMJ Open* 2014;4:e004596. Zeker in het licht van de actuele discussie rond ‘personalized medicine’ is het ontbreken van bewijs voor klinisch nut van een test-plus-behandeling-strategie een belangrijk probleem. Juist bij ‘personalized medicine’ zal zich de vraag naar het klinische nut voordoen. De selectie van de patiëntengroep voor de behandeling vindt immers plaats met behulp van een medische test.

¹²⁸ Het standpunt over de medische test PCA3 bij diagnostische interventies voor prostaatkanker van 22 augustus 2011 is te vinden op de website van ons instituut: www.zorginstituutnederland.nl.

EBRO-classificatie en aanbevelingen te gebruiken. Zoals in hoofdstuk 3 vermeld, zijn wij ertoe overgegaan om de EBRO-methode geleidelijk te vervangen door GRADE. Ook voor de beoordeling van het klinisch nut van medische tests is dat het geval.

In 2011 hebben wij al wel aangegeven bij de beoordeling van systematische reviews over de accuratesse van medische tests gebruik te willen maken van het Quadas instrument en niet van de EBRO-classificatie, omdat deze laatste geen goede oplossing biedt bij vragen over de kwaliteit en toepasbaarheid van een test. Zoals we destijds al aankondigden is sindsdien een gereviseerde versie van dit instrument verschenen: QUADAS-2.¹²⁹

Binnen de GRADE Working Group wordt het uitgangspunt ‘klinisch nut’ bij de beoordeling van medische tests onderschreven en internationaal wordt gewerkt aan een nadere methodologische uitwerking hiervan. Zo wordt, binnen het door de Europese commissie gesteunde DECIDE project (Developing and Evaluating Communication Strategies to Support Informed Decisions and Practices Based on Evidence), Workpackage 4 uitgevoerd. Doel is strategieën te ontwikkelen voor de weergave van evidence based aanbevelingen over medische tests. De Universiteit van Amsterdam is verantwoordelijk voor de coördinatie van dit Workpackage.¹³⁰

Ons instituut zal deze methodologische uitwerking uiteraard volgen en zo nodig gebruiken voor de eigen beoordelingen.

¹²⁹ Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med* 2011;155:529-36.

¹³⁰ <http://www.decide-collaboration.eu/WP4>

6 Ingangsdatum en gevolgen van standpunt over ‘de stand van de wetenschap en praktijk’

6.1 Ingangsdatum

Een standpunt over ‘de stand van de wetenschap en praktijk’ heeft gevolgen voor de rechten van verzekerden. Met name als er sprake is van een positief standpunt kan het van belang zijn dat wij aangeven op welk moment voldaan is aan het criterium ‘de stand van de wetenschap en praktijk’. Dat is het moment waarop de zorg tot de te verzekeren prestaties is gaan behoren (indien ook aan de overige voorwaarden is voldaan). Het gaat er dan om dat wordt vastgesteld op welk moment de gegevens die bepalend, doorslaggevend zijn voor de omslag (de conclusie dat de zorg is gaan voldoen aan ‘de stand van de wetenschap en praktijk’) door publicatie openbaar zijn geworden.¹³¹ Het kan bijvoorbeeld zijn dat de gegevens uit een bepaald wetenschappelijk onderzoek of de binnen de beroepsgroep bereikte consensus de doorslag geven. De publicatiedatum van de betreffende gegevens is dan de datum van omslag. Van geval tot geval moet dit worden vastgesteld.

Hierna gaan wij nader in op de gevolgen van een standpunt voor de rechten van verzekerden.

6.2 Gevolgen van een standpunt voor rechten verzekerden

Verzekerden ontlenen hun recht op (vergoeding van) zorg aan de zorgverzekering (polis) die zij met een zorgverzekeraar hebben gesloten. In alle polissen is de voorwaarde opgenomen dat zorg – wil deze onder dekking van de polis vallen – in ieder geval moet voldoen aan ‘de stand van de wetenschap en praktijk’. Wordt door middel van een standpunt vastgesteld dat bepaalde zorg wel of niet voldoet aan ‘de stand van de wetenschap en praktijk’ dan werkt dat direct door in de polis. Aanpassing van de polis is daar niet voor nodig. Verder is relevant dat het moment waarop de verzekerde de medische behandeling ondergaat, bepalend is voor het recht op aanspraak. Is de behandeling op dat moment conform ‘de stand van de wetenschap en praktijk’, dan valt de behandeling – als ook aan eventuele andere gestelde voorwaarden is voldaan – onder de dekking van de zorgverzekering van betrokkene.¹³² Het kan zijn dat dat pas later, als de verzekerde de behandeling al heeft gekregen, duidelijk is respectievelijk wordt vastgesteld. Wat heeft dat voor consequenties?

Hoe de verschillende situaties die zich voor kunnen doen uit voor de verzekerden uitpakken, zetten wij hierna kort op een rij.

6.2.1 Standpunt is in overeenstemming met de uitvoeringspraktijk

Het kan zijn dat een standpunt over ‘de stand van de wetenschap en praktijk’ de bestaande vergoedingspraktijk bevestigt en dat dus de feitelijke situatie overeenstemt met de uitkomst van de beoordeling. Twee situaties zijn te onderscheiden:

- het standpunt luidt dat de interventie voldoet aan ‘de stand van de wetenschap en praktijk’ en de praktijk is al dat zorgverzekeraars de interventie ten laste van de basisverzekering vergoeden;
- het standpunt luidt dat de interventie niet voldoet aan ‘de stand van de wetenschap en praktijk’ en de praktijk is dat zorgverzekeraars de interventie niet ten laste van de basisverzekering vergoeden.

Het bepalen van een omslagpunt (in geval van een positief standpunt) is hier eigenlijk niet relevant. Het standpunt is een bevestiging van hetgeen in de praktijk al plaatsvindt. Deze situaties zullen in de regel niet tot problemen leiden.

¹³¹ Ten tijde van de Ziekenfondswet was het bestuursrecht van toepassing. De hoogste rechterlijke instantie was toen de Centrale Raad van beroep (CRvB). Volgens uitspraak van de CRvB is bepalend het moment waarop de resultaten van het voor de omslag relevante wetenschappelijke onderzoek door publicatie aan de beroepsgroep bekend is gemaakt. Zie CRvB 19 januari 2006, RZA 2006/80.

¹³² Onder de werking van de Zvw, waarbij sprake is van privaatrechtelijke schadeverzekeringen, is het moment waarop de verzekerde de medische behandeling ondergaat bepalend. Zie Groot GRJ de. De stand van de wetenschap en praktijk. Tijdschr Gezondheidsrecht 2006;30:326-50.

6.2.2 Standpunt is niet in overeenstemming met de uitvoeringspraktijk

Het kan ook zijn dat een standpunt over 'de stand van de wetenschap en praktijk' afwijkt van de bestaande vergoedingspraktijk. Behandelingen die in de toekomst liggen (die na de datum van inwerkingtreding van het standpunt gestart worden) zullen conform het nieuwe standpunt afgewikkeld moeten worden. Het is aan zorgverzekeraars om voor verdere afwikkeling van zaken te zorgen. Welke opstelling kiest men voor reeds afgewezen, reeds vergoede respectievelijk lopende behandelingen? Er zijn twee situaties te onderscheiden:

- Het kan zijn dat niet altijd direct wordt onderkend dat een interventie op een bepaald moment is gaan voldoen aan 'de stand van de wetenschap en praktijk', maar dat pas later wordt vastgesteld dat de interventie reeds op een eerder tijdstip onder de dekking van de basisverzekering is komen te vallen. Aan ons standpunt zal dan in de regel terugwerkende kracht worden verbonden. Doet zo'n situatie zich voor dan kan dit betekenen dat een zorgverzekeraar die in het verleden geweigerd heeft de zorg te vergoeden, alsnog de kosten moet vergoeden aan de betreffende verzekerden.
- Het kan zich ook voordoen dat de zorgverzekeraars ervan uitgingen dat de zorg voor vergoeding ten laste van de basisverzekering in aanmerking kwam, terwijl achteraf is komen vast te staan dat de zorg niet voldoet aan 'de stand van de wetenschap en praktijk'. In de regel zullen wij in zo'n geval aan ons standpunt geen terugwerkende kracht verbinden. Wij gaan ervan uit dat zorgverzekeraars in dat geval de kosten van reeds vergoede behandelingen niet van verzekerden zullen terugvorderen. Nieuwe behandelingen (d.w.z. behandelingen die starten na de ingangsdatum van het standpunt) zullen conform het nieuwe standpunt afgewikkeld moeten worden. Voor lopende behandelingen is het advies aan zorgverzekeraars om een redelijke afbouwtermijn in acht te nemen (zie ook paragraaf 6.3.2).¹³³

6.3 Overige gevolgen van een standpunt

Zonder uitputtend te willen zijn, geven wij nog wat andere aspecten weer die kunnen samenhangen met een standpunt over 'de stand van de wetenschap en praktijk'.

6.3.1 Zorgverzekeringen (polissen)

Zoals gezegd, werkt een standpunt over 'de stand van de wetenschap en praktijk' direct door in de polis. Zorgverzekeraars behoeven niet per se de polis daarvoor aan te passen. Toch kan er voor een zorgverzekeraar aanleiding zijn om – als duidelijk is geworden dat bepaalde zorg onder de dekking is gaan vallen – de polis aan te passen. Een zorgverzekeraar kan ervoor kiezen om in de polis op te nemen dat de verzekerde zich voor de betreffende zorg moet wenden tot een bepaald soort zorgaanbieder. Ook kan een zorgverzekeraar ervoor kiezen om m.b.t. de betreffende zorg een toegangsvoorwaarde in de polis op te nemen, zoals bijvoorbeeld het vereiste van voorafgaande toestemming. Dit soort beperkende voorwaarden kunnen uiteraard niet met terugwerkende kracht worden geïntroduceerd.

Hoe zit het met polisaanpassingen? Naast de jaarlijkse zorgpolis die zorgverzekeraars hun verzekerden aanbieden, mogen zowel de polis als de bijbehorende reglementen door zorgverzekeraars tussentijds worden gewijzigd. Zorgverzekeraars moeten hun verzekerden hier schriftelijk melding van doen door hun verzekerden een erratum bij de polis te zenden of hiertoe een mededeling te doen via hun websites. Een tussentijdse wijziging ten nadele van de verzekerde die niet voortvloeit uit gewijzigde wet- of regelgeving betekent een opzegmogelijkheid voor de verzekerde. Het tussentijds stellen van beperkende voorwaarden in de polis, zoals de aanspraak beperken tot bepaalde zorgaanbieders of het introduceren van een toegangsvoorwaarde, in reactie op een standpunt over 'de stand van de wetenschap en praktijk', is niet aan te merken als 'voortvloeiend uit gewijzigde wet- of regelgeving', en vormt dus een opzegmogelijkheid.

6.3.2 Contractering van zorg

Het kan zich voordoen dat zorg onder de dekking is gaan vallen (omdat is vastgesteld dat de zorg voldoet aan 'de stand van de wetenschap en praktijk'), terwijl de zorgverzekeraar die naturapolissen aanbiedt, de zorg niet heeft ingekocht. De zorgverzekeraar zal daar dan alsnog in moeten voorzien. Ditzelfde geldt voor zorgverzekeraars die restitutiepolissen aanbieden met gecontracteerde zorgaanbieders. Overigens

¹³³ Voor medisch specialistische zorg kan als redelijk gelden dat de DBC die voor de datum van inwerkingtreding van het standpunt is geopend, mag worden afgemaakt.

is het zo dat zorgverzekeraars niet verplicht zijn om met alle zorgaanbieders die de betreffende zorgvorm leveren, contracten te sluiten. Zorgverzekeraars hebben contracteervrijheid.

Als een standpunt duidelijk maakt dat bepaalde zorg niet (meer) behoort tot het basispakket, houdt dit niet in dat deze zorg niet (meer) verleend zou mogen worden door zorgaanbieders. Een standpunt geeft slechts aan of de betreffende zorg al dan niet tot de basisverzekering behoort en of de zorg al dan niet laste van de basisverzekering vergoed mag worden.

Het feit dat een behandeling niet langer voor vergoeding in aanmerking komt ten laste van de basisverzekering, kan er wel toe leiden dat de verzekerde en zijn behandelaar in nauw overleg besluiten de behandeling te beëindigen dan wel af te bouwen. Uiteraard zal dat dan op zorgvuldige en medisch verantwoorde wijze moeten gebeuren. Ook zorgverzekeraars hebben er belang bij dat het proces van beëindiging of afbouw als gevolg van het staken van de vergoeding voor hun verzekerden zorgvuldig verloopt. Het ligt dan ook in de rede dat zij in voorkomend geval bij deze besluitvorming betrokken zijn.

6.3.3 Prestatiebeschrijving/tarieven NZa

Zorg die onder de dekking van de Zvw of de Wlz valt, moet kunnen worden gedeclareerd en betaald. De NZa stelt daarvoor de ingevolge de Wet marktordening gezondheidszorg (Wmg) vereiste prestatiebeschrijvingen en tarieven vast. Voor Zvw-zorg waarvan de tarieven vrij onderhandelbaar zijn (zoals het B-segment van de medisch specialistische zorg en fysiotherapie), stelt de NZa alleen prestatiebeschrijvingen vast (bijvoorbeeld in de vorm van DBC's als het gaat om medisch specialistische zorg). Voor zorg die qua tarieven niet vrij onderhandelbaar is (bijvoorbeeld het A-segment van de medisch specialistische zorg en huisartsenzorg), stelt de NZa ook de tarieven vast.

Het kan zich voordoen dat bepaalde zorg onder de basisverzekering is gaan vallen (omdat de zorg voldoet aan 'de stand van de wetenschap en praktijk'), maar dat een NZa-prestatiebeschrijving en – afhankelijk van de situatie – ook een NZa-tarief ontbreken. In deze zaken moet dan zo snel mogelijk worden voorzien. Om dat te bespoedigen treden wij al tijdens het traject van voorbereiding van het standpunt in contact met de NZa. Het is vervolgens aan de zorgaanbieders en de zorgverzekeraars om de Wmg-kwesties verder met de NZa te regelen.

Vaststelling rapport

De Raad van bestuur van Zorginstituut Nederland heeft dit rapport op 19 januari 2015 vastgesteld. Voor de opstelling van het rapport is mede gebruik gemaakt van het commentaar op conceptversies van WAR, ACP en externe partijen.

Het rapport komt in de plaats van het rapport “Beoordeling de stand van de wetenschap en praktijk” van 2007.⁸

Wij hebben het rapport toegezonden aan de minister van VWS met het verzoek om met het in het rapport opgenomen beoordelingskader in te stemmen.

Zorginstituut Nederland

Plv. Voorzitter Raad van Bestuur

dr. A. Boer

Literatuurlijst

AGREE Next Steps Consortium. AGREE II-Instrument: Appraisal Instrument for Guidelines, Research and Evaluation. Update 2013.

Ali U, Boermeester M. Beoordelen van bewijs voor nieuwe behandelingen. Ned Tijdschr Geneeskd 2014;158:A7281.

Boluyt N, Rottier BL, Langendam MW. Richtlijnen worden transparanter met GRADE-methode. Ned Tijdschr Geneeskd. 2012;156:A4379.

Commissie Keuzen in de Zorg (Commissie Dunning). Kiezen en delen. Rapport. Den Haag, 1991.

CRvB 19 januari 2006, RZA 2006/80.

CVZ. Beleidsregel Toetsingskader kwaliteitsstandaarden en meetinstrumenten. Staatscourant 2014, nr. 9953.

CVZ. Beoordeling stand van de wetenschap en praktijk. Diemen, 2007. Rapportnr. 254.

CVZ. Beoordelingskader hulpmiddelenzorg. Diemen, 2008. Rapportnr. 258.

CVZ. Leidraad Kwaliteitsstandaarden. Diemen, 2015.

CVZ. Medische tests. Beoordeling stand van de wetenschap en praktijk. Diemen, 2011. Rapportnr. 293.

CVZ. Onderwerpen Pakketbeheer 2014. Een risicogerichte selectie. Diemen, 2014.

CVZ. Pakketbeheer in de praktijk deel 3. Diemen, 2013.

CVZ. Pakketbeheer specialistische geneesmiddelen. Diemen, 2013.

CVZ. Reglement Wetenschappelijke Adviesraad. Diemen, 2013.

CVZ, Ministerie van VWS. Procedure beoordeling extramurale geneesmiddelen. Diemen, 2012.

Dijen M van. Bevordering van kwaliteit in de zorg. De zorgtaken van het Zorginstituut Nederland. PS Documenta 2013; 3: 221-36.

Du Bois RM, Weycker D, Albera C, et al. Six-minute walk test in idiopathic pulmonary fibrosis: test validation and minimal clinically important difference. Am J Respir Crit Care Med 2011;183:1231-7.

EMA. Guideline on the evaluation of medicinal products in the treatment of primary osteoporosis. London, 2006.

EMGO+, Maastricht University. Verkenning effectmaat voor de care sector. Rapportage in opdracht van het College voor Zorgverzekeringen (CVZ). Amsterdam, 2009.

Everdingen JJE van, Dreesens DHH, Burgers JS. Handboek evidence-based richtlijnontwikkeling. ze dr. Houten: Bohn Stafleu van Loghum, 2014.

EUnetHTA. Guidelines: Endpoints used for relative effectiveness assessment of pharmaceuticals. Clinical endpoints. Composite endpoints. Surrogate endpoints. Comparators and comparisons: direct and indirect comparisons. Health-related quality of life and utility measures. February 2013.

EUnetHTA. JA WP5: Relative effectiveness assessment (REA) of pharmaceuticals, July 2011.

Glasziou P, Chalmers I, Rawlins M, et al. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;334:349-51.

Groot GRJ de. De stand van de wetenschap en praktijk. *Tijdschr Gezondheidsrecht* 2006;30: 326-50.

Heymans J, Kleijnen S, Verstijnen I. Passend bewijs bij het bepalen van effectiviteit. *Ned Tijdschr Geneesk* 2013;157:A5479.

Kamerstukken II 2011/12, 33243, nr. 3

Korevaar DA, Bossuyt PMM, Hooft L. Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports. *BMJ Open* 2014;4:e004596.

Kwaliteitsinstituut voor de Gezondheidszorg CBO. Evidence-based Richtlijnontwikkeling. Handleiding voor werkgroepleden. Utrecht, 2007.

Kwok BC, Pua YH, Mamun K, et al. The minimal clinically important difference of six-minute walk in Asian older adults. *BMC Geriatrics* 2013;13:23.

Lachmann R, Schoser B. The clinical relevance of outcomes used in late-onset Pompe disease: can we do better? *Orphanet J Rare Dis* 2013;8:160.

Mathai SC, Puhan MA, Lam D, et al. The minimal important difference in the 6-minute walk test for patients with pulmonary arterial hypertension. *Am J Respir Crit Care Med* 2012;186: 428-33.

Millum J, Grady C. The ethics of placebo-controlled trials: methodological justifications. *Contemp Clin Trials* 2013;36:510-4.

Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med* 2002;347:81-8.

NHG-Standaard Osteoporose. *Huisarts Wet* 2005;48:559-70.

NICE. Developing NICE guidelines: the manual. 2014. Beschikbaar via <http://www.nice.org.uk/article/pmg20/resources/non-guidance-developing-nice-guidelines-the-manual-pdf>.

Nota van Toelichting bij het Besluit zorgverzekering (Stb. 2005, 389)

Puhan MA, Schünemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630.

Raad voor de Volksgezondheid en Zorg. Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid. Den Haag: Centrum voor ethiek en gezondheid, 2007. Signalering ethiek en gezondheid 2007/4.

Sackett DL, Rosenberg WMC, Gray JAM, et al. Evidence based medicine: what it is and what it isn't. *BMJ* 1996; 312:71-2.

Scholten RJPM, Offringa M, Assendelft WJJ van. Inleiding in evidence-based medicine. 4e dr. Houten: Bohn Stafleu van Loghum, 2014.

Schünemann H, Hill S, Guyatt G, et al. The GRADE approach and Bradford Hill's criteria for causation. *J Epidemiol Community Health* 2011;65:392-5.

Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med* 2013;369:2515-24.

Stang A, Hense HW, Jöckel KH, et al. Is it always unethical to use a placebo in a clinical trial? *PLoS Med* 2005;2:177-80.

Tulder van, M, Malmivaara A, Hayden J, Koes B. Statistical significance versus clinical importance. *Spine* 2007;32:1785-90.

Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med* 2011;155:529-36.

World Medical Association. Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. 2013.

Bijlage

Passendbewijsvragenlijst

NB. In deze tweede versie zijn er enkele wijzigingen ten opzichte van de eerder in 2013 gepubliceerde versie in het NTVG

Referentie

Heymans JM, Kleijnen S, & Verstijnen IM (2013) Passend bewijs bij het bepalen van effectiviteit van interventies. Ned Tijdschr Geneeskd. 2013;157:A5479.

Pakketbeheervragen

VRAAG 1: Eerder beoordeeld; variant op interventie?

Betreft de voorliggende interventie slechts een kleine aanpassing aan de gangbare interventie die reeds tot de StW&P behoort (én al beoordeeld is voor dezelfde indicatie)?

Wanneer de voorliggende behandeling een kleine aanpassing betreft aan een reeds tot de StW&P behorende, gangbare behandeling is het hoogste niveau van bewijs niet altijd noodzakelijk of is een beoordeling van scratch niet noodzakelijk. Chirurgische technieken boeken over het algemeen vooruitgang door een aantal opeenvolgende, kleine aanpassingen aan behandelingen (6).

Een voorbeeld van een interventie die vooruitgang boekt door een aantal opeenvolgende, kleine aanpassingen aan een reeds tot de StW&P behorende behandeling, is autologe kraakbeencelimplantatie in de knie (7). In eerste instantie werden losse chondrocyten geïmplantéerd via een open artrotomie, en afgedekt door middel van een periostlap (ACI-P). Deze ACI-techniek had echter technische en klinische nadelen. Daarom werd ACI-P verder ontwikkeld en ontstond ACI-C, waarbij in plaats van een periostlap een afdekking van collageen werd gebruikt. Vervolgens ging de ontwikkeling nog weer verder waarbij de chondrocyten op een vooraf op maat gesneden collageen-matrix worden gedrupeerd (bijvoorbeeld ChondroCelect) (ACT-C). Hierdoor zou een optimaler gebruik van de cellen mogelijk zijn en het celverlies gereduceerd worden. Een andere gelijktijdige ontwikkeling was dat de chondrocyten gekweekt werden in een driedimensionale matrix (M-ACI). De M-ACI techniek heeft weer als voordeel dat het door middel van een arthroscopie of miniartrotomie kan worden uitgevoerd. De opeenvolgende aanpassingen waren echter steeds op een zelfde pathofysiologisch werkingsmechanisme gebaseerd.

Ja, een vergelijkbare interventie is al eerder beoordeeld, de resultaten zijn daarom waarschijnlijk extrapoléerbaar naar de huidige interventie. De argumentatie in de beoordeling kan daarom in principe gebaseerd worden op de oude beoordeling (tenzij er nieuwe evidence is).

Nee, er is geen sprake van een kleine aanpassing aan een, reeds tot de StW&P behorende gangbare behandeling, daarom moet de beoordeling van scratch gedaan worden (alsof er geen eerdere beoordeling is geweest) en daarom is net als bij andere beoordelingen in principe het hoogste niveau van bewijs gewenst, dus dat is een reden om (zo mogelijk) **gerandomiseerde behandeltoewijzing** te wensen én eventuele **blinding van de effectbeoordelaar(s), patiënten en behandelaars**. Een gerandomiseerde behandeltoewijzing impliceert de aanwezigheid van een controlegroep.

Weet niet

VRAAG 2: Eerder beoordeeld; zelfde interventie andere/smaller indicatie?

Is de interventie voor een andere indicatie positief beoordeeld?

Wanneer de voorliggende interventie al positief beoordeeld is voor een nauwere indicatie, is het hoogste niveau van bewijs niet altijd noodzakelijk, onder voorwaarde dat het zeer aannemelijk is dat de werking berust op grond van een zelfde pathofysiologisch/ farmacologisch werkingsmechanisme.

Een voorbeeld uit de farmaceutische zorg is cholesterol-verlagers. Het effect van cholesterol-verlagers is aangetoond bij jonge mannen. De vervolgvraag is of dit bewijs geëxtrapoleerd kan worden naar oudere vrouwen, met andere woorden spelen leeftijd en geslacht een mogelijke rol in de pathofysiologische verklaring van de werking van de interventie? Als leeftijd en geslacht géén rol spelen dan gaat de pathofysiologische verklaring ook volledig op in het 'nieuwe' ruimere indicatiegebied met oudere vrouwen. In dat geval volstaat argumentatie ten aanzien

van de extrapolatie van het ‘oude’ bewijs. Wanneer in het voorbeeld leeftijd wél een rol speelt in de pathofysiologische verklaring of wanneer het bij voorbaat niet kan worden uitgesloten dat leeftijd géén rol speelt, dan is er geen reden om klinisch bewijs niet relevant te achten en lijkt voor deze interventie-indicatie combinatie het hoogste niveau van bewijs gewenst, i.e. een gerandomiseerde toewijzing van patiënten aan de verschillende armen van het onderzoek met eventuele blinding van de effectbeoordelaar(s), patiënten en behandelaars.

Ja, deze interventie is eerder beoordeeld voor een andere/smaller indicatie. Kies je het JA antwoord dan verschijnt een extra tabblad (P-VRAAG2b) die ingaat op de **vervolgvraag** over extrapoleerbaarheid/indicatieverruiming.

Nee, deze interventie is niet eerder voor een andere indicatie positief beoordeeld, daarom moet de beoordeling van scratch gedaan worden (alsof er geen eerdere beoordeling is geweest) en daarom is net als bij andere beoordelingen in principe het hoogste niveau van bewijs gewenst, dus dat is een reden om (zo mogelijk) **gerandomiseerde behandeltoewijzing** te wensen én eventuele **blinding van de effectbeoordelaar(s), patiënten en behandelaars**. Een gerandomiseerde behandeltoewijzing impliceert de aanwezigheid van een controlegroep.

Weet niet

VRAAG 2b: Extrapolatie/Indicatieverruiming?

Zijn de resultaten voor de eerder beoordeelde smallere indicatie (vraag 2) extrapoleerbaar naar de indicatie die momenteel voorligt? (indicatieverruiming/indicatieshift)

Wanneer de voorliggende interventie al positief beoordeeld is voor een nauwere indicatie, is het hoogste niveau van bewijs niet altijd noodzakelijk of is een beoordeling van scratch niet altijd noodzakelijk, onder voorwaarde dat het zeer aannemelijk is dat de werking berust op grond van een zelfde pathofysiologisch werkingsmechanisme.

Ja, de pathofysiologische verklaring gaat ook op in het nieuwe indicatiegebied, de resultaten zijn daarom waarschijnlijk extrapoleerbaar naar het ruimere indicatie gebied. De argumentatie in de beoordeling kan daarom in principe gebaseerd worden op de oude beoordeling (tenzij er nieuwe evidence is).

Nee, er zijn zodanige cruciale verschillen tussen ‘het oude’ en ‘het nieuwe’ indicatiegebied dat betwijfeld kan worden of de pathofysiologische verklaring ook op gaat in het nieuwe indicatiegebied, daarom moet de beoordeling van scratch gedaan worden (alsof er geen eerdere beoordeling is geweest) en is net als bij andere beoordelingen in principe het hoogste niveau van bewijs gewenst, dus dat is een reden om **gerandomiseerde behandeltoewijzing** te wensen én eventuele **blinding van de effectbeoordelaar(s), patiënten en behandelaars**. Een gerandomiseerde behandeltoewijzing impliceert de aanwezigheid van een controlegroep.

Weet niet

Zorginhoudelijke vragen

VRAAG 1: Natuurlijk beloop?

Verminderen bij dit ziektebeeld de klachten na verloop van tijd vanzelf (natuurlijk beloop)?

Wanneer de klachten na verloop van tijd vanzelf verminderen (vanwege het natuurlijke beloop), weet je zonder controlegroep niet of na een interventie de verbetering toe te schrijven is aan de interventie of aan het natuurlijke beloop van het ziektebeeld (of aan beiden). Omdat bij de controlegroep alleen het natuurlijke beloop plaatsvindt (in geval van placebo), krijg je alléén door een vergelijking tussen de controlegroep en de interventiegroep inzicht in het daadwerkelijke effect van de interventie. Niet-gecontroleerde studies hebben de neiging de effectiviteit van een behandeling te overschatten, onder andere omdat zij geen rekening houden met het gunstige, natuurlijke beloop van de aandoening (2). Een voorbeeld is vertebroplastiek (VP) en ballonkyphoplastiek (BK) voor de behandeling van compressie wervelfracturen die klachten geven. Klachten, zoals pijn en vermindering van mobiliteit

en functioneren als gevolg van deze pijn, verdwijnen bij de meeste patiënten (grotendeels) na verloop van tijd (gemiddeld binnen 1-3 maanden) spontaan (3).

Ja, er is sprake van een ziektebeeld waarbij de klachten na verloop van tijd vanzelf verminderen, dus dat is een reden om een **controlegroep** te wensen.

Nee, er is geen sprake van een ziektebeeld waarbij de klachten na verloop van tijd vanzelf verminderen.

Weet niet

VRAAG 2: Confounding by Indication?

Zijn er bekende patiënt- of ziektekenmerken die de prognose van de patiënt beïnvloeden (risico van confounding-by-indication)?

Wanneer een behandelaar bewust of onbewust geneigd is om de interventie van zijn voorkeur toe te wijzen aan patiënten met een relatief gunstige prognose, of juist aan de meest ernstig zieke patiënten, dan is er sprake van confounding by indication (4). Een behandelaar zou kennis over deze kenmerken mee kunnen laten spelen bij zijn keuze voor een behandeling. Deze selectie 'aan de poort' kan worden voorkomen door het random (a-select) toewijzen van patiënten aan de verschillende armen van het onderzoek.

Een voorbeeld waaruit de noodzaak van randomiseren blijkt, is het aantonen van het effect van zogenaamde "drug-eluting stents" (DES). DES reduceren het risico van restenose in vergelijking met zogenaamde "bare metal stents" (BMS). Een aantal observationele studies toonde een reductie aan in mortaliteit en niet-fataal myocardinfarct, terwijl deze resultaten in gerandomiseerde studies niet werden waargenomen. In de observationele studies werd het effect overschat, omdat de DES werden geplaatst bij patiënten met een relatief gunstige prognose. Met andere woorden, patiënt- of ziektekenmerken die de toewijzing van het type stent beïnvloeden, lijken de mortaliteitswinst van DES in observationele studies te verklaren (5).

Ja, er zijn bekende patiënt- of ziektekenmerken die de prognose van de patiënt beïnvloeden (risico van confounding-by-indication), dus dat is een reden om **gerandomiseerde behandeltoewijzing** te wensen. Een gerandomiseerde behandeltoewijzing impliceert de aanwezigheid van een **controlegroep**. Verder zou je de **effectbeoordelaar** (zo mogelijk) willen **blinderen** **tav** **patientkenmerken**, omdat verwachtingen dan geen rol spelen bij de beoordeling van het resultaat. Zo mogelijk wil je ook de **behandelaar blinderen**.

Nee, er zijn geen bekende patiënt- of ziektekenmerken die de prognose van de patiënt beïnvloeden (risico van confounding-by-indication), dus dat is geen reden om gerandomiseerde behandeltoewijzing te wensen.

Weet niet

VRAAG 3a: Plausibiliteit pathofysiologische verklaring

Is er sprake van een plausibel causaal verband tussen de pathofysiologie van de aandoening en de werking van de interventie?

Een duidelijk en plausibel causaal verband tussen de pathofysiologie van de aandoening en de werking van de interventie wordt meestal aangetoond dmv (langdurig) nauwkeurig onderzoek met rigide onderzoeksdesigns. Dit in tegenstelling tot vage pathofysiologische verklaringen die vaak het product zijn van toevallige observaties. Bij vage pathofysiologische verklaringen is het onduidelijk of de interventie de veroorzaker is van het effect. Anders gezegd, het is onduidelijk of er sprake is van een causaal verband of dat er alleen sprake is van coincidentie. Deze situatie vraagt om een meer scrupuleuze onderzoekszet waarin een eventueel causaal verband kan worden aangetoond en die zoveel mogelijk vrij is van de invloed van verwachtingen. Een voorbeeld komt uit autisme onderzoek. Op basis van observaties en onderzoek zonder controlegroep (ref1) ontstond het idee dat het MMR vaccin (mazelen, bof, rode hond), een darmontsteking zou kunnen veroorzaken die er voor zou zorgen dat normaliter niet permeabele peptiden toegang krijgen tot het bloed en de hersenen. Dit zou tot autisme leiden. Doordat er geen controle groep was kon niet vastgesteld worden of er een causal verband was of dat er sprake was coincidentie. Hoop en de wens om de oorzaak van autisme te vinden, vergrootte de bereidheid om de observaties causaal te interpreteren. (ref2)

Ja, Wanneer er sprake is van een duidelijk en plausibel causaal verband tussen de pathofysiologie van de aandoening en de werking van de interventie. Je krijgt een **vervolgvraag** over het effect als je JA antwoordt op deze plausibiliteitsvraag.

Nee, Het pathofysiologische werkingsmechanisme is niet heel plausibel/logisch/sluitend en dat vormt een (extra) reden om **gerandomiseerde behandeltoewijzing** te wensen én zo mogelijk **blinding** van de **effectbeoordelaar(s)**, **patiënten** en **behandelaars**. Een gerandomiseerde behandeltoewijzing impliceert de aanwezigheid van een controlegroep.

Weet niet

VRAAG 3b: Grootte/Directheid effect?

Is er sprake van een vrijwel alles-of-niets effect en/of een direct effect?

Wanneer er sprake is van een duidelijk en plausibel causaal verband tussen de pathofysiologie van de aandoening en de werking van de interventie, EN dit verband wordt ondersteund door een vrijwel alles-of- niets effect of direct effect, dan is het hoogste niveau van bewijs niet altijd noodzakelijk. Bij afwezigheid van een plausibel causaal verband blijft het hoogste niveau van bewijs gewenst. Een plausibel causaal verband tussen de werking van de interventie en de pathofysiologie van de aandoening wordt ondersteund door een (positief) effect van de interventie dat direct op de toepassing volgt (bijvoorbeeld insuline suppletie bij diabetes) en/ of een 'alles-of-niets effect'. Aan het 'alles-of-niets effect' wordt volgens Sacket (8) voldaan wanneer bijvoorbeeld alle patiënten stierven voor het beschikbaar komen van de interventie, maar sommige patiënten nu overleven, of waarneer sommige patiënten stierven voor het beschikbaar komen van de interventie, maar nu allen overleven. De 'directheid' van het effect en het 'alles-of-niets effect' verkleinen de kans dat het effect in zijn geheel veroorzaakt wordt door (onbekende) confounders (9). "Directheid" en "grootte" van het effect leiden er tezamen met de aanwezigheid van een plausibel causaal verband tussen de pathofysiologie van de aandoening en de werking van de interventie toe dat het vertrouwen in de effectiviteit van de interventie toeneemt, waardoor de wenselijkheid van gerandomiseerde behandeltoewijzing, evenals eventuele blinding van de effectbeoordelaar(s), patiënten en behandelaars afneemt.

Ja, er is sprake van een plausibele pathofysiologisch werkingsmechanisme en bovendien van een alles-of-niets effect of een direct effect. Dit wordt niet zichtbaar in de figuur hiernaast (want dit zijn argumenten vóór), maar wel op het tabblad WENSPROFIEL zie je nu (in het rood) het tegenargument Plausibiliteit&Grootte/Directheid van het effect verschijnen.

Nee, er is weliswaar sprake van een plausibele pathofysiologisch werkingsmechanisme maar er is geen sprake van een alles-of-niets effect noch van een direct effect. De lat blijft waar die ligt.

Weet niet

VRAAG 4: Subjectieve uitkomstmaten

Wordt bij dit ziektebeeld een mogelijke uitkomst (de O in de PICO) van de interventie gemeten door middel van subjectieve uitkomstmaten?

NB bedenk dat je voor iedere PICO en dus voor iedere uitkomst deze vragenlijst opnieuw moet invullen

Wanneer een mogelijk effect van de interventie wordt gemeten met subjectieve uitkomstmaten, kan dit overschatting van de effectiviteit versterken (2). Effectbeoordelaars worden door middel van blinding onwetend gehouden over de toegewezen interventie, waardoor zij hun enthousiasme/ antipathie voor een interventie niet (onbewust) kunnen laten meespelen bij de bepaling van het effect van de interventie. Blinding van de effectbeoordelaar(s) is bij subjectieve uitkomstmaten daarom gewenst. Voorbeelden van subjectieve uitkomstmaten zijn pijn en ongemak van een interventie. Bij dit soort subjectieve uitkomsten is de effectbeoordelaar de patiënt zelf. Niet alleen de patiënt, maar ook onafhankelijke effectbeoordelaars kunnen te maken hebben met subjectieve inschattingen van het effect, zoals de interpretatie van röntgenfoto's. Zonder controlegroep en blinding van de effectbeoordelaar(s) kan het effect van de interventie vertekend zijn. Bedenk dat je bij GRADE meerdere uitkomstmaten kan beoordelen, bij de subjectieve daarvan moet je dus specifiek opletten of er blinding van de effectbeoordelaar(s) tav de toegewezen interventie heeft plaatsgevonden.

Ja, er is sprake van (onder andere) subjectieve uitkomstmaten, dus dat is een reden om een **controlegroep** te wensen is en waarbij (bij de subjectieve maten) **blinding van de**

effectbeoordelaar(s) heeft plaatsgevonden. N.B. Een controlegroep impliceert niet per sé randomisatie, je kan bijvoorbeeld een historische controle groep hebben of parallelle cohorten.

Nee, er is geen sprake van subjectieve maten, alleen van objectieve maten.

Weet niet

Contextvragen

VRAAG 1: Patiëntvoorkeur

Hebben de patiënten een uitgesproken voorkeur voor de interventie (of voor de controlebehandeling) die de navolging van het protocol kan beïnvloeden?

Wanneer het enthousiasme voor de interventie (of de controlebehandeling) van patiënten de navolging van het protocol beïnvloeden, dan kan er een vertekening in de onderzoeksresultaten optreden. Factoren die de voorkeur van patiënten voor een interventie kunnen beïnvloeden zijn onder andere belasting, potentieel risico van de interventie en vertrouwdsheid met de interventie (13). "Door blinding van de patiënt wordt voorkomen dat hij - omdat hij op de hoogte is van de aard van de toegewezen behandeling - bewust of onbewust een grotere navolging van het protocol zal hebben" (10). Blinding is dus gewenst, en, zeker als blinderen niet mogelijk is, is een controle groep wenselijk om differential drop-out te detecteren (dat er in de arm van onderzoek waar men niet in gelooft meer patiënten uitvallen), bovendien is randomisatie wenselijk om te zorgen dat die twee armen in principe gelijkwaardig bedield zijn qua mensen met een bepaalde voorkeur. Bovendien wil je, zeker als patiënten blinderen niet mogelijk is, de effectbeoordelaar blinderen ten aanzien van de twee armen van onderzoek. Nieuw gerandomiseerd onderzoek is vaak met dergelijke 'onwillige' patiënten niet mogelijk, dus houdt er bij je beoordeling rekening mee dat er waarschijnlijk geen nieuwe evidence zal komen.

Ja, de voorkeur van de patiënten voor de interventie of de controlebehandeling kan de navolging van het protocol beïnvloeden, dus dat is een (extra) reden om **blinding van de patiënt, blinding van de effectbeoordelaar, randomisatie** en een **controlegroep** te wensen.

Nee, de patiënten hebben geen voorkeur.

Weet niet

VRAAG 2: Behandelaarsvoorkeur

Hebben de behandelaars een uitgesproken voorkeur voor de interventie (of voor de controlebehandeling) die de navolging van het protocol kan beïnvloeden?

Wanneer het enthousiasme voor de interventie (of de controlebehandeling) van de behandelaars de navolging van het protocol beïnvloeden, dan kan er een vertekening in de onderzoeksresultaten optreden. Door blinding van de behandelaar wordt voorkomen dat hij omdat hij op de hoogte is van de aard van de toegewezen behandeling - bewust of onbewust een andere navolging van het protocol zal hebben. Het zou zo kunnen zijn dat artsen door hun zorgen over één van de behandelingen additionele medicatie voorschrijven. Ook wordt door blinding van de behandelaar voorkomen dat hij een bepaald enthousiasme zal uitstralen (selectieve vergroting van het placebo-effect) (10).

Een voorbeeld dat het belang van blinding benadrukt is injectie- en denervatietechniek bij patiënten met chronische aspecifieke lage rugpijn. De beroepsgroep van anesthesiologen steunt het uitvoeren van injectie- en denervatietechnieken (11). Door het enthousiasme voor de injectie- en denervatietechnieken is er zonder blinding van de behandelaar een grote kans op een vertekening van de effectiviteit.

Ja, de voorkeur van de behandelaars voor de interventie of de controlebehandeling kan de navolging van het protocol beïnvloeden, dus dat is een reden om **blinding van de behandelaar, blinding van de effectbeoordelaar, randomisatie** en een **controlegroep** te wensen.

Nee, de behandelaars hebben geen voorkeur.

Weet niet

Haalbaarheidsvragen

VRAAG 1: Langdurigheid Uitkomsten

Is er sprake van uitkomsten in de verre toekomst en zijn intermediaire uitkomsten niet predictief?

Geblindeerd onderzoek waarbij de behandelaars en/of patiënt niet we(e)t(en) of de patiënt aan de interventie of aan de vergelijkende behandeling is blootgesteld, leent zich over het algemeen niet voor lange termijn onderzoek. Op den duur moeten behandelaars/ onderzoekers en de patiënt vaak weten aan welke interventie de patiënt is blootgesteld, bijvoorbeeld vanwege vervolgbehandelingen of om de oorzaak van mogelijke bijwerkingen te achterhalen, en moet de blinding worden opgeheven. Indien er intermediaire uitkomstmaten beschikbaar zijn, kan blinding in stand gehouden worden tot deze uitkomstmeting.

Ja, er is sprake van uitkomsten in de verre toekomst en/of er is sprake van een langdurige interventie (en intermediaire uitkomsten bieden geen oplossing). **Blinden behandelaar en patiënten** is dan vaak onmogelijk.

Nee, er is geen sprake van uitkomsten in de verre toekomst of van een langdurige interventie, intermediaire of directe uitkomsten zijn mogelijk.

Weet niet

VRAAG 2: Patiënt Equipose

Heeft de interventie of de controlebehandeling een zodanige reputatie dat patiënten niet mee willen werken aan randomiseren (patiënt equipose)?

Soms willen patiënten niet meewerken aan onderzoek omdat de nieuwe interventie (of de oude) een zodanige goede reputatie heeft, dat de haalbaarheid van het rekruteren en randomiseren van patiënten in gevaar komt (12). Factoren die de voorkeur van patiënten voor een interventie kunnen beïnvloeden zijn onder andere belasting, risico van de interventie en vertrouwde met de interventie (Ref2). Nieuw gerandomiseerd onderzoek is met een dergelijke 'onwillige' patiënten vaak niet mogelijk. Wél kunnen vroeger patiënten behandeld zijn geweest met de nu 'niet-favoriete' interventie (historische controlegroep) of kunnen ze hun eigen controle zijn.

Een voorbeeld. Bij een grootschalige studie onder vrouwen met een verhoogd risico op borstkanker vanwege een familiale of genetische predispositie bleek dat randomiseren niet mogelijk was door gebrek aan patiënt equipose. In de studie werd onderzocht of MRI als screeningsinstrument effectiever zou zijn dan de standaard mammografie. Bij het opzetten van de studie bleek dat randomiseren niet haalbaar was: reeds eerder opgezette, kleinschalige, diagnostische studies lieten inmiddels veelbelovende resultaten voor de MRI zien. Vrouwen wilden zich om deze reden niet laten randomiseren, mede omdat in een aantal centra inmiddels de mogelijkheid bestond voor het krijgen van een MRI. Ook artsen uit deelnemende centra waren voor de start van de studie van mening dat randomiseren geen kans van slagen zou hebben. In de studie is toen gekozen voor 'de patiënt als eigen controle'. Er werd niet gerandomiseerd, deelnemers aan de studie kregen zowel een MRI als een mammografie.

Ja, er is sprake van een dusdanig favoriete behandeling dat patiënten niet willen meewerken aan **randomisatie** omdat ze daarmee het risico lopen te worden ingedeeld in de niet favoriete **controlegroep**. Wanneer een historische controlegroep al wél bestaat, of wanneer een patiënt als zijn eigen controle kan dienen, pas dan het antwoord bij Controlegroep nog aan naar Moeilijk/Mogelijk.

Nee, er is geen sprake van een dusdanig favoriete behandeling, dat randomiseren (en daarmee het inrichten van een Controlegroep) moeilijk of onmogelijk is.

Weet niet

VRAAG 3: Klinische Equipose

Heeft de interventie of de controlebehandeling een zodanige reputatie dat behandelaren niet mee willen werken aan randomiseren (klinische equipose)?

Soms willen behandelaren niet meewerken aan onderzoek omdat de nieuwe interventie een zodanig goede reputatie heeft dat men het niet meer verantwoord vindt nog een andere (vaak oude) interventie toe te passen of een placebo. Factoren die de voorkeur van behandelaren voor een interventie kunnen beïnvloeden zijn onder

andere belasting, risico van de interventie en vertrouwde met de interventie (ref1). Een historisch controlegroep waarin de alternatieve (oude) behandeling is toegepast bestaat vaak wel. Een voorbeeld. In het CVZ-advies van 2007 (ref2) werd de nieuwe Metal-on-metal Hip Resurfacing Arthroplasty (MoM-HRA) vergeleken met de oudere behandeling Total Hip Replacement (TRT). Het behandelrisico van de twee interventies liep volgens de behandelaren zodanig uiteen dat nieuw onderzoek waarin MoM direct met TRT zou worden vergeleken en waarin dus nog TRT toegepast zou worden binnen één van de twee onderzoekarmen, niet meer verantwoord was. TRT werd echter vóór de introductie van MoM-HRA toegepast dus een historische controle groep bestond al wel.

Ja, er is sprake van een dusdanig favoriete behandeling dat behandelaren geen controlegroep meer willen inrichten. behandeling aan **randomisatie** omdat ze daarmee het risico lopen te worden ingedeeld in de niet favoriete **controlegroep**. Wanneer een historische controlegroep al wél bestaat, of wanneer een patiënt als zijn eigen controle kan dienen, pas dan het antwoord bij Controlegroep nog aan naar Moeilijk/Mogelijk.

Nee, er is geen sprake van een dusdanig favoriete behandeling, dat randomiseren (en daarmee het inrichten van een Controlegroep) moeilijk of onmogelijk is.

Weet niet

VRAAG 4: Grootte Indicatiegebied

Is er sprake van een zeer klein indicatiegebied (P in PICO)?

Een zeer klein indicatiegebied bemoeilijkt het verkrijgen van statistisch betrouwbare resultaten ten aanzien van de effectiviteit m.b.v. gerandomiseerd (geblindeerd) onderzoek. Om te bepalen of er sprake is van een zeer klein indicatiegebied moet men niet alleen denken aan de omvang van de patiëntenpopulatie in Nederland. Vaak is het combineren van internationale studies mogelijk**. Er zijn echter situaties denkbaar waarbij internationale verschillen in behandelpraktijken te ver uiteenlopen om grensoverschrijdend onderzoek te verrichten, zoals ongestandaardiseerde chirurgische procedures. Anderzijds kan de patiënt in sommige gevallen zijn eigen controle zijn. Een voorbeeld. Naast de beschikbaarheid van verschillende studies (van lager niveau van bewijs) was het zeer kleine indicatiegebied bij intra-oculaire tumoren, chordomen/ chondrosarcomen en bij pediatrische tumoren een reden om niet het hoogste niveau van bewijs te wensen bij het aantonen van de StW&P van protonentherapie.

Ja, er is sprake van een zeer klein indicatiegebied, dus de haalbaarheid van **randomisatie** en daarmee **controlegroepen** is beperkt.

Nee, er is geen sprake van een zeer klein indicatiegebied

Weet niet

VRAAG 5: Rug-tegen-de-muur

Is de interventie de laatste mogelijkheid voor uitbehandelde patiënten met een (zeer) slechte (fatale) prognose en/of zeer slechte kwaliteit van leven? (rug-tegen-de-muur)

Soms biedt een nieuwe interventie de enige hoop voor patiënten. Zonder de interventie hebben deze patiënten sowieso een zeer slechte (fatale) prognose en/of zeer slechte kwaliteit van leven. De nieuwe behandeling kan de enige mogelijkheid zijn voor patiënten met een contra-indicatie voor de standaardbehandeling, of voor patiënten bij wie andere behandelingen niet effectief bleken. We spreken dan van een rug-tegen-de-muur situatie. Iets doen geeft een kans op stabilisatie of een verbetering, niets doen geeft dat niet. Hierbij kan men denken aan een verbetering ten aanzien van overleving, maar de verbetering kan ook betrekking hebben op kwaliteit van leven.

Voor de patiënten die geen baat hebben bij de standaardbehandeling en voor wie niets anders resteert dan de betreffende interventie, kan beperkte bewijskracht ten aanzien van de effectiviteit acceptabel worden geacht. Vaak is blinding dan nog wel.

Ja, er is sprake van een rug-tegen-de-muur situatie, waarbij **randomisatie** en een **controlegroep** een deel van de patiënten de enige mogelijkheid tot stabilisatie of verbetering zal onthouden.

Nee, er is geen sprake van een rug-tegen-de-muur situatie.

Weet niet

VRAAG 6: Urgentie

Bemoeilijkt de mate van urgentie van de interventie het uitvoeren van gerandomiseerd (geblindeerd) onderzoek?

Sommige interventies dienen met zoveel spoed te worden uitgevoerd, waardoor urgent ingrijpen noodzakelijk is. De urgentie van een interventie kan het randomiseren van patiënten bemoeilijken, inclusief het verkrijgen van een informed consent (ref1). Een urgente interventie maakt gerandomiseerd (geblindeerd) onderzoek echter niet noodzakelijkerwijs ongeschikt. Zelfs bij een aneurysma van de buikaorta is gerandomiseerd onderzoek uitgevoerd. Als het aneurysma scheurt, is een spoedoperatie noodzakelijk. Recentelijk is er in de regio Amsterdam een trial uitgevoerd waarin de resultaten van een minimaal invasieve aanpak vergeleken werden met een grote acute buikoperatie (AJAX-trial). In een gebied met 1,3 miljoen inwoners werden 105 patiënten in twee jaar geïncludeerd (Ref2).

Ja, de mate van urgentie van de interventie bemoeilijkt het uitvoeren van gerandomiseerd (geblindeerd) onderzoek, dus de haalbaarheid van **randomiseren** en **blinderen van patiënten en behandelaars** wordt beperkt.

Nee, er is geen sprake van een urgente situatie

Weet niet

VRAAG 7: Multifactorieel karakter

Bemoeilijkt het multifactoriële karakter van de interventie het uitvoeren van gerandomiseerd (geblindeerd) onderzoek?

Gerandomiseerd (geblindeerd) onderzoek wordt bemoeilijkt wanneer de interventie een multifactorieel karakter heeft: de interventie bestaat uit verschillende componenten, en deze kunnen los van elkaar de uitkomst (zoals gedefinieerd in de O van de PICO) beïnvloeden. Sommige van deze componenten kunnen niet gestandaardiseerd worden. Dergelijke interventies worden ook wel complex genoemd. Net als bij de urgentie van een interventie is er geen sprake van een absolute onmogelijkheid van gerandomiseerd (geblindeerd) onderzoek, maar eerder van een barrière.

Een veelgebruikt voorbeeld van een complexe interventie is de behandeling van beroerte in de zogenaamde 'stroke unit'. Bij de evaluatie van een stroke unit door middel van gerandomiseerd (geblindeerd) onderzoek moet onder andere rekening worden gehouden met de expertise van verschillende hulpverleners en individueel medicijngebruik van patiënten (Ref1). Ook het karakter van een chirurgische interventie kan resulteren in beperkingen ten aanzien van het uitvoeren van gerandomiseerd (geblindeerd) onderzoek (Ref2). Ten eerste wordt chirurgie uitgevoerd door een team van chirurg(en), anesthesist(en), verpleegkundige(n) etc. De vaardigheid, ervaring en keuzes van al deze individuen hebben invloed op de uitkomst van de behandeling. Ook is het moeilijk om standaardisatie van patiëntenzorg voor en na de interventie te realiseren (Ref3).

Ja, de multifactoriële karakter bemoeilijkt het uitvoeren van gerandomiseerd (geblindeerd) onderzoek, dus de haalbaarheid van **randomiseren** en **blinderen van patiënten en behandelaars** wordt beperkt.

Nee, er is geen sprake van een multifactorieel karakter.

Weet niet

VRAAG 8: Eventueel nog aanvullende redenen die onderzoekskenmerken moeilijk of onmogelijk maken

Adresseer hier de andere aspecten die het uitvoeren van gerandomiseerd geblindeerd onderzoek bemoeilijken.

Colofon

Volgnummer

2014116583

Contactpersonen

mw. mr. P.C. Staal, afdeling Pakket, telefoon +31 (0)20 797 87 33

mw. J. Heymans, arts, afdeling Pakket, telefoon +31 (0)20 797 85 92

Auteurs

mw. mr. P.C. Staal

mw. J. Heymans, arts

mw. dr. G. Ligtenberg

mw. J.T.M. Derksen, gynaecoloog

mw. drs. B.T.L.E. Couwenbergh

Uitgebracht aan de minister van VWS

