

Handreiking voor het meten
van kwaliteit van zorg met

Patient Reported Outcome Measures



Handreiking voor het meten van kwaliteit van zorg met **Patient Reported Outcome Measures**

Paul van Kessel

Mattanja Triemstra

Dolf de Boer

OKTOBER 2014

**HANDREIKING VOOR HET METEN VAN KWALITEIT VAN ZORG MET
PATIENT REPORTED OUTCOME MEASURES**

*Dit document is een product van samenwerking tussen
het NIVEL en Zorginstituut Nederland*

AUTEURS

Paul van Kessel (NIVEL)
Mattanja Triemstra (NIVEL)
Dolf de Boer (NIVEL)

REDACTIE

Laura Koopman (Zorginstituut Nederland)
Kim Boerman (Zorginstituut Nederland)
Femke Vlems (Zorginstituut Nederland)
Dolf de Boer (NIVEL)
Paul van Kessel (NIVEL)
Mattanja Triemstra (NIVEL)

ISBN 978-94-6122-273-2

www.nivel.nl
nivel@nivel.nl
Telefoon (030) 2 729 700
Fax (030) 2 729 729

© 2014 NIVEL, Postbus 1568, 3500 BN Utrecht
Eerste druk oktober 2014
Vormgeving: Jan van Waarden/RAM vormgeving

*Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar
gemaakt worden door middel van druk, fotokopie, microfilm of op welke
andere wijze dan ook zonder voorafgaande schriftelijke toestemming van
het NIVEL te Utrecht. Het gebruik van cijfers en/of tekst als toelichting of
ondersteuning in artikelen, boeken en scripties is toegestaan, mits de
bron duidelijk wordt vermeld.*

Inhoud

Voorwoord	5
1 Inleiding	7
1.1 Achtergrond	7
1.2 Doel, positionering en doelgroep	8
1.3 Verschillende doelen van PROMs en prioritering	9
1.4 Opbouw en leeswijzer	10
2 PROMs: definiëring en stappenplan	11
2.1 PRO, PROM en kwaliteitsindicator	11
2.2 Generiek stappenplan en relatie toetsingscriteria	11
3 Selectie van PROs	14
3.1 Patiëntbetrokkenheid	14
3.2 Relatie met kwaliteit van zorg	15
3.3 Is een PRO wel geschikt?	15
4 Selectie van bestaande PROMs	17
4.1 Identificeren van bestaande PROMs	17
4.2 Evaluatie en selectie van een PROM	18
4.2.1 Conceptueel- en meetmodel	19
4.2.2 Inhoudsvaliditeit	19
4.2.3 Constructvaliditeit	21
4.2.4 Cross-culturele validiteit	21
4.2.5 Betrouwbaarheid	22
4.2.6 Responsiviteit	23
4.2.7 Hanteerbaarheid en interpreteerbaarheid	24
5 Besluitvorming over ontwikkeling of doorontwikkeling	27
5.1 Doorontwikkeling	27
5.2 Ontwikkelen van een nieuwe PROM	28

6	Vergelijkbaarheid en onderscheidend vermogen: van PROM naar indicator	31
6.1	Vergelijkbaarheid	31
6.1.1	Registratievergelijkbaarheid	32
6.1.2	Populatievergelijkbaarheid	33
6.1.3	Steekproef- en responsvergelijkbaarheid	34
6.2	Onderscheidend vermogen	35
6.2.1	Van PROM naar indicator	36
6.2.2	Keuze voor het aggregatieniveau	37
6.2.3	Vaststellen van het onderscheidend vermogen	37
6.2.4	Aantallen zorgaanbieders en patiënten bij bepalen van het onderscheidend vermogen	38
7	Aanbieden aan Register: besluitvorming, werkinstructie en aanbiedingsproces	40
7.1	Besluiten om de PROM en indicator aan te bieden aan het Register	40
7.2	Werkinstructie	40
7.3	Het proces van aanbieden en toetsen	41
	Literatuur	44
	Bijlagen	
	Bijlage 1 Dankwoord	47
	Bijlage 2 Onderzoeksopzet en meetmomenten	49

Voorwoord

Voor u ligt de *Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures*. Nationaal en internationaal is het gebruik van *Patient Reported Outcome Measures* (PROMs), als bron voor indicatoren van kwaliteit van zorg, in opkomst. Met PROMs is het mogelijk om de effectiviteit van een behandeling te meten vanuit het perspectief van de patiënt. Recente ontwikkelingen laten zien dat PROMs sterk kunnen bijdragen aan kwaliteitsverbetering. Zorginstituut Nederland juicht deze ontwikkelingen toe, omdat deze bijdragen aan de zoektocht naar betere/effectievere indicatoren.

De handreiking is bedoeld voor alle partijen die betrokken (willen) zijn bij het ontwikkelen of gebruiken van PROMs als indicator(en) voor kwaliteit van zorg. Het was een uitdaging om de handreiking voor alle partijen begrijpelijk vorm te geven. De inleidende hoofdstukken (1 en 2) zijn zo geschreven dat deze voor een breed lezerspubliek te begrijpen zijn. In de daaropvolgende hoofdstukken worden methodologische begrippen besproken waarbij wij adviseren om deze thema's aanvullend te bespreken met experts op het gebied van ontwikkeling, beoordeling en gebruik van PROMs.

Het NIVEL heeft in samenwerking met Zorginstituut Nederland de handreiking ontwikkeld die helpt bij het selecteren en (door-)ontwikkelen van PROMs vragen(lijsten). Dit document dient tevens als achtergronddocument van het Toetsingskader kwaliteitsstandaarden en meetinstrumenten. De handreiking biedt inzicht in de te volgen stappen bij het selecteren of ontwikkelen van PROMs. In deze handreiking ligt de focus op PROMs die worden gebruikt als indicator voor kwaliteit van zorg. Bij het opstellen van de handreiking is nadrukkelijk contact gezocht met verschillende methodologische experts op het gebied van PROMs, patiëntenorganisaties, zorgaanbieders en zorgverzekeraars om enerzijds alle kennis over het ontwikkelen en selecteren van PROMs te bundelen en anderzijds te zorgen dat het document breed toepasbaar is en draagvlak heeft in Nederland (zie Bijlage 1). Wij willen iedereen die mondeling dan wel schriftelijk een bijdrage heeft geleverd, daarvoor hartelijk danken.

Laura Koopman

Zorginstituut Nederland

1 Inleiding

1.1 Achtergrond

De afgelopen jaren zijn forse inspanningen geleverd om de kwaliteit van zorg transparant te maken voor patiënten¹, zorgaanbieders en zorgverzekeraars, met als doel kwaliteit van zorg te borgen of te verbeteren. Daarnaast draagt transparantie van kwaliteit van zorg bij aan het versterken van de positie van de patiënt. Patiënten krijgen namelijk meer en betere toegang tot informatie over de zorg. Tegelijkertijd vormen zij zelf een belangrijke en waardevolle informatiebron: de patiënt weet het best hoe de zorg voor hem was en wat het heeft opgeleverd. Het is dan ook van groot belang om patiënten te betrekken bij het meten van kwaliteit van zorg.

De overheid heeft een belangrijke rol gespeeld bij het transparant maken van kwaliteit van zorg. Hiertoe heeft de overheid onder meer het voormalige programma Zichtbare Zorg in het leven geroepen om de ontwikkeling van kwaliteitsindicatoren te initiëren, stimuleren en faciliteren. Tevens is de Consumer Quality Index (CQI of CQ-index) door de overheid ondersteund als standaard voor het meten van kwaliteit van zorg vanuit het patiëntenperspectief. Met de komst van Zorginstituut Nederland neemt de overheid een andere positie in en komt het initiatief voor het transparant maken van kwaliteit van zorg in grotere mate bij partijen in de zorg (patiëntenorganisaties, zorgaanbieders en zorgverzekeraars) te liggen.

Het Zorginstituut stimuleert de partijen in de zorg om hun rol bij transparantie van kwaliteit op te pakken en zorgt ervoor dat helder is wat goede zorg is. Het instituut schept ook helderheid over de voorwaarden waar informatie over kwaliteit van zorg aan moet voldoen. Hiertoe is een Toetsingskader kwaliteitsstandaarden en meetinstrumenten opgesteld waarin verschillende procedurele criteria voor de totstandkoming van kwaliteitsstandaarden en meetinstrumenten zijn opgenomen (Zorginstituut Nederland, 2014d). Daarnaast zorgt het Zorginstituut ervoor dat partijen toegang hebben tot begrijpelijke en betrouwbare informatie over de kwaliteit van zorg. Op deze manier kan de patiënt informatie vinden over de geleverde zorg en eventueel een keuze maken voor een passende zorgaanbieder; zo kunnen de patiënt en professional samen goede beslissingen nemen over behandeling en ondersteuning en zo kunnen zorgverzekeraars goede zorg inkopen.

¹ In deze handreiking wordt de term 'patiënt' gehanteerd. Waar 'patiënt' staat kan ook 'cliënt' gelezen worden.

De partijen in de zorg kunnen meetinstrumenten en indicatoren aanbieden aan het Zorginstituut die vervolgens toetst of deze voldoen aan het Toetsingskader.

Het Toetsingskader geeft aan waaraan meetinstrumenten moeten voldoen voor opname in het landelijk Register van het Zorginstituut. Belangrijke criteria zijn dat meetinstrumenten en indicatoren zijn ontwikkeld en/of geselecteerd in samenwerking met alle relevante partijen (patiëntenorganisaties, zorgaanbieders en zorgverzekeraars) en dat deze partijen de instrumenten en indicatoren gezamenlijk aanbieden aan het Register. Daarnaast bekijkt het instituut hoe de validiteit, betrouwbaarheid, vergelijkbaarheid en het onderscheidend vermogen van meetinstrumenten en indicatoren zijn onderbouwd (Zorginstituut Nederland, 2014d). Door de relevante partijen te betrekken krijgen deze partijen de mogelijkheid om zorg te dragen voor de aandachtspunten die zij belangrijk vinden, bijvoorbeeld door erop te letten dat het meetinstrument of de indicator het patiëntenperspectief voldoende vertegenwoordigt (patiëntenorganisaties), voldoende recht doet aan de kwaliteit van zorg die wordt geleverd (zorgaanbieders), of voldoende geschikt is voor zorginkoop (zorgverzekeraars).

Een relatief nieuwe ontwikkeling betreft het gebruik van *Patient-Reported Outcome Measures* (PROMs) als bron voor indicatoren van kwaliteit van zorg. PROMs worden uitgevraagd bij patiënten en sluiten daardoor nauw aan bij het besef dat patiënten een centrale bron zijn van informatie over kwaliteit van zorg. PROMs zijn complementair aan de CQ-index omdat PROMs zich primair richten op uitkomsten van zorg, terwijl de CQ-index zich meer richt op structuur- en procesaspecten van zorg. Het gebruik van PROMs als bron voor indicatoren van kwaliteit van zorg heeft onder meer veel aandacht gekregen door het 'PROMs initiative' in het Verenigd Koninkrijk waar op zeer grote schaal ervaring is opgedaan met kwaliteitsindicatoren gebaseerd op PROMs voor verschillende electieve ingrepen (Devlin et al., 2010). In Nederland zijn ook verschillende initiatieven ontplooid. Zo hebben zorgverzekeraars een serie pilots met PROMs uitgevoerd en worden PROMs momenteel ook in een groeiend aantal klinische registraties ingebed.

1.2 Doel, positionering en doelgroep

Rond PROMs bestaat een rijke onderzoekstraditie met veel literatuur over methoden, technieken en eisen aan validiteit en betrouwbaarheid. Met deze handreiking biedt Zorginstituut Nederland een overzicht, zodat voor partijen duidelijk wordt welke kennis uit de PROMs literatuur op welke wijze kan helpen bij het voldoen aan de criteria voor meetinstrumenten uit het Toetsingskader. Deze handreiking richt zich op het

selecteren van relevante uitkomsten, het beoordelen en selecteren van PROMs waarmee die uitkomsten worden gemeten en het definiëren van indicatoren gebaseerd op PROMs. Hierbij is gebruik gemaakt van bestaande consensus uit de wetenschappelijke literatuur over PROMs (Mokkink et al., 2010; NQF, 2013; Reeve et al., 2013) en is eveneens aangesloten bij de bestaande documentatie rond het Toetsingskader (Koolman et al., 2012; Zorginstituut Nederland, 2014b).

Deze handreiking is bedoeld voor (samenwerkingsverbanden van) patiëntenorganisaties, zorgverzekeraars en zorgaanbieders die indicatoren gebaseerd op PROMs willen ontwikkelen en gebruiken als indicator voor kwaliteit van zorg. Volgen deze partijen de PROMs handreiking, dan kunnen zij voldoen aan de criteria uit het Toetsingskader. Bij de uitvoering van verschillende onderdelen van de handreiking is echter ook expertise vereist met betrekking tot het ontwikkelen en beoordelen van vragenlijsten en het beoordelen van indicatoren. Het is dan ook noodzakelijk dat partijen die een PROM als indicator voor kwaliteit willen gebruiken en aan willen bieden aan het Zorginstituut, deze expertise daarbij betrekken. Deze handreiking ondersteunt daarmee de interactie tussen de partijen in de zorg en experts op het gebied van het ontwikkelen en beoordelen van vragenlijsten en indicatoren.

1.3 *Verschillende doelen van PROMs en prioritering*

Bij het gebruik van PROMs zijn verschillende gebruiksdoelen te onderscheiden waaronder het monitoren van patiënten in de dagelijkse klinische praktijk, als uitkomstmaat in clinical trials en als indicator voor kwaliteit van zorg. Binnen het gebruik als indicator voor kwaliteit van zorg zijn ook verschillende doelen te onderscheiden zoals publieke informatie voor patiënten, toezichtinformatie of zorginkoopinformatie. Over de mate waarin het mogelijk is om verschillende gebruiksdoelen van PROMs te combineren, lopen de meningen uiteen. Het combineren van gebruiksdoelen betekent vaak dat geen van de doelen optimaal wordt gediend. Het scheiden van gebruiksdoelen lost dit op, maar betekent wel dat er meerdere metingen moeten plaatsvinden binnen dezelfde patiëntenpopulaties met net iets andere vragenlijsten en methoden. De primaire focus van het Toetsingskader en het Register, en daarmee ook van deze handreiking, is het gebruik van indicatoren voor kwaliteit van zorg gericht op publieke informatie die valide, betrouwbaar en vergelijkbaar is.

1.4 Opbouw en leeswijzer

In hoofdstuk twee worden PROMs en enkele gerelateerde begrippen gedefinieerd en wordt een globaal stappenplan geschetst om tot een selectie van PROMs te komen en deze te gebruiken als basis voor indicatoren voor kwaliteit van zorg. In dit hoofdstuk wordt ook de relatie gelegd tussen de stappen die in dit document staan beschreven en de criteria uit het Toetsingskader kwaliteitsstandaarden en meetinstrumenten. Hoofdstuk drie tot en met zes vormen de nadere uitwerking van die stappen. Hoofdstuk 7 schetst het aanbiedings- en toetsproces dat voorafgaat aan opname in het Register.

2 PROMs: definiëring en stappenplan

2.1 PRO, PROM en kwaliteitsindicator

Patient-Reported Outcomes (PROs) zijn uitkomsten van zorg die betrekking hebben op de gezondheid van de patiënt en die de patiënt zelf rapporteert (FDA, 2006). Voorbeelden van PROs zijn symptomen, functioneren en gezondheidsgerelateerde kwaliteit van leven. Wanneer een instrument wordt gebruikt om een PRO te meten, wordt het ook wel *Patient Reported Outcome Measure* (PROM) genoemd. Met PROMs is het mogelijk om de effectiviteit van een behandeling te meten vanuit het perspectief van de patiënt. Als dit instrument als basis dient voor een indicator, waarmee de kwaliteit van een zorgaanbieder of behandelaar in een numerieke score wordt uitgedrukt, wordt het een prestatiemaat of (kwaliteits)indicator genoemd. Een term die aansluit bij internationale literatuur is *Patient Reported Outcome - Performance Measure* (PRO-PM; NQF, 2013).

2.2 Generiek stappenplan en relatie toetsingscriteria

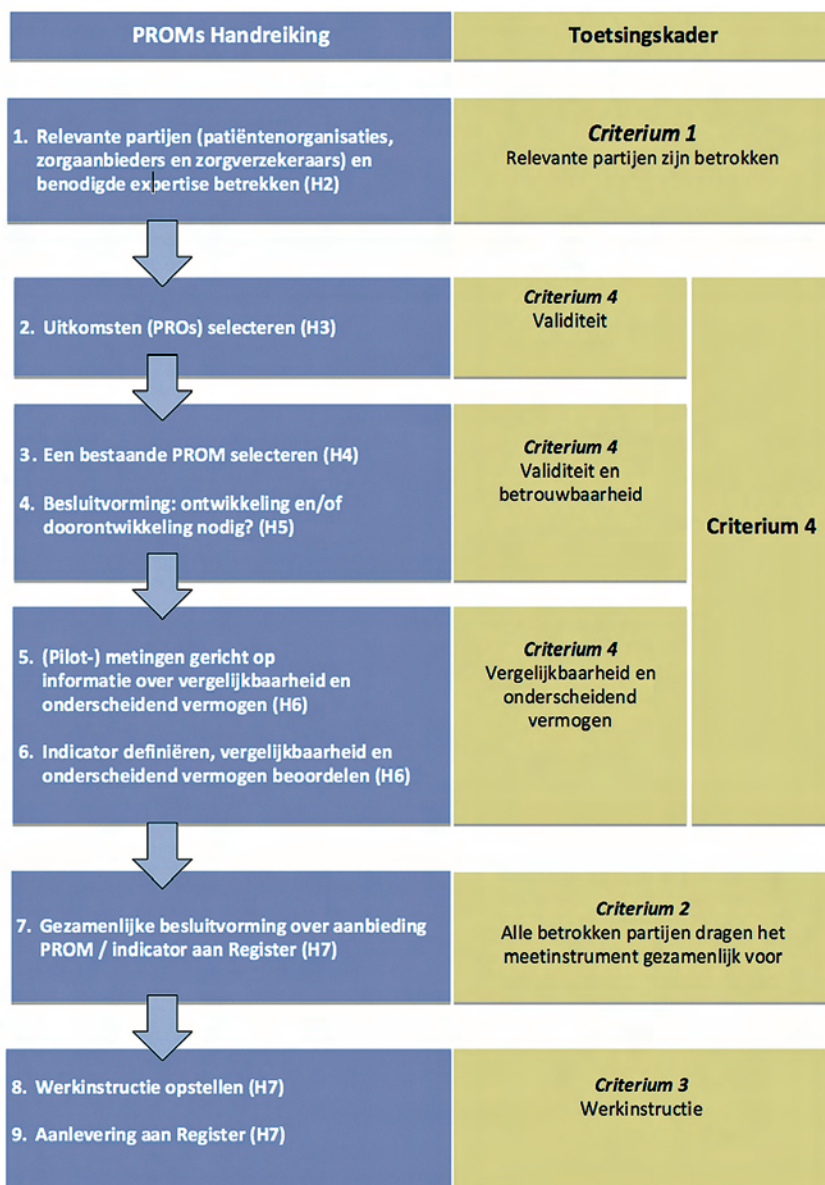
In figuur 2.1 zijn de stappen weergegeven die worden doorlopen om tot een adequate uitkomstindicator en bijbehorend meetinstrument te komen. Aan de rechterkant van de figuur is weergegeven op welk criterium uit het Toetsingskader kwaliteitsstandaarden en meetinstrumenten de stappen betrekking hebben. Vanuit de gedachte dat deze stappen uiteindelijk leiden tot een indicator die wordt aangeboden aan het Register van Zorginstituut Nederland, is de eerste stap om patiëntenorganisaties, zorgaanbieders en zorgverzekeraars te betrekken. Belangrijke eisen uit het Toetsingskader zijn namelijk dat deze partijen betrokken zijn bij de totstandkoming van een meetinstrument of indicator en deze gezamenlijk aanbieden aan het Register (Zorginstituut Nederland, 2014d). Dit kan bijvoorbeeld worden geborgd in de vorm van een begeleidingscommissie die de keuzes en besluiten bij de verschillende stappen neemt. Het ligt voor de hand om in dit stadium ook de experts met kennis van het ontwikkelen en beoordelen van vragenlijsten en indicatoren te betrekken.

De stappen 2 t/m 6 zijn relevant voor verschillende onderdelen van criterium 4 uit het Toetsingskader (validiteit, betrouwbaarheid, vergelijkbaarheid en onderscheidend vermogen). Stap 2 betreft het vaststellen van de uitkomsten (PROs) die belangrijk zijn voor de betreffende patiëntenpopulatie en die worden beïnvloed door de kwaliteit van zorg (hoofdstuk 3). Vervolgens worden bestaande meetinstrumenten die de

uitkomst kunnen meten voor de betreffende patiëntenpopulatie bij elkaar gezocht en beoordeeld op validiteit en betrouwbaarheid (stap 3; hoofdstuk 4).

Figuur 2.1

Schematisch overzicht van stappen om tot een PROM en een indicator te komen in relatie tot de toetsingscriteria uit het Toetsingskader kwaliteitsstandaarden en meetinstrumenten van het Zorginstituut



In het ideale scenario leidt voorgaande tot de keuze voor een PROM met bewezen validiteit en betrouwbaarheid. Dan ligt het voor de hand direct met deze PROM verder te gaan. Het zal echter ook geregeld voorkomen dat over de validiteit en betrouwbaarheid van een PROM onvoldoende bekend is of dat informatie hierover niet is gebaseerd op een Nederlandse populatie. In een enkel geval kan het gebeuren dat er überhaupt geen passende bestaande PROM beschikbaar is. Dergelijke scenario's vragen om zorgvuldige besluitvorming over ontwikkeling of doorontwikkeling van een PROM alvorens verder te gaan. Stap 4 betreft dan ook de besluitvorming hierover (zie hoofdstuk 5).

Wanneer een PROM voorhanden is die (voldoende) geschikt is als basis voor een indicator vindt een pilotmeting plaats op grond waarvan de indicator kan worden gedefinieerd en de vergelijkbaarheid en het onderscheidend vermogen kunnen worden beoordeeld (stap 5 en 6; hoofdstuk 6). Vervolgens kunnen de betrokken patiëntenorganisaties, zorgaanbieders en zorgverzekeraars beslissen of zij de PROM en de indicator, eventueel met aanpassingen, gezamenlijk willen aanbieden aan het Register van het Zorginstituut (stap 7; hoofdstuk 7). Als dat het geval is dienen de partijen een werkinstructie op te stellen met instructies en specificaties voor het gebruik van de PROM en de indicator (stap 8; hoofdstuk 7). Vervolgens kunnen zij het aanbiedingsformulier samen met de PROM, de indicator en de werkinstructie ter toetsing aanbieden (stap 9; hoofdstuk 7).

Aanbevelingen

Betrek de relevante partijen uit de zorg (patiëntenorganisaties, zorgaanbieders en zorgverzekeraars) bij de stappen uit het stappenplan. Dit is van belang voor criterium 1 en 2 van het Toetsingskader (relevante partijen zijn betrokken en dragen gezamenlijk voor). Betrek eveneens expertise met betrekking tot het ontwikkelen en beoordelen van vragenlijsten en indicatoren.

3 Selectie van PROs

Dit hoofdstuk richt zich op de selectie van de uitkomsten zelf (PROs). Voordat er een meetinstrument kan worden geselecteerd of (door-)ontwikkeld, moet bekend zijn welke uitkomsten dat meetinstrument in kaart behoort te brengen. Van groot belang hierbij is dat uitkomsten worden geselecteerd die 1) er volgens patiënten toe doen en 2) die worden beïnvloed door de kwaliteit van zorg. De selectie van een PRO vormt daarmee de basis voor de inhoudsvaliditeit van de PROM of de indicator die daaraan wordt ontleend. Deze stap is dan ook essentieel voor criterium 4 uit de het Toetsingskader kwaliteitsstandaarden en meetinstrumenten waarin de validiteit wordt getoetst. Uiteraard is het ook van belang om patiënten(-organisaties), zorgaanbieders en zorgverzekeraars te betrekken bij deze stap en de geselecteerde PROs gezamenlijk vast te stellen (Zorginstituut Nederland, 2014d).

3.1 Patiëntbetrokkenheid

Om vast te stellen welke uitkomsten er voor patiënten toe doen is onderzoek onder patiënten nodig. Kwalitatief onderzoek ligt hierbij het meest voor de hand. Een veelgebruikte methode is het doen van focusgroeps gesprekken met patiënten.

Focusgroeps gesprekken zijn uitermate geschikt om snel helder te krijgen welke uitkomsten voor patiënten belangrijk zijn, omdat patiënten op elkaar kunnen reageren door elkaar aan te vullen of juist tegen te spreken (Brédart et al., 2014). Een aandachtspunt bij focusgroepen is het betrekken van deelnemers met verschillende leeftijden en opleidingsniveaus evenals niet-westerse migranten. Centrale vragen die gesteld kunnen worden aan patiënten, kunnen gaan over het effect of resultaat dat men verwacht van een behandeling, over de problemen die men verholpen of verminderd zou willen zien en/of wat men door de klachten of symptomen minder of niet kan doen. Naast de symptomen en fysieke beperkingen die patiënten kunnen noemen, kan ook worden gedacht aan uitkomsten zoals zelfmanagement of therapietrouw, vooral bij patiënten met een chronische aandoening.

Een andere manier om patiënten te betrekken is het doen van interviews (Brédart et al., 2014). Over het algemeen is dit echter wel arbeidsintensiever en er is geen onderlinge interactie tussen patiënten mogelijk. Daardoor duurt het vaak langer voordat de kern van het patiëntenperspectief op relevante uitkomsten duidelijk wordt. Een ander, meer kwantitatief alternatief, kan zijn om patiënten te vragen een serie

mogelijk relevante uitkomsten te prioriteren. Het risico hieraan is dat een uitkomst die specifiek relevant is voor die patiëntengroep niet wordt opgenomen in de selectie van te prioriteren uitkomsten en dus wordt gemist.

Vornoemde benaderingen om vast te stellen welke uitkomsten patiënten belangrijk vinden, sluiten elkaar niet uit en zijn wellicht ook niet uitputtend. Het is aan te raden om relevante expertise met betrekking tot (kwalitatieve) onderzoeksmethoden te betrekken bij de keuze en/of uitvoering van een benadering.

Een aandachtspunt in dit stadium is de vraag of de PRO generiek is of juist specifiek voor de betreffende patiëntengroep. Tijdens focusgroeps-gesprekken of interviews wordt vaak al duidelijk of patiënten spreken in termen van specifieke symptomen die horen bij hun aandoening of dat patiënten vooral spreken in generieke termen. Wanneer het vooral gaat over algemene aspecten van gezondheid kan het namelijk nuttig zijn om een generieke PRO te selecteren. Dit kan voordelen bieden omdat deze breed toepasbaar zijn bij meerdere patiëntengroepen (zie ook paragraaf 4.1).

3.2 *Relatie met kwaliteit van zorg*

Wanneer het meten van een PRO uiteindelijk moet leiden tot informatie over de kwaliteit van zorg, is het van essentieel belang dat een PRO wordt gekozen die beïnvloed wordt door de kwaliteit van zorg (Koolman et al., 2012). Patiënten kunnen het beste aangeven wat zij zien als de meest belangrijke uitkomsten waarop zij verwachten dat kwaliteit van zorg van invloed is. Daarnaast is van belang dat de verschillende partijen die de PROM en de indicator uiteindelijk gaan aanbieden aan het Register, de relatie tussen de uitkomst en de kwaliteit van zorg voldoende aannemelijk achten. Mogelijke aanvullende manieren om zicht te krijgen op de relatie tussen kwaliteit van zorg en uitkomsten zijn literatuurstudies en expert meetings, maar ook aanvullende wetenschappelijk onderzoek kan hierbij worden gebruikt.

3.3 *Is een PRO wel geschikt?*

Het is denkbaar dat bij het selecteren van relevante uitkomsten blijkt dat er geen geschikte PRO beschikbaar is, bijvoorbeeld als er geen plausibele relatie tussen relevante PROs en kwaliteit van zorg valt te leggen. Of als patiënten een zorguitkomst van belang vinden die niet door henzelf kan worden gerapporteerd. Het is dus van groot belang om bij de

selectie van PROs de vraag te stellen of de uitkomsten die in dit proces naar voren komen wel als een PRO kunnen worden gemeten en of voldoende aannemelijk is dat de PRO wordt beïnvloed door de kwaliteit van zorg. In het uiterste geval zou het proces hier kunnen stoppen, maar het is ook denkbaar dat wordt gekozen voor aanvullend onderzoek naar de mogelijkheid een bepaalde uitkomst te meten als een PRO, of om de relatie met kwaliteit van zorg beter te onderbouwen.

Aanbevelingen

Vraag patiënten wat relevante uitkomsten zijn en betrek ook patiëntenorganisaties, zorgaanbieders en zorgverzekeraars bij het vaststellen van de relevante uitkomsten. Kies een uitkomst waarvan plausibel is dat deze kan worden beïnvloed door kwaliteit van zorg. Deze aanbevelingen dragen bij aan de inhoudsvaliditeit van de PROM en hebben daarmee betrekking op het onderdeel 'validiteit' uit criterium 4 van het Toetsingskader.

Meer lezen?

- Brédart et al. (2014) gaan in op de technieken (zoals communicatie- en luistervaardigheden) die een interviewer dient te beheersen bij focusgroepen en cognitieve interviews. Zij beschrijven tevens praktische voorbeelden en geven enkele voorbeeldvragen. In een artikel van Brod et al. (2009) worden best practices en een overzicht gepresenteerd van de kwalitatieve onderzoeksmethoden om de inhoudsvaliditeit vast te stellen (zie ook paragraaf 4.2.2).
- Het *Handboek Ontwikkelen CQ-index* bevat een werkinstructie voor het uitvoeren van focusgroepen die niet specifiek gericht is op PROMs, maar wel veel praktische aanknopingspunten biedt (Zorginstituut Nederland, 2014e).
- Voor meer informatie over het betrekken van patiëntenorganisaties biedt ook de *Tool Patiëntenparticipatie in richtlijnontwikkeling aanknopingspunten* (Kwaliteitsinstituut voor de Gezondheidszorg CBO en NPCF, 2010).

4 Selectie van bestaande PROMs

Dit hoofdstuk laat zien op welke wijze een PROM kan worden geselecteerd en wat daarbij belangrijke aandachtspunten zijn. De selectie van een PROM houdt nauw verband met de onderdelen ‘validiteit’ en ‘betrouwbaarheid’ uit criterium 4 van het Toetsingskader kwaliteitsstandaarden en meetinstrumenten. Bij de validiteit van de PROM gaat het er met name om dat de uitkomst (PRO) die is geselecteerd op een herkenbare en coherente manier terugkomt in de PROM.

De betrouwbaarheid heeft betrekking op de mate waarin een meting met de PROM vrij is van meetfout. Uiteraard is van belang dat patiënten(-organisaties), zorgaanbieders en zorgverzekeraars bij de selectie van een PROM zijn betrokken en het resultaat onderschrijven, zodat kan worden voldaan aan criterium 1 en 2 van het Toetsingskader (de relevante partijen zijn betrokken en dragen gezamenlijk voor). De materie uit dit hoofdstuk vereist de nodige inhoudelijke expertise over het ontwikkelen en beoordelen van vragenlijsten. Het is dan ook van belang om die expertise hierbij te betrekken.

4.1 Identificeren van bestaande PROMs

Gezien de veelheid aan reeds ontwikkelde meetinstrumenten is de eerste stap om na te gaan welke PROMs er al zijn alvorens zelf een nieuw instrument te gaan ontwikkelen. Dit voorkomt een wildgroei aan meetinstrumenten. Daarnaast is het gebruik van een bestaande PROM aanzienlijk goedkoper dan het ontwikkelen van een nieuwe PROM.

Er zijn verschillende manieren om te inventariseren welke PROMs beschikbaar zijn voor de betreffende patiëntengroep en/of uitkomsten. Veelal wordt literatuuronderzoek uitgevoerd (Mokkink et al., 2009), liefst systematisch. Soms kan gebruik worden gemaakt van een reeds bestaand systematisch literatuuronderzoek (Kenniscentrum Meetinstrumenten, 2014). Er zijn ook verschillende online informatie-systemen beschikbaar voor het vinden van PROMs. Voorbeelden hiervan zijn de *Patient Reported Outcomes Measurement Group* (University of Oxford; <http://phi.uhce.ox.ac.uk/>), de *Patient-Reported Outcome and Quality Of Life Instruments Database* (www.proqolid.org/) en *Meetinstrumenten in de Zorg* (www.meetinstrumentenzorg.nl/). Daarnaast is een overkoepelend systeem beschikbaar dat bestaat uit generieke PROMs voor de meeste elementen van fysiek, mentaal en sociaal functioneren (PROMIS; Cella et al., 2010). Andere, meer pragmatische benaderingen zijn om onder experts of onder beroepsgroepen te inventariseren welke PROMs zij als relevant zien, of om de inventarisatie

te beperken tot instrumenten die nu al in meer of mindere mate worden gebruikt binnen een sector of beroepsgroep. Deze benaderingen zijn uiteraard minder grondig dan systematisch literatuuronderzoek met als risico dat goede instrumenten worden gemist.

Bij de inventarisatie van PROMs is het van belang om vast te stellen of de inventarisatie zich richt op generieke of specifieke PROMs of beide. Een generieke PROM kan onafhankelijk van de aandoening van de patiënt worden ingezet en bevat dus geen termen of vragen die specifiek zijn voor een bepaalde patiëntengroep. Generieke PROMs zijn vaak nuttig om verschillende patiëntengroepen te vergelijken of voor zeer groot-schalige metingen onder patiëntengroepen met uiteenlopende gezondheidsproblemen. Specifieke instrumenten daarentegen, zijn juist ontwikkeld om te gebruiken bij een specifieke patiëntengroep en bevatten vaak termen of vragen die alleen voor die patiëntengroep relevant zijn. Specifieke PROMs zijn doorgaans ook beter in staat om veranderingen in de PRO te meten (McKenna, 2011). De keuze voor een specifieke of generieke PROM is onder meer van belang voor zoekstrategieën en/of voor de vraag waarmee experts worden benaderd bij het inventariseren van relevante PROMs.

4.2 Evaluatie en selectie van een PROM

Als duidelijk is welke PROMs beschikbaar zijn en aansluiten bij de uitkomsten die men wil meten, dan kan de kwaliteit van deze PROMs worden beoordeeld om tot een selectie te komen. Voor de beoordelingscriteria sluiten we aan bij de consensus binnen de *International Society for Quality of Life Research* (ISOQOL; Reeve et al., 2013). Er is gekozen voor de ISOQOL omdat deze is gebaseerd op consensus van experts en omdat het gebruik maakt van andere belangrijke overzichtsdocumenten van onder andere FDA (2006), *Scientific Advisory Committee of the Medical Outcomes Trust* (2002) en COSMIN (Mokkink, et al., 2010). ISOQOL heeft deze documenten gebruikt tezamen met onderzoek onder hun leden om een minimale standaard op te stellen waaraan PROMs moeten voldoen. De eigenschappen van PROMs die de ISOQOL hierbij betreft zijn hieronder weergegeven met tussen haakjes steeds het criterium van het Toetsingskader waar dat aspect betrekking op heeft:

- Conceptueel- en meetmodel
- Inhoudsvaliditeit (Criterium 4: validiteit)
- Constructvaliditeit (Criterium 4: validiteit)
- Cross-culturele validiteit (Criterium 4: validiteit)
- Betrouwbaarheid (Criterium 4: betrouwbaarheid)
- Responsiviteit (Criterium 4: validiteit)
- Hanteerbaarheid en interpreterbaarheid

Het overzicht laat zien dat de eigenschappen die ISOQOL onderscheidt met name terugkomen in criterium 4 van het Toetsingskader. De eigenschappen conceptueel- en meetmodel en hanteerbaarheid hebben niet direct betrekking op de toetsingscriteria, maar zijn indirect wel van belang. Hier komen we op terug bij de uitwerking van deze eigenschappen in paragraaf 4.2.1 tot en met 4.2.7.

Voor alle eigenschappen geldt dat beschikbare evidentie uit de literatuur lang niet altijd gebaseerd zal zijn op Nederlandse populaties. Wanneer uit een studie bij een Amerikaanse populatie blijkt dat een PROM op alle eigenschappen gunstig scoort, betekent dit nog niet dat dit ook geldt voor de vertaalde PROM in een Nederlandse populatie. In de gevallen waarin zo'n PROM toch wordt geselecteerd dienen de eigenschappen van die PROM opnieuw te worden bekeken aan de hand van (pilot-) studies onder Nederlandse patiënten. Daarnaast is de kwaliteit van de studies waarin één of meer eigenschappen van een PROM zijn beoordeeld een punt van aandacht. Bij twijfels hierover kan de COSMIN (*COnsensusbased Standards for the selection of health Measurement INstruments*) worden geraadpleegd, die onder meer heeft geleid tot een checklist met bijbehorend beoordelingssysteem waarmee de kwaliteit van studies naar meetinstrumenten kan worden beoordeeld (Mokkink, Terwee, Knol, et al., 2010; Terwee et al., 2012).

4.2.1 Conceptueel- en meetmodel

De eerste belangrijke eigenschap van een PROM die ISOQOL onderscheidt is het conceptueel- en meetmodel. In feite gaat het hier om het antwoord op de vragen: 'Welke uitkomsten (PROs) meet de PROM?' en 'Welke items/vragen horen bij welke uitkomst?'. Het conceptueel en meetmodel bevat dus informatie die relevant is voor het beoordelen van de inhoudsvaliditeit van de PROM (zie paragraaf 4.2.2). Dit moet duidelijk zijn gedocumenteerd en verantwoord, bijvoorbeeld in de vorm van een artikel. Voor de verantwoording wordt vaak factoranalyse gebruikt, een statistische techniek waarmee wordt onderbouwd dat verschillende sets van vragen ook daadwerkelijk verschillende uitkomsten meten. Van belang is uiteraard dat de PROMs die worden geëvalueerd uitkomsten meten die overeenstemmen met de uitkomsten (PROs) die in stap 2 zijn geselecteerd.

4.2.2 Inhoudsvaliditeit

Inhoudsvaliditeit is de mate waarin de inhoud van de PROM een adequate afspiegeling is van de te meten uitkomst (Mokkink, et al., 2010). Een eerste borging van inhoudsvaliditeit vindt al plaats wanneer PROMs worden gekozen die uitkomsten meten die patiënten belangrijk en relevant vinden (zie ook hoofdstuk 3 – Selectie van PROs). De inhoudsvaliditeit kan verder worden onderbouwd door patiënten en zorginhoude-

lijke experts te vragen of zij vinden dat de te meten uitkomsten op een herkenbare en complete manier terugkomen in de PROM. Cognitieve interviews onder patiënten vormen hierbij een nuttig hulpmiddel waarmee wordt bekeken of patiënten de vragen eenduidig interpreteren, of die interpretatie past bij de uitkomst die wordt gemeten, of zij relevante aspecten missen etc. (zie ook box 4.1).

In veel gevallen is het zo dat de PROM ontwikkeld is door enkel zorgverleners, zonder directe input van patiënten. Alvorens zo'n PROM te selecteren is het zo mogelijk nog belangrijker om te verifiëren dat patiënten de uitkomsten die zij belangrijk vinden, voldoende herkennen in de PROM.

Box 4.1

Cognitieve interviews

Tijdens een cognitief interview wordt een vragenlijst voorgelegd aan patiënten die deze hardop denkend invullen. Door het stellen van verdiepende vragen, verkrijgt men inzicht in de manier waarop respondenten de PROM interpreteren. Meer specifiek kunnen cognitieve interviews worden gebruikt om:

- aan patiënten te vragen of zij de items relevant vinden voor de te meten uitkomst en of er nog belangrijke items missen (Brod et al., 2009; Frost et al., 2007). Dit geeft inzicht in de inhoudsvaliditeit van de PROM;
- van een vertaalde vragenlijst vast te stellen of patiënten de vragen interpreteren zoals deze in de oorspronkelijke taal bedoeld zijn (Beaton et al., 2000). Dit geeft inzicht in de cross-culturele validiteit van een vertaalde vragenlijst;
- te beoordelen in hoeverre verschillende patiënten de vragen uit de PROM op dezelfde manier interpreteren. Dit geeft inzicht in de betrouwbaarheid van een vragenlijst (Buers et al., 2013; Knafl et al., 2007);
- inzicht te krijgen in de hanteerbaarheid voor patiënten, doordat duidelijk wordt of patiënten niet teveel belast worden door de lengte van de vragenlijst of door eventuele gevoelige vragen in de PROM.

Wanneer de bevindingen van een cognitief interview worden gebruikt om de PROM te verbeteren, bevordert dit de validiteit, betrouwbaarheid of hanteerbaarheid van de PROM.

4.2.3 *Constructvaliditeit*

Constructvaliditeit heeft betrekking op het toetsen van vooraf opgestelde hypothesen over de grootte en richting van onderlinge correlaties tussen de PROM en andere relevante meetinstrumenten. Meer specifiek gaat het over de mate waarin een meetinstrument correleert met andere meetinstrumenten die hetzelfde meten (convergente validiteit) en niet of minder correleert met instrumenten die wat anders meten (divergente validiteit) (Mokkink et al., 2010). Zo wordt bijvoorbeeld een correlatie verwacht tussen twee verschillende PROMs die betrekking hebben op pijn; indien die correlatie niet wordt gevonden roept dit twijfels op over de validiteit van (één van) beide PROMs. Idealiter is op basis van eerder onderzoek informatie beschikbaar over de convergente en/of divergente validiteit. Is dit niet het geval, dan kan hierover aanvullende informatie worden verzameld bij de eerste (pilot-)metingen. Het risico daarbij is uiteraard dat relaties tussen de PROM en andere instrumenten niet voldoen aan de verwachtingen, waarmee de constructvaliditeit in een laat stadium alsnog ter discussie komt te staan. In dit licht is het aantrekkelijk om een PROM te selecteren waarvan de constructvaliditeit reeds is onderzocht.

4.2.4 *Cross-culturele validiteit*

Als een PROM recent is vertaald, is het van groot belang dat de vertaling van een PROM op een zorgvuldige, professionele manier is gedaan. Naast een goede vertaling, is cross-culturele adaptatie ook van belang. Bepaalde items in een vragenlijst uit het buitenland kunnen een andere betekenis hebben in Nederland. Hier dient rekening mee te worden gehouden bij het vertalen van een PROM. Tevens is het bij een vertaling van groot belang om te borgen dat het taalniveau van de PROM eenvoudig is en dat de PROM door verschillende bevolkingsgroepen op dezelfde wijze wordt geïnterpreteerd. Een breed geaccepteerde methode die rekening houdt met voornoemde aandachtspunten is de forward-backward vertaalmethode (zie ook box 4.2; Beaton et al., 2000). Indien in Nederland al veel ervaring is opgedaan met een vertaalde PROM en daaruit blijkt dat deze in de Nederlandse populatie voldoet aan de verschillende criteria uit dit hoofdstuk, dan weegt het minder zwaar als de vertaling niet strikt volgens de regels der kunst is uitgevoerd.

Box 4.2

De forward-backward vertaalmethode (Beaton et al., 2000)

- 1 Voorwaartse vertaling: Meerdere onafhankelijke vertalers, bij voorkeur vertalers met de doeltaal als moedertaal, vertalen de oorspronkelijke versie naar de doeltaal (Nederlands).
- 2 Expertpanel: De twee vertalers uit stap 1 komen tot consensus voor wat betreft eventuele inadequate vertalingen die zijn gemaakt in stap 1.
- 3 Terugwaartse vertaling: Het instrument wordt terug vertaald naar oorspronkelijke taal door een onafhankelijke vertaler die de oorspronkelijke taal als moedertaal heeft, maar geen kennis heeft van de originele vragenlijst.
- 4 Expertgroep: een expertgroep bereikt consensus over de vertalingen en zorgt ervoor dat er geen discrepanties meer zijn tussen het bron- en het doelinstrument.
- 5 Pre-test door middel van cognitieve interviews: Het instrument wordt getest op de doelpopulatie door middel van cognitieve interviews om te kijken of de vertaling geïnterpreteerd wordt zoals het origineel was bedoeld. Daarnaast dragen cognitieve interviews bij aan verschillende andere eigenschappen van een PROM (zie box 4.1).

4.2.5 *Betrouwbaarheid*

Betrouwbaarheid is de mate waarin een meting vrij is van meetfout (COSMIN: Mokkink, et al., 2010). Wanneer een meting vrij is van meetfout, levert een meting in dezelfde omstandigheden hetzelfde resultaat op. De betrouwbaarheid drukt dus uit hoe precies er wordt gemeten. Idealiter is de betrouwbaarheid hoog zodat de PRO zo precies en goed mogelijk gemeten wordt. Een manier om in vragenlijstonderzoek de betrouwbaarheid te bevorderen, is door met meerdere vragen dezelfde PRO te meten. Voor betrouwbaarheid van uitkomsten die met meerdere vragen worden gemeten (uni-dimensionele schalen) wordt standaard gekeken naar de *Cronbachs alfa* (Reeve et al., 2013; Sijtsma, 2009). Een gebruikelijke conventie is dat de minimale betrouwbaarheid van schalen 0,70 is voor vergelijkingen op groepsniveau (Nunnally, 2010). Het is belangrijk dat het aantal waarnemingen waarop Cronbachs alfa gebaseerd is groot genoeg is; anders geeft Cronbachs alfa geen goede indicatie van de betrouwbaarheid (idealiter meer dan 500; Drenth en Sijtsma, 2012). Een veelgenoemde alternatieve methode voor het testen van betrouwbaarheid is de test-hertest betrouwbaarheid waarmee wordt bekeken of een herhaalde meting onder gelijke omstandigheden hetzelfde resultaat geeft. Deze vorm van het testen van betrouwbaarheid stuit bij PROMs echter op kritiek omdat men er vanuit gaat dat de uitkomst die wordt gemeten juist kan veranderen. Dit zou voor een onder-

schatting van de betrouwbaarheid zorgen. Daarnaast kunnen methoden uit de Item Respons Theorie (IRT) bijdragen aan de betrouwbaarheid van een PROM (zie box 4.3).

Box 4.3

Item Respons Theorie

Item respons theorie (IRT) is een verzameling moderne en complexe technieken die in aanvulling op *Cronbachs alfa* kunnen bijdragen aan de betrouwbaarheid van metingen. Toepassing van IRT wordt door ISOQOL niet gezien als minimale eis, maar over het algemeen wordt er preciezer gemeten met instrumenten die zijn ontwikkeld met IRT methoden (Reeve et al., 2013). De bekendste vorm van IRT is Rasch-analyse. Het belangrijkste kenmerk van IRT modellen is dat deze modellen de vragen (items) in een vragenlijst kunnen rangschikken op hun moeilijkheidsgraad. Vijf kilometer hardlopen is bijvoorbeeld aanzienlijk moeilijker voor mensen dan uit bed komen. Dit is uiteraard zo overduidelijk dat daar geen IRT model aan te pas hoeft te komen, maar er zijn genoeg minder duidelijke vergelijkingen te bedenken waar IRT nodig is om de rangorde in moeilijkheid van vragen te bepalen. Een ander belangrijk voordeel is dat ook personen kunnen worden gerangschikt op hun niveau van functioneren. Dergelijke inzichten en mogelijkheden kunnen bijdragen aan de betrouwbaarheid (Embretson & Reise, 2000; Reeve et al., 2013) en validiteit (Hambleton et al., 1991) van een PROM. Voor meer informatie over IRT verwijzen we naar een artikel van Edelen en Reeve (2007), waarin de meest belangrijke elementen worden besproken over het toepassen van IRT bij vragenlijstontwikkeling.

4.2.6 Responsiviteit

Responsiviteit is het vermogen van een PROM om veranderingen in de te meten uitkomst te detecteren (Mokkink et al., 2010). Responsiviteit wordt gezien als onderdeel van validiteit, vanuit de gedachte dat een PROM die een uitkomst goed meet, ook veranderingen in deze uitkomst kan meten. ISOQOL ziet responsiviteit als een minimale eis in longitudinale studies met PROMS (Reeve et al., 2013). Om de responsiviteit vast te stellen, zijn aanvullende criteria of *anchors* nodig. Een voorbeeld van een anchorvraag is: *‘In welke mate zijn uw klachten veranderd in vergelijking tot de situatie van vlak voor uw behandeling?’* (Ostelo et al., 2003). Vervolgens wordt bekeken of mensen die deze vraag positiever beantwoorden ook een gunstiger ontwikkeling in hun PROM-score laten zien. Ook kan de responsiviteit van een PROM blijken uit de relatie met een andere PROM waarvan de responsiviteit al is vastgesteld (Revicki et al., 2006).

Bij het gebruik van PROMs voor het meten van kwaliteit van zorg gaat het meestal om de mate waarin zorg leidt tot een verbetering van functioneren of verdere teruggang in functioneren voorkomt/vertraagt. De responsiviteit van een PROM is hiervoor uiteraard van essentieel belang.

4.2.7 Hanteerbaarheid en interpreteerbaarheid

De eigenschap hanteerbaarheid gaat in op verschillende aspecten van een vragenlijst:

- de lengte van de PROM;
- de frequentie van dataverzameling;
- het vereiste niveau van taal- en gezondheidsvaardigheden;
- de aansluiting bij fysieke en cognitieve mogelijkheden van de patiëntengroep.

Informatie over de hanteerbaarheid van een PROM volgt doorgaans uit cognitieve interviews waarbij patiënten wordt gevraagd een vragenlijst in te vullen en hierbij hardop na te denken en waarbij verdiepende vragen worden gesteld over de PROM (zie box 4.1; Beatty en Willis, 2007; Willis, 2005). Patiënten kunnen in de cognitieve interviews aangeven of de PROM begrijpelijk is, of de lengte te behappen is etc.

Aangezien ook de meest hanteerbare PROMs nog enige belasting met zich meebrengen voor patiënten, wordt vaak besproken of deze belasting in verhouding staat met het nut. Het onderbouwen van het nut begint met kwalitatief onderzoek om te bepalen welke uitkomst (PRO) patiënten relevant vinden. Dit is bij het selecteren van de uitkomst al aan de orde geweest (zie paragraaf 3.1). Een andere manier om het nut voor patiënten te vergroten kan zijn om patiënten zelf een directe terugkoppeling van de eigen resultaten te geven, of de resultaten te bewerken tot publieke informatie.

Interpreteerbaarheid gaat over de betekenis van scores van patiënten op de PROM voor verschillende partijen in de zorg en draagt daarmee bij aan het nut van een PROM. Interpreteerbaarheid begint bij duidelijkheid over welke scores als hoog en welke scores als laag gezien mogen worden. Informatie over het verschil in PROM-scores tussen een patiëntenpopulatie en een gezonde populatie kan verder bijdragen aan interpreteerbaarheid evenals informatie over welke veranderingen in PROM-scores er volgens patiënten toe doen. In dit verband wordt ook wel gesproken van een minimaal relevant verschil. De interpreteerbaarheid van een PROM is van belang bij de selectie van een PROM, maar dient opnieuw te worden bekeken als een indicator wordt gebaseerd op de PROM. In hoofdstuk 6 (van PROM naar indicator) komen we hier dan ook op terug en wordt eveneens het fenomeen minimaal relevant verschil toegelicht (zie paragraaf 6.2).

Hanteerbaarheid en interpreteerbaarheid komen niet direct terug in het Toetsingskader kwaliteitsstandaarden en meetinstrumenten. Indirect speelt het wel een rol bij de validiteit van een PROM, want een PROM die niet te begrijpen is voor de doelgroep kan ook niet valide zijn. Daarnaast is in het verleden gebleken dat lange vragenlijsten op minder draagvlak kunnen rekenen; dit kan een risico vormen voor criterium 2 (patiëntenorganisaties, zorgaanbieders en zorgverzekeraars dragen het meetinstrument of de indicator gezamenlijk voor). Ook een goede interpreteerbaarheid kan bijdragen aan draagvlak. In dit licht is het dan ook heel zinvol om stil te staan bij de hanteerbaarheid en interpreteerbaarheid van een PROM.

Aanbevelingen

Evalueer de geïdentificeerde PROMs met betrekking tot het conceptueel- en meetmodel, inhoudsvaliditeit, constructvaliditeit, cross-culturele validiteit, betrouwbaarheid, responsiviteit, hanteerbaarheid en interpreteerbaarheid. Betrek de relevante partijen die voornemens zijn om de PROM aan te bieden voor het Register. Indien een PROM wordt gekozen op grond van de evaluatie, documenteer dan in het bijzonder informatie met betrekking tot de evaluatiecriteria voor de geselecteerde PROM. Door het zorgvuldig doorlopen van deze stappen is de kans om tot een valide en betrouwbare indicator te komen het grootst. Dit draagt bij aan de validiteit en betrouwbaarheid waar met criterium 4 van het Toetsingskader op getoetst wordt.

Meer lezen?

Er zijn verscheidene artikelen beschikbaar waarin meer informatie over kwalitatieve onderzoeksmethoden en de selectie van PROs wordt beschreven. In onderstaande (niet volledig uitputtende) lijst wordt een selectie gepresenteerd.

- In artikelen van Lasch et al. (2010), Brod et al. (2009) en Magasi (2012) worden kwalitatieve methoden en best practices besproken voor het onderbouwen van inhoudsvaliditeit.
- Een artikel van Patrick et al. (2011) beschrijft de relevantie van cognitieve interviews en er zijn in dat artikel werkwijzen opgenomen. Beatty en Willis (2007) en Willis (2005) beschrijven meer informatie over de achterliggende theorie van cognitieve interviews. Ook is meer informatie over de procedure en werkwijze van cognitieve interviews te vinden in het *Handboek Ontwikkelen CQ-index* in de vorm van praktische werk-instructies (Zorginstituut Nederland, 2014c).

- Voor meer informatie over constructvaliditeit kan *Measurement in Medicine* geraadpleegd worden van De Vet et al. (2011).
- Over cross-culturele validiteit beschrijven Herdman et al. (1998) in een wetenschappelijk artikel een model om in verschillende talen dezelfde concepten te meten, aangepast aan de cultuur en taal. De forward-backwardmethode wordt beschreven in een wetenschappelijk artikel van Beaton et al. (2000).
- Meer informatie over responsiviteit is te vinden in wetenschappelijke artikelen van Revicki et al. (2006; 2008).

5 Besluitvorming over ontwikkeling of doorontwikkeling

In het ideale geval zijn één of meer PROMs beschikbaar met goede eigenschappen en complete informatie over die eigenschappen op basis van een Nederlandse populatie. Als de betrokken patiëntenorganisaties, zorgaanbieders en zorgverzekeraars zich ook kunnen vinden in die PROM ligt het voor de hand de PROM(s) in gebruik te nemen en er indicatoren op te baseren. Het zal echter ook regelmatig voorkomen dat informatie ontbreekt over één of meer van de eigenschappen van een PROM, of dat een PROM voor één of meer van die eigenschappen niet voldoet. Met name ernstige twijfels aan de validiteit of betrouwbaarheid van een PROM kunnen aanleiding zijn tot (door-) ontwikkeling aangezien deze eigenschappen ook worden getoetst wanneer instrumenten of indicatoren worden aangeboden aan het Register van het Zorginstituut. Voorts zal voorkomen dat informatie over de eigenschappen van een PROM niet zijn gebaseerd op een Nederlandse populatie. In het uiterste geval is ook denkbaar dat er geen bestaande PROMs beschikbaar zijn voor een uitkomst of dat over de eigenschappen van die PROMs weinig tot niets bekend is. Het is dus niet vanzelfsprekend dat de inventarisatie, evaluatie en selectie van PROMs direct leidt tot een bevredigend instrument. Het is dan ook verstandig in deze fase zorgvuldig stil te staan bij het nut of de noodzaak van ontwikkeling of doorontwikkeling.

5.1 Doorontwikkeling

Doorontwikkeling kan bestaan uit het verzamelen van (aanvullende) informatie over de eigenschappen van een PROM, het verifiëren van informatie over die eigenschappen binnen een Nederlandse populatie of pogingen verbeteringen aan te brengen in de eigenschappen van een PROM. De agenda voor doorontwikkeling wordt gevormd door de eigenschappen uit paragraaf 4.2.1 t/m 4.2.7 waar twijfels over bestaan. In deze secties worden ook de technieken genoemd die hierbij kunnen worden gebruikt.

Voor verschillende eigenschappen van een PROM kan doorontwikkeling tegelijkertijd informatie opleveren over het gebruik van de PROM als indicator. Dit geldt met name als de doorontwikkeling een dataverzameling vereist, bijvoorbeeld in het geval van onderzoek naar constructiviteit of het vaststellen van de betrouwbaarheid in een Nederlandse populatie. De pilotmeting kan dan zodanig worden opgezet dat tegelijkertijd inzicht ontstaat in de mate waarin de indicator verschillen laat

zien tussen zorgaanbieders en welke aandachtspunten er zijn voor de vergelijkbaarheid (zie paragraaf 6.1 en 6.2 voor meer informatie). De belangrijkste consequentie voor de opzet van de pilotmeting is het streven naar deelname van 30 zorgaanbieders en 30 respondenten per zorgaanbieder (Zorginstituut Nederland, 2014b). Het is ook mogelijk om direct over te gaan op grootschalige (landelijke) metingen gericht op het gebruik van PROMs als indicator en deze metingen een pilotstatus te geven totdat op basis van de data verschillende eigenschappen van de PROM beter zijn onderbouwd.

In hoeverre doorontwikkeling noodzakelijk is hangt af van de aard en omvang van de informatie die ontbreekt over de eigenschappen van een PROM of de tekortkomingen die hierin zijn geconstateerd. De partijen die voornemens zijn de PROM als indicator aan te bieden aan het Zorginstituut, dienen hier een besluit over te nemen, idealiter bijgestaan door experts op het gebied van het ontwikkelen of beoordelen van vragenlijsten.

5.2 Ontwikkelen van een nieuwe PROM

Wanneer geen PROM voorhanden is voor de te meten uitkomsten of alleen PROMs waarvan weinig of niets bekend is over de eigenschappen, kan het ontwikkelen van een nieuwe PROM worden overwogen. Dit vraagt echter een aanzienlijke hoeveelheid tijd en middelen. Het is dan ook van belang om de noodzaak ervan zorgvuldig vast te stellen.

Aangezien de verwachting is dat het ontwikkelen van een volledig nieuwe PROM in veel gevallen niet nodig zal zijn, volstaan we hier met een globaal overzicht van het ontwikkelproces en verwijzen we naar andere literatuur voor meer gedetailleerde informatie. In figuur 5.1 wordt kort weergegeven wat de gebruikelijke fases zijn die worden doorlopen bij het ontwikkelen van een PROM. Normaal gesproken wordt de eerste stap uit dit schema voorafgegaan door het vaststellen van de relevante uitkomsten door focusgroepen of interviews bij patiënten. Aangezien deze stap in het kader van dit document reeds is uitgevoerd bij de selectie van PROs (zie hoofdstuk 3) komt deze stap in de figuur niet terug. Ook staan de enkele methoden en werkwijzen die doorgaans voor die fases worden gehanteerd kort genoemd.

De precieze invulling van iedere stap kan per ontwikkeltraject verschillen en vereist zorgvuldige afstemming met experts op het gebied van vragenlijstconstructie. Evenals bij de doorontwikkeling van een PROM geldt bij het ontwikkelen van een nieuwe PROM dat het mogelijk is om

Figuur 5.1

Schematisch overzicht van de ontwikkeling van een PROM¹

1 Inhoud en vorm vaststellen (validiteit)	2 Kwalitatieve testfase	3 Kwantitatieve testfase	4 Aanpassing aan instrument
Doelen -Items genereren -Vaststellen van de manier van datacollectie, score-administratie, herinneringsperiode, antwoordcategorieën Werkwijze -Expert meetings / interviews met stakeholders en patiënten, literatuur review, transcripties van focusgroepen	Doelen -Verificatie bij patiënten van concepten, antwoordcategorieën, herinneringsperiode en leesbaarheid -Belasting onderzoeken bij patiënten Werkwijze -Cognitieve interviews bij patiënten. Zie Willis (2005) en Procedure Cognitief Interview (Zorginstituut Nederland, 2014b)	Doelen -Vaststellen van betrouwbaarheid, dimensionaliteit, responsiviteit, constructvaliditeit -Vaststellen minimaal relevant verschil Werkwijze -ICC, SEM, factoranalyses, IRT-analyses. Richtlijnen zijn te vinden in bijvoorbeeld De Vet et al. (2011). -Responsiviteit vaststellen met behulp van ankervraag of andere PROM (zie paragraaf 4.2.6) -Ideaaliter bestaande PROMs en/of andere variabelen meenemen t.b.v. constructvaliditeit	Doelen -Naar aanleiding van de kwantitatieve testfase dient de PROM aangepast te worden Werkwijze -Denk hierbij aan het schrappen van items met matige psychometrische eigenschappen, items die te veel lijken op andere items of items die niet bij een schaal blijken te passen.

¹ Uitgangspunt voor deze figuur is dat de relevante uitkomsten (PROs) in een eerder stadium reeds zijn vastgesteld (zie ook hoofdstuk 3).

de dataverzameling zodanig op te zetten dat ook inzicht wordt verkregen in de eigenschappen van indicatoren die worden gebaseerd op de PROM (zie hoofdstuk 6 voor een toelichting). De belangrijkste consequentie voor de opzet van de dataverzameling is het streven naar deelname van 30 zorgaanbieders en 30 respondenten per zorgaanbieder (Kreft, 1996; Zorginstituut Nederland, 2014b).

Aanbevelingen

Besluit of een PROM moet worden doorontwikkeld of dat een nieuwe PROM moet worden ontwikkeld. Indien een nieuwe PROM wordt ontwikkeld, zorg dan dat het traject resulteert in informatie met betrekking tot de evaluatiecriteria uit paragraaf 4.2, zodat de validiteit en betrouwbaarheid van de PROM zijn onderbouwd. Dit is van belang voor criterium 4 van het Toetsingskader kwaliteitsstandaarden en meetinstrumenten. Overweeg om een kwantitatieve testfase zodanig uit te breiden dat tegelijkertijd inzicht ontstaat in de eigenschappen van de indicator zoals beschreven in hoofdstuk 6.

Meer lezen?

Als naslagwerken doen we enkele suggesties, maar er zijn vele andere standaardwerken die geraadpleegd kunnen worden:

- Een praktisch boek over meten in de gezondheidszorg is geschreven door De Vet et al. en heet *Measurement in Medicine* (2011). Hierin staat beschreven welke aspecten van belang zijn bij de ontwikkeling of beoordeling van een meetinstrument in de gezondheidszorg.
- Verder is een nuttig document gepubliceerd door de *Food and Drug Administration* (FDA; 2006). In dit document staat op een wat meer theoretische wijze beschreven hoe de FDA PROMs beoordeelt en evalueert.
- Delen van het *Handboek Ontwikkelen CQ-index* (Zorginstituut Nederland, 2014b) bieden aanknopingspunten, omdat er elementen zijn in de ontwikkeling van de CQ-index die overeenkomen met de ontwikkeling van PROMs, zoals het houden van focusgroepen, het doen van cognitieve interviews of het uitvoeren van een kwantitatieve testfase.

6 Vergelijkbaarheid en onderscheidend vermogen: van PROM naar indicator

Een goede, valide en betrouwbare PROM vormt de beste basis voor een goede indicator. Voor de specifieke gebruiksdoelen publieke informatie en zorginkoopinformatie zijn er echter nog aanvullende criteria naast validiteit en betrouwbaarheid, te weten: de mate waarin resultaten van verschillende zorgaanbieders vergelijkbaar zijn (vergelijkbaarheid) en de mate waarin de indicator verschillen laat zien tussen zorgaanbieders (onderscheidend vermogen). Vergelijkbaarheid en onderscheidend vermogen komen ook terug in criterium 4 van het Toetsingskader kwaliteitsstandaarden en meetinstrumenten, maar vanuit de literatuur over PROMs zal hier doorgaans weinig over bekend zijn om twee redenen:

- 1 Informatie over deze criteria vraagt een heel specifieke opzet van data-verzameling waar de meeste studies naar PROMs niet aan voldoen omdat die zich niet richten op het vergelijken van zorgaanbieders.
- 2 Studies die zich wel richten op het gebruik van PROMs voor het vergelijken van zorgaanbieders doen dat vaak in een ander land of andere setting.

Om de vergelijkbaarheid en het onderscheidend vermogen van een indicator goed te kunnen beoordelen is dus vrijwel altijd een nieuwe data-verzameling nodig.

In dit hoofdstuk gaan we eerst in op de criteria ‘vergelijkbaarheid’ en ‘onderscheidend vermogen’. Vervolgens geven we aan welke eisen en aandachtspunten hieruit voortvloeien voor de opzet van de dataverzameling. De partijen die de PROM en de indicator willen aanbieden aan het Zorginstituut dienen uiteindelijk vast te stellen of de vergelijkbaarheid en het onderscheidend vermogen voldoende zijn voor het gebruiksdoel dat zij voor ogen hebben. De materie uit dit hoofdstuk vereist echter ook expertise over het ontwikkelen en beoordelen van indicatoren en het is dan ook noodzakelijk die expertise te betrekken bij het beoordelen van de vergelijkbaarheid en het onderscheidend vermogen van een indicator.

6.1 *Vergelijkbaarheid*

Bij een indicator is het van belang dat er geen sprake is van vertekening waardoor zorgaanbieders niet zuiver te vergelijken zijn. Voor het onderscheid tussen verschillende vormen van vertekening sluiten we aan bij de indicatorstandaard die spreekt van registratievergelijkbaarheid, popula-

tievergelijkbaarheid en steekproef-/responsvergelijkbaarheid (Koolman et al., 2012). Hierbij geven we aan op welke wijze deze vormen van vertekening tot uiting kunnen komen bij een indicator.

6.1.1 Registratievergelijkbaarheid

Registratievergelijkbaarheid heeft betrekking op de mate waarin meetfouten vergelijkingen tussen zorgaanbieders vertekenen. In het geval van een indicator gaat het primair om de mate waarin dataverzameling bij alle zorgaanbieders op dezelfde manier plaatsvindt. Hierbij is van belang dat procedures en specificaties voor dataverzameling en analyse helder en eenduidig zijn vastgelegd en bij alle zorgaanbieders op dezelfde gestandaardiseerde manier worden toegepast. Een van de criteria uit het Toetsingskader die hieraan is gerelateerd, is de beschikbaarheid van een werkinstructie die moet worden ingediend als de PROM of indicator wordt aangeboden aan het Register (Zorginstituut Nederland, 2014d). Belangrijke methodologische keuzes die hier ook bij worden gemaakt zijn de keuze voor het moment in de behandeling waarop patiënten wordt gevraagd de PROM in te vullen, evenals de keuze voor meerdere meetmomenten of een enkel meetmoment met vragen over het huidige functioneren en het functioneren in het verleden. Meer informatie over deze vraagstukken is te vinden in bijlage 2.

In het geval van één centraal georganiseerde dataverzameling is registratievergelijkbaarheid relatief eenvoudig te borgen. Het risico op systematische verschillen in het meetproces tussen zorgaanbieders is dan namelijk betrekkelijk gering. Wanneer iedere zorgaanbieder zelf metingen uitvoert of wanneer de gegevens van metingen door verschillende partijen worden verwerkt, zijn de risico's voor registratievergelijkbaarheid aanzienlijk groter. In het laatste geval kan de eenduidigheid van de instructies van het meetproces ook beter worden beoordeeld, want dan kan bijvoorbeeld een helpdesk worden ingericht voor vragen van de verschillende partijen die de instructies en specificaties van het meetproces moeten toepassen. Een bundeling van deze vragen laat zien in hoeverre de (werk-)instructies eenduidig interpreteerbaar en toepasbaar zijn. Als er sprake is van verschillende dataverwerkende bureaus, registraties of software applicaties loont het om te bekijken of er op dat niveau grote verschillen zijn in resultaten. Dit kan worden bekeken door per bureau, registratie of softwareapplicatie de gemiddelde score voor de indicator te bepalen. Is er sprake van één of meer extreme scores, dan is dat aanleiding om na te gaan of die score het gevolg kan zijn van vertekening door een andere wijze van dataverzameling, registratie of codering.

6.1.2 *Populatievergelijkbaarheid*

Populatievergelijkbaarheid heeft betrekking op de mate waarin verschillen tussen de patiëntenpopulaties van zorgaanbieders de vergelijking van een indicator tussen zorgaanbieders vertekenen. Dit kan bijvoorbeeld gebeuren als een zorgaanbieder relatief veel patiënten heeft met comorbiditeit die als gevolg daarvan minder goed reageren op een behandeling; de zorgaanbieder krijgt dan onterecht een lagere score op de indicator. In dit verband wordt vaak gesproken over casemixcorrectie, een statistische correctie voor verschillen in de kenmerken van patiëntenpopulaties die van invloed zijn op de uitkomst en waar een zorgaanbieder niets aan kan doen (Koolman et al., 2012; Zaslavsky et al., 2000; Zorginstituut Nederland, 2014b; zie ook box 6.1 voor meer uitleg over rekening houden met casemix).

Om de populatievergelijkbaarheid te borgen is het van groot belang te inventariseren welke kenmerken van de patiëntenpopulatie de vergelijikbaarheid kunnen vertekenen. Veel gebruikte casemix variabelen zijn de demografische kenmerken leeftijd, geslacht, opleiding, SES en leefsituatie/burgerlijke staat. Ook klinische kenmerken komen in aanmerking. Voorbeelden hiervan zijn kenmerken van de aandoening zoals ernst en duur van een aandoening en co-morbiditeit. In dit verband wordt de eerste score, voorafgaand aan de behandeling, ook regelmatig gebruikt als casemix variabele. Het functioneren van een patiënt voorafgaand aan een behandeling hangt namelijk samen met de mate waarin vooruitgang of achteruitgang mogelijk is (Coles, 2010).

De keuze van de casemix variabelen is van groot belang voor de acceptatie van de resultaten en vindt bij voorkeur plaats op basis van literatuur, empirische gegevens en de mening van experts (Deutsch et al., 2012). Idealiter leidt de inventarisatie tot een wat uitgebreidere set met mogelijke casemix variabelen die op grond van nader onderzoek wordt teruggebracht (Coles, 2010). Zo kan bijvoorbeeld blijken dat een veronderstelde casemixvariabele uiteindelijk toch niet samenhangt met de indicator, of dat er geen verschillen zijn tussen de populaties van zorgaanbieders met betrekking tot die variabele waardoor het geen zin heeft voor die variabele te corrigeren.

Rekening houden met casemix

De meest simpele vorm om rekening te houden met casemix is stratificatie. Hierbij worden respondenten gesplitst in groepen op basis van de casemix factoren. Dit houdt in dat (bijvoorbeeld) de score voor de laagopgeleide mannen van zorgaanbieder A wordt vergeleken met de score voor laagopgeleide mannen van zorgaanbieder B. Het opsplitsen in groepen zorgt echter voor een lager aantal waarnemingen per groep, waarmee verschillen tussen zorgaanbieders minder snel significant worden. Daarom wordt in veel gevallen regressieanalyse gebruikt zodat er statistisch kan worden gecorrigeerd voor casemix factoren zonder concessies te hoeven doen aan de groepsgrootte voor statistische toetsen (zie ook paragraaf 6.2.3).

6.1.3 Steekproef- en responsvergelijkbaarheid

Steekproef- en responsvergelijkbaarheid hebben betrekking op de representativiteit van de patiënten die worden uitgenodigd om de PROM in te vullen en de representativiteit van de groep die dit ook daadwerkelijk doet. Het borgen van steekproefvergelijkbaarheid kan plaatsvinden door steekproeven te controleren op representativiteit voor bijvoorbeeld leeftijd en geslacht en zo nodig een nieuwe steekproef te trekken. De respons-vergelijkbaarheid is lastiger te borgen. Een zo hoog mogelijke respons draagt over het algemeen bij aan responsvergelijkbaarheid en manieren om de respons te bevorderen, zoals bijvoorbeeld het afstemmen van dataverzamelmethode op de doelgroep, kunnen hierbij helpen. Toch biedt dit geen garanties: sommige patiëntengroepen zijn nu eenmaal minder bereidwillig om vragenlijsten in te vullen en ook bij een hoge respons kan het voorkomen dat de respondenten niet representatief zijn voor de doelgroep. Het blijft daarom van belang om in alle gevallen ook de representativiteit van de respons per zorgaanbieder te toetsen. Dit vereist dat er een registratie is van respondenten en non-respondenten, inclusief enkele achtergrondkenmerken, zodat kan worden bekeken hoe hoog de respons is en of respondenten representatief zijn voor de totale groep die is benaderd. Voornoemde kwesties spelen uiteraard ook voor de CQ-index, waarvoor reeds procedures en werkinstructies bestaan die ook aanknopingspunten bieden voor PROMs. Zo zijn er bijvoorbeeld onderdelen over dataverzameling met behulp van interviews (Zorginstituut Nederland, 2013b), standaard verzend-schema's voor schriftelijke en/of online dataverzamelingen (Zorginstituut Nederland, 2013a) en steekproeftrekking, respons en representativiteit. Ook de bescherming van de privacy komt daarbij uitgebreid aan bod. De genoemde CQI richtlijnen bieden aanknopingspunten

ten, maar zijn niet in alle gevallen zonder meer toepasbaar voor PROMS; dit hangt af van de wijze waarop metingen met PROMS worden ingericht.

Aanbevelingen

Registratievergelijkbaarheid: Zorg dat het proces van dataverzameling en bewerking overal zo gestandaardiseerd mogelijk verloopt.

Evalueer de eenduidigheid van meetinstructies in overleg met de gebruikers of op basis van de vragen die hierover zijn gesteld. Bekijk op grond van data of er aanwijzingen zijn voor registratieverschillen tussen partijen die data verzamelen en onderzoek die aanwijzingen.

Populatievergelijkbaarheid: Identificeer op basis van literatuur en in overleg met experts factoren die de populatievergelijkbaarheid kunnen vertekenen en dus in aanmerking komen als casemixadjuster. Neem deze potentiële casemixadjusters mee in de dataverzameling zodat deze kunnen worden getoetst op relevantie en significantie.

Steekproef- en responsvergelijkbaarheid: Kies een dataverzamelmethode die aansluit bij de doelgroep teneinde de respons te bevorderen. Zorg voor een gedegen en verifieerbare registratie van steekproef en respons zodat de representativiteit kan worden getoetst.

Voornoemde aanbevelingen geven inzicht in de vergelijkbaarheid van de indicator en helpen om de vergelijkbaarheid te verhogen. Dit is van belang voor het onderdeel vergelijkbaarheid van criterium 4 uit het Toetsingskader.

6.2 Onderscheidend vermogen

Bij het vergelijken van zorgaanbieders (of enig ander organisatorisch verband) gaat het niet meer om onderscheid maken tussen individuen, maar tussen zorgaanbieders. In dit verband wordt ook wel gesproken van onderscheidend vermogen. De Indicatorstandaard spreekt van 'statistisch betrouwbaar onderscheiden' (Koolman et al., 2012). In het algemeen is de gedachte dat naarmate een indicator meer onderscheidend is, deze ook informatiever is en de resultaten bijgevolg beter geschikt zijn als publieke informatie of zorginkoopinformatie. Vanuit deze gedachte worden eisen gesteld aan het onderscheidend vermogen van indicatoren. Indicatoren die hier bij lange na niet aan voldoen kunnen nog steeds heel accuraat en valide zijn als alle zorgaanbieders ook min of meer dezelfde kwaliteit van zorg leveren. Maar dergelijke indicatoren worden niet of nauwelijks informatief en bruikbaar gevonden voor patiëntkeuze of zorginkoop.

6.2.1 Van PROM naar indicator

Alvorens het onderscheidend vermogen van een indicator te kunnen beoordelen moet de indicator worden gedefinieerd. Dit kan op verschillende manieren. In het algemeen is het idee dat de indicator weergeeft in hoeverre zorg van invloed is op de uitkomst (PRO), bijvoorbeeld dat de uitkomst verbetert door zorg, of – bij bijvoorbeeld chronische aandoeningen – dat de zorg achteruitgang van de uitkomst vertraagt of vermindert. In dit licht gaat het dus om veranderingen in een uitkomst in de tijd. In het geval van twee tijdstippen kan de indicator worden gedefinieerd als het verschil in het gemiddelde of de somscores tussen de twee tijdstippen. Een indicator kan ook gedefinieerd worden als het al dan niet behalen van een vooraf vastgesteld verschil tussen de twee tijdstippen, zoals bijvoorbeeld het percentage patiënten dat een minimaal relevant verschil heeft behaald (zie box 6.2 voor meer informatie over minimaal relevant verschil).

Box 6.2

Minimaal relevant verschil

Het minimaal relevant verschil is de kleinste verandering in de score op de PROM die belangrijk wordt gevonden door patiënten (De Vet et al., 2007). De aanbevolen methode voor het bepalen van een minimaal relevant verschil is door het te schatten op basis van anchor-based methoden (Revicki et al., 2008). Daarmee is het bepalen van het minimaal relevant verschil nauw verwant aan het bepalen van de responsiviteit (zie paragraaf 4.2.6). Anchor-based methoden maken gebruik van een ander instrument of een anchor vraag om te bepalen wanneer een verschil relevant is (Jaeschke et al., 1989). Een andere invalshoek zijn de distribution-based methoden. Distribution-based methoden richten zich op het kleinst waarneembare verschil dat buiten de meetfout kan worden waargenomen (De Vet & Terwee, 2010). Distribution-based methoden gaan daarmee voorbij aan de vraag of dat kleinst waarneembare verschil wel relevant wordt gevonden door patiënten.

Voor verschillende behandelingen is het eveneens interessant om te kijken of de meting voorafgaand aan de behandeling verschilt tussen zorgaanbieders. Dergelijke verschillen kunnen er bijvoorbeeld op duiden dat sommige zorgaanbieders bij een geringer niveau van klachten een behandeling inzetten dan anderen. Ook hierbij zijn verschillende varianten mogelijk waaronder bijvoorbeeld de gemiddelde score voorafgaand aan de behandeling of het percentage patiënten dat voorafgaand aan de behandeling een bepaalde norm of afkapwaarde overschrijdt.

Een eenduidige conventie om een indicator te definiëren is vooralsnog niet voorhanden en de keuze kan ook afhangen van wat het beste aansluit bij de belevingswereld van de gebruiker (Hildon et al., 2012). Daarnaast is het verstandig om van enkele varianten vast te stellen in welke mate deze verschillen tussen instellingen laten zien.

6.2.2 *Keuze voor het aggregatieniveau*

Een belangrijke keuze bij het vaststellen van het onderscheidend vermogen van een indicator is het aggregatieniveau waarop het onderscheidend vermogen wordt bekeken. Voorbeelden van aggregatieniveaus die tot op heden veel voorkomen zijn instelling, afdeling, zorgketen, concern, locatie etc. Belangrijke overwegingen bij het kiezen van een aggregatieniveau zijn op welk niveau kwaliteitsverschillen worden verwacht en op welk niveau de resultaten bruikbaar zijn. Idealiter worden meerdere aggregatieniveaus meegenomen bij (pilot-)metingen zodat in ieder geval kan worden bekeken op welke niveaus de indicator meer of minder verschillen in kwaliteit laten zien.

6.2.3 *Vaststellen van het onderscheidend vermogen*

Wat betreft de statistische modellen voor het bepalen van het onderscheidend vermogen van een indicator tussen zorgaanbieders (of enig ander aggregatieniveau) hebben multilevel modellen de voorkeur. Binnen de CQ-index zijn deze modellen vrijwel uitsluitend gebruikt voor het vergelijken van zorgaanbieders (Zorginstituut Nederland, 2014a) en ook bij zorginhoudelijke indicatoren worden idealiter multilevel modellen toegepast, hoewel deze voor zorginhoudelijke indicatoren in een enkel geval niet passen (Koolman et al., 2012; Reitsma et al., 2011). In het kort houden multilevel modellen rekening met het feit dat patiënten binnen één zorgverlener meer met elkaar gemeen hebben dan patiënten van verschillende zorgverleners. Het statistisch model dat gebruikt wordt, dient hier rekening mee te houden en multilevel modellen doen dat ook. Een tweede voordeel is dat de scores van zorgaanbieders met weinig waarnemingen accurater worden geschat. De voordelen van multilevel analyse voor het vergelijken van zorgaanbieders zijn elders uitvoeriger beschreven (De Boer et al., 2010) evenals de werkwijze (Zorginstituut Nederland, 2013c).

Bij gebruik van multilevel modellen voor het vergelijken van zorgaanbieders kan de betrouwbaarheid van schattingen op het niveau van zorgaanbieders worden berekend als de intra-class reliability waarbij waarden van 0,70 en hoger duiden op een goed onderscheidend vermogen (Zorginstituut Nederland, 2013c). In de Indicatorstandaard worden verschillen tussen zorgaanbieders op basis van effect size (Cohens D) ingedeeld in klein, medium en groot. Vervolgens wordt met behulp van

poweranalyse voor kleine, medium en grote verschillen bekeken of er voldoende waarnemingen zijn om 80% van die verschillen significant aan te tonen (Koolman et al., 2012). Beide manieren van kijken zijn zeer gevoelig voor het aantal waarnemingen per zorgaanbieder.

6.2.4 Aantallen zorgaanbieders en patiënten bij bepalen van het onderscheidend vermogen

Voor het bepalen van het discriminerend vermogen is het uitgangspunt dat een enigszins substantiële groep van zorgaanbieders moet deelnemen aan de dataverzameling. Naarmate meer zorgaanbieders deelnemen ontstaat namelijk beter inzicht in welke verschillen tussen zorgaanbieders voorkomen. Op basis van literatuur en uitgebreide ervaring met deze thematiek streeft de CQ-index naar deelname van 30 zorgaanbieders en 30 respondenten per zorgaanbieder voor pilotmetingen (Kreft, 1996; Zorginstituut Nederland, 2013c, 2014b). Ook voor PROMs is dit een goede basis om de multilevel modellen te kunnen toepassen die de voorkeur hebben voor verschillende analyses (zie paragraaf 6.2.3) en biedt het voldoende inzicht in effecten van potentiële casemix adjusters (zie paragraaf 6.1.2). Voorts geeft een dataset van deze omvang een goed algemeen beeld van variatie tussen zorgaanbieders en kan berekend worden hoeveel respondenten per aanbieder nodig zijn om uiteindelijk significante verschillen te kunnen aantonen (Zorginstituut Nederland, 2013c). Als de variatie tussen zorgaanbieders beperkt is, kunnen bij het uiteindelijke gebruik van de indicator vele honderden respondenten per zorgaanbieder nodig zijn (De Boer et al., 2011). Dergelijke situaties kunnen consequenties hebben voor de bruikbaarheid van een indicator en komen aan de orde wanneer het onderscheidend vermogen wordt beoordeeld (zie paragraaf 6.2.3).

Aanbevelingen

Zorg voor een (pilot-)dataverzameling met in ieder geval 30 zorgaanbieders en 30 respondenten per zorgaanbieder om inzicht te krijgen in het onderscheidend vermogen en te kunnen berekenen hoeveel respondenten per zorgaanbieder uiteindelijk nodig zijn om onderscheid te kunnen maken tussen benedengemiddeld, gemiddeld en bovengemiddeld scorende aanbieders.

Kies als aggregatieniveau een niveau waarop verschillen worden verwacht. Includeer bij voorkeur verschillende aggregatieniveaus in de metingen om te kunnen toetsen op welk niveau verschillen het best zichtbaar worden.

Er zijn verschillende manieren om een indicator te definiëren die mogelijk verschillen in hun onderscheidend vermogen. Bekijk daarom bij voorkeur het onderscheidend vermogen van verschillende definiëringen van de indicator.

Het doorlopen van deze stappen draagt bij aan een indicator die instellingen kan onderscheiden in kwaliteit. Dit is relevant voor criterium 4 uit het Toetsingskader kwaliteitsstandaarden en meetinstrumenten (onderscheidend vermogen).

Meer lezen?

- In het *Handboek Eisen en Werkwijzen CQI-metingen* wordt voor de CQ-index beschreven hoe het onderscheidend en de populatievergelijkbaarheid kunnen worden getoetst. De werkinstructie *Vergelijkende analyses en casemix-adjustment* biedt hiervoor vele aanknopingspunten die ook bruikbaar zijn voor PROMs (Zorginstituut Nederland, 2014b).
- Meer informatie over interpretatie van verschillscores is te vinden in overzichtsstudies van Crosby et al. (2003), King (2011) en Wyrwich et al. (2013).
- Er zijn verschillende andere algemene documenten over methodologische kwesties omtrent PROMs bij het meten van kwaliteit van zorg: Cella et al. (2013), NQF (2013) en Deutsch et al (2012).
- De indicatorstandaard en het AIRE-instrument zijn documenten over het beoordelen van allerlei typen kwaliteitsindicatoren en kunnen ook worden gebruikt voor PROMs (De Koning et al., 2007; Koolman et al., 2012).

7 Aanbieden aan Register: besluitvorming, werkinstructie en aanbiedingsproces

7.1 *Besluiten om de PROM en indicator aan te bieden aan het Register*

Uiteindelijk dienen patiëntenorganisaties, zorgaanbieders en zorgverzekeraars te besluiten of zij gezamenlijk de PROM en de indicator willen aanbieden aan het Register. De informatie die in de voorgaande stappen is verzameld over de PROM en de indicator speelt uiteraard een belangrijke rol bij dit besluit, maar het is aan de partijen zelf om af te wegen of die informatie voldoende basis biedt om de PROM en de indicator aan te bieden. Hierbij is uiteraard ook denkbaar dat wordt gekozen voor doorontwikkeling, bijvoorbeeld om de vergelijkbaarheid of het onderscheidend vermogen te bevorderen alvorens over te gaan tot het aanbieden aan het Register.

Als patiëntenorganisaties, zorgaanbieders en zorgverzekeraars besluiten om de PROM en indicator gezamenlijk aan te bieden, vullen zij een aanbiedingsformulier in. Op grond van dit formulier toetst het Zorginstituut of de PROM en de indicator voldoen aan de criteria uit het Toetsingskader kwaliteitsstandaarden en meetinstrumenten. De informatie die nodig is om het aanbiedingsformulier in te vullen volgt grotendeels uit de voorgaande stappen. Daarnaast is een werkinstructie nodig (Criterium 3 uit het Toetsingskader).

7.2 *Werkinstructie*

De werkinstructie is een belangrijk instrument voor het borgen van registratievergelijkbaarheid. Box 7.1 geeft de minimale inhoud van een werkinstructie weer. De werkinstructies kunnen voor een belangrijk gedeelte worden gevuld met de keuzes die zijn gemaakt voor de opzet van een eerste (pilot-)meting. In de meeste gevallen leveren de eerste metingen belangrijke inzichten op voor de werkinstructies. Bijvoorbeeld als blijkt dat er meer respondenten per zorgaanbieder nodig zijn of als de indicator anders wordt gedefinieerd naar aanleiding van de resultaten. Kortom, de keuzes die belangrijk zijn voor de werkinstructie moeten (opnieuw) worden vastgesteld voor de werkinstructie die uiteindelijk samen met de PROM en de indicator wordt aangeboden aan het Zorginstituut.

Box 7.1

Minimale inhoud van een werkinstructie

De werkinstructie bevat ten minste de volgende onderdelen (Zorginstituut Nederland, 2014d)

- omschrijving van het doel van het meetinstrument;
- definitie over welke patiëntengroep gegevens worden verzameld (in- en exclusiecriteria);
- beschrijving van de bron, de dataverzamelmethode of registratiewijze;
- indien van toepassing: beschrijving welke informatiestandaard wordt gehanteerd bij de elektronische registratie van de voor dit meetinstrument noodzakelijke data;
- beschrijving van de benodigde correcties voor populatieverschillen;
- beschrijving van de benodigde data-analyse om te komen tot indicatorscores.

7.3 Het proces van aanbieden en toetsen

In de figuur op de volgende pagina staat een overzicht van het Zorginstituut waarin het toetsproces is weergegeven. De figuur heeft zowel betrekking op meetinstrumenten en indicatoren als op kwaliteitsstandaarden en bevat een globaal overzicht van het aanbiedings- en toetsproces. In het kort bestaat het proces uit het online aanmelden en indienen van de PROM en de indicator, tezamen met het aanbiedingsformulier en de werkinstructie. Het Zorginstituut wijst een contactpersoon toe en voert de toetsing uit, al dan niet met advies van de Adviescommissie Kwaliteit van het Zorginstituut. Goedgekeurde PROMs en indicatoren worden in het Register geplaatst. Als een PROM of indicator nog niet is goedgekeurd verschijnt het beoordelingsformulier in de kwaliteitsbibliotheek met daarin informatie die helpt om de PROM of de indicator te verbeteren.

Wanneer een indicator in gebruik is genomen en is opgenomen in het Register, is het uiteraard van belang deze periodiek te evalueren. Een voor de hand liggende manier om dit vorm te geven is een jaarlijkse evaluatie op grond van de data die dat jaar zijn verzameld en de ervaringen van de partijen in de zorg met het gebruik van die data. Dit zorgt ervoor dat wensen en mogelijkheden voor het verfijnen van de indicator, de dataverzameling en het gebruik van de indicator tijdig aan het licht komen. Tevens wordt op die manier duidelijk of blijvend wordt voldaan aan de criteria uit het Toetsingskader.

Hoe kom ik in het Register?

of De stappen voor het aanbieden van een kwaliteitsstandaard of meetinstrument



Start u binnenkort met het ontwikkelen van een kwaliteitsstandaard of meetinstrument, of wilt u een bestaande in het Register plaatsen? Dit stappenplan helpt u op weg.

Aanmelden en uploaden

- Elke organisatie die een kwaliteitsproduct (mede)beheert of ontwikkelt kan een account aanvragen op Zorginzicht.nl.
- Maak voor uw kwaliteitsproduct een pagina aan in de Kwaliteitsbibliotheek.
- Upload het kwaliteitsproduct en eventuele verwante producten bijvoorbeeld een cliëntenversie of meetinstructie door te drukken op de knop 'Dien in Register'. Er verschijnt een aanbiedingsformulier.

Algemene informatie

In dit onderdeel vult u de algemene gegevens in, bijvoorbeeld het onderwerp van de standaard of instrument, de betrokken partijen en een contactpersoon. Wat is de titel van het product? Welke partijen zijn betrokken bij de ontwikkeling van het product?

Kwaliteitsproducten worden op 4 criteria beoordeeld. De beoordeling geeft Zorginstituut Nederland inzicht in de toestand van het gebruik van het kwaliteitsproduct.

1

Criterium 1 Gezamenlijk ontwikkelen

Zijn alle relevante partijen betrokken bij de ontwikkeling? Wie zijn de vertegenwoordigers?

2

Criterium 2 Gezamenlijk voordragen

Is de procedure voor het gezamenlijk indienen vastgelegd? Dragen betrokken partijen het product gezamenlijk voor?

3

Criterium 3 Beschrijven zorgproces

Hoe zijn de inhoud van de zorg en de organisatie van het zorgproces beschreven in de kwaliteitsstandaard? Bevat het een beschrijving van het zorgproces vanuit het perspectief van de cliënt? Of is er een werkinstructie die beschrijft hoe het meetinstrument gebruikt dient te worden?

4

Criterium 4 Volledig en valide

Is de kwaliteitsstandaard volledig? Is er bijvoorbeeld een cliëntenversie, of is het meetinstrument betrouwbaar en valide?

Indienen

Contact & Planning

Binnen twee weken na indiening hoort u wie bij Zorginstituut Nederland uw contactpersoon is. De contactpersoon zorgt voor de verdere afhandeling van uw aanbestedingsproces.

Wist u dat...

we voor kwaliteitsstandaarden en meetinstrumenten de verzamelterm kwaliteitsproduct gebruiken?

Wist u dat...

u het aanbiedingsformulier niet in één keer hoeft in te vullen. Het kan ook tussentijds worden opgeslagen.

Wist u dat...

kwaliteitsstandaarden en meetinstrumenten die al voor 2014 zijn ontwikkeld onder een overgangsregeling vallen? In dat geval vervallen criteria 1 en 2. Het product moet wel ten minste door drie partijen voorgedragen worden. Criteria 3 en 4 moeten wel ingevuld worden.

Wist u dat...

u met vragen over het aanbestedingsproces altijd terecht kunt bij de helpdesk?

Mail naar:
helpdesk@zini.nl

Beoordeling

Zorginstituut Nederland beoordeelt het aanbiedingsformulier en de bijlagen. Indien nodig adviseert de Adviescommissie Kwaliteit over toelating tot het Register. Iedere beoordeling wordt door Zorginstituut Nederland genoteerd in een beoordelingsformulier. Ook als het product (nog) niet wordt goedgekeurd voor opname in het Register.

Ja, dan in het Register

Het kwaliteitsproduct komt samen met aanbiedingsformulier en beoordelingsformulier in het Register.

Nee, nog niet goedgekeurd

Dan komt het kwaliteitsproduct met het beoordelingsformulier in de kwaliteitsbibliotheek. In deze bibliotheek staan alle kwaliteitsproducten. Het beoordelingsformulier helpt u bij het verder ontwikkelen van het kwaliteitsproduct.

Aanbevelingen

Besluit met alle relevante partijen (patiëntenorganisaties, zorgaanbieders en zorgverzekeraars) of de PROM en indicator worden aangeboden aan het Register. Stel de werkinstructie op waarmee wordt voldaan aan criterium 3 uit het Toetsingskader. Vul het aanbiedingsformulier in met behulp van de informatie over de toetsingscriteria die vergaard is bij het uitvoeren van het stappenplan en biedt de PROM en de indicator aan voor het Register.

Meer lezen?

Gedetailleerde informatie over het aanbiedings- en toetsproces is te vinden op de websites van Zorginstituut Nederland, te weten: onder ‘Toetsingskader en Register’ op www.zorginstituutnederland.nl/kwaliteit en op www.zorginzicht.nl.

- Aaronson N, Choucair A, Elliott T, Greenhalgh J, Halyard M, et al.. *User's guide to implementing patient-reported outcomes assessment in clinical practice*. Milwaukee: International Society for Quality Life Research, 2011
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 2000; 25(24):3186-91
- Beatty PC, Willis GB. Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly*, 2007; 71(2):287-311
- Beatty PC. Answerable Questions: Advances in the Methodology for Identifying and Resolving Questionnaire Problems in Survey Research. *Dissertation Abstracts International, Doctoral Dissertation*, 2003; (64):9
- Boer D de, Delnoij D, Rademakers J. The discriminative power of patient experience surveys. *BMC Health Serv Res*, 2011; Dec 6;11:332
- Boer D de, Hoek L van der, Delnoij D, Groenewegen P. *Kleine zorgaanbieders in multilevel vergelijkende analyses: de CQI Verpleging, Verzorging en Thuiszorg*. Utrecht: NIVEL, 2010
- Brédart A, Marrel A, Abetz-Webb L, Lasch K, Acquadro C. (2014). Interviewing to develop Patient-Reported Outcome (PRO) measures for clinical research: eliciting patients' experience. *Health Qual Life Outcomes*, 2014; 2(1):15
- Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res*, 2009; 18(9):1263-78
- Buers C, Triemstra M, Bloemendal E, Zwijnenberg NC, Hendriks M, Delnoij DM. The value of cognitive interviewing for optimizing a patient experience survey. *Int J Soc Res Method*, 2013; 1-16
- Cella D, Hahn EA, Jensen SE, Butt Z, Nowinsky CJ, Rothrock N. (2013, 9/28/2013). *Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings*. Retrieved 12/19/2013, 2013, verkregen via http://www.qualityforum.org/Projects/n-r/Patient-Reported_Outcomes/Commissioned_Paper_1.aspx
- Cella D, Riley W, Stone A, Rothrock N, Reeve B, et al.. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*, 2010; 63(11):1179-94
- Coles J. *PROMS risk adjustment methodology - guide for general surgery and orthopaedic procedures*. Herts (UK): Northgate Information Solutions Ltd & CHKS Ltd., 2010
- Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*, 2003; 56(5):395-407
- Deutsch L, Smith L, Gage B, Kelleher C, Garfinkel D. *Patient-reported outcomes in performance measurement: commissioned paper on pro-based performance measures for healthcare accountable entities*. Washington, DC: National Quality Forum, 2012
- Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. *Health Econ*, 2010; 19(8):886-905
- Drenth PJD, Sijtsma K. *Testtheorie*. Houten: Bohn Stafleu Van Loghum, 2012
- Edelen MO, Reeve BB. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*, 2007; 16(1):5-18
- Embretson SE, Reise SP. *Item response theory for psychologists*. East Sussex: Psychology Press, 2000
- FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*, 2006; 4(1):79
- Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value in Health*, 2007; 10:S94-S105
- Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Boston:

- Sage Publications Ltd., 1991
- Hildon Z, Neuburger J, Allwood D, Meulen J van der, Black N. Clinicians' and patients' views of metrics of change derived from patient reported outcome measures (PROMs) for comparing providers' performance of surgery. *BMC Health Serv Res*, 2012; 12(1):171
- Hox JJ. *Multilevel analysis: Techniques and applications*. East Sussex: Psychology Press, 2002
- Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J Consult Clin Psychol*, 1999; 67(3):300
- Jaesckhe R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Controlled Clin Trials*, 1989; 10(4):407-15
- Kamper SJ, Ostelo RW, Knol DL, Maher CG, Vet HC de, Hancock MJ. Global Perceived Effect scales provided reliable assessments of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. *J Clin Epidem*, 2010; 63(7):760-6
- Kenniscentrum Meetinstrumenten. *Systematic reviews of measurement properties*. 2014, verkregen via http://www.kmin-vumc.nl/systematic-reviews-of-measurement-instruments_7_16.html
- King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*, 2011; 11(2):171-84
- Knafl K, Deatrick J, Gallo A, Holcombe G, Bakitas M, Dixon J, Grey M. The analysis and interpretation of cognitive interviews for instrument development. *Res Nurs Health*, 2007; 30(2):224-34
- Koning JS de, Kallewaard M, Klazinga NS. *Prestatie-indicatoren langs de meetlat—het AIRE instrument*. TSG, 2007; 85(5):261-4
- Koolman X, Zuidgeest M, Visser J, Appelman M. (2012). *Indicatorstandaard. Methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren in de zorg*. Den Haag: SiRM, 2012
- Kreft IGG. *Are multilevel techniques necessary? An overview, including simulation studies*. Unpublished Report, California State University, Los Angeles, 1996
- Kwaliteitsinstituut voor de Gezondheidszorg CBO, Nederlandse Patiënten Consumenten Federatie. *Blauwdruk patiëntenparticipatie in richtlijnontwikkeling. Leidraad voor richtlijnmakers*. Utrecht: CBO/NPCF, 2010
- Lasch KE, Marquis P, Vigneux M, Abetz L, Arnould B, et al.. (2010). PRO development: rigorous qualitative research as the crucial foundation. *Qual Life Res*, 2010; 19(8):1087-96
- Magasi S. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual Life Res*, 2012; 21(5)
- McKenna SP. Measuring patient-reported outcomes: moving beyond misplaced common sense to hard science. *BMC Medicine*, 2011; 9(1):86
- Meyer T, Richter S, Raspe H. Agreement between pre-post measures of change and transition ratings as well as then-tests. *BMC Med Res Methodol*, 2013; 13(1):52
- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, et al.. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*, 2010; 10(1):22
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford P@., et al.. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidem*, 2010; 63(7):737-45
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, et al.. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res*, 2009; 18(3): 313-33
- NQF. *Patient Reported Outcomes (PROs) in Performance Measurement*. 2013, verkregen via http://www.quality-forum.org/Publications/2012/12/Patient-Reported_Outcomes_Final_Report.aspx
- Nunnally JC. *Psychometric theory*. Noida (India): Tata McGraw-Hill Education, 2010
- Ostelo RW, Vet HC de, Vlaeyen JW, Kerckhoffs MR, Berfelo WM, et al.. Behavioral graded activity following first-time lumbar disc surgery: 1-year results of a randomized clinical trial. *Spine*, 2003; 28(16):1757-65
- Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, et al.. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding.

- Value in Health*, 2011; 14(8):978-88
- Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, et al.. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*, 2013; 1-17
- Reitsma J, Schipper M, Remmerswaal RE. *Statistisch betrouwbaar onderscheiden*. Barneveld: Significant, 2011
- Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health and Qual Life Outcomes*, 2006; 4(1):70
- Revicki DA, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidem*, 2008; 61(2):102-9
- Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*, 2002; 11(3):193-205
- Sijtsma K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 2009; 74(1):107-20
- Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med*, 1999; 48(11):1507-15
- Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, Vet HC de. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*, 2012; 21(4):651-7
- Vet HC de, Ostelo RWJG, Terwee CB, Roer N van der, Knol DL, et al.. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res*, 2007; 16(1), 131-42
- Vet HCW de, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide*. Cambridge University Press, 2011
- Vet HCW de, Terwee CB. The minimal detectable change should not replace the minimal important difference. *J Clin Epidem*, 2010; 63(7):804-5
- Willis GB. *Cognitive interviewing: A tool for improving questionnaire design*. New York: Sage, 2005
- Wyrwich K, Norquist J, Lenderking W, Acaster S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research*, 22(3), 475-483.
- Zaslavsky AM, Zaborski L, Cleary PD. Does the effect of respondent characteristics on consumer assessments vary across health plans? *Med Care Res Rev*, 2000; 57(3):379-94
- Zorginstituut Nederland. *Handboek Eisen en Werkwijzen CQI-metingen*. 2014a, verkregen via <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/handboek+cqi-metingen>
- Zorginstituut Nederland. *Handboek Ontwikkelen CQI-index*. 2014b, verkregen via <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/handboek+ontwikkelen+cqi-index>
- Zorginstituut Nederland. *PRO 03 Schriftelijke en/of online dataverzameling*. 2013a, verkregen via <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/handboek+cqi-metingen>
- Zorginstituut Nederland. *PRO 04 Mondelinge dataverzameling*. 2013b, verkregen via <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/handboek+cqi-metingen>
- Zorginstituut Nederland. *Procedure Cognitieve Interviews*. 2014c, verkregen via <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/handboek+ontwikkelen+cq-index#Cognitievetest>
- Zorginstituut Nederland. *Toetsingskader kwaliteitsstandaarden en meetinstrumenten*. Diemen: Zorginstituut Nederland, 2014d
- Zorginstituut Nederland. *Werkinstructie Kwalitatief onderzoek onder de doelgroep*. Diemen: Zorginstituut Nederland, 2014e
- Zorginstituut Nederland. *Werkinstructie: Vergelijkende analyses en casemix-adjustment*. 2013c, verkregen via <http://www.zorginstituutnederland.nl/kwaliteit/toetsingskader+en+register/cq-index/handboek+cqi-metingen>

Bijlage 1

Dankwoord

De handreiking is mede tot stand gekomen met input van patiëntenorganisaties, zorgaanbieders, zorgverzekeraars en methodologische experts. Deze input is vergaard middels een openbare raadpleging en discussiebijeenkomst, tijdens bijeenkomsten van expertgroepen en in enkele individuele contacten. De auteurs danken iedereen die middels schriftelijke of mondelinge reactie heeft bijgedragen aan deze handreiking.

- **Reacties openbare raadpleging**

Sebastiaan Baan (Landelijk Platform GGZ)

Sandra Beurskens (Zuyd Hogeschool/Maastricht University)

Kirsten Bornebroek (NPCF)

Myriam Crijns (v&vn)

Mostafa Filali (Pharos)

Ron van Heerde (Fysiotherapeut)

Mieke Hollander (Directeur Gezondheidszorg College – ROC Midden Nederland)

Noor Holst (Mediquest)

Renate Kieft (v&vn)

Helena Kosec (Pharos)

Perla Marang-van de Mheen (Leids Universiteit Medisch Centrum, afdeling Medische Besliskunde)

Wieneke Mekkink (Kenniscentrum Meetinstrumenten, vumc)

Pauline Poel (Achmea)

Heleen Post (NPCF)

Marleen Ploegmakers (Kennisinstituut van Medisch Specialisten)

Inge van der Putten (v&vn)

David Samoocha (NETQ Healthcare BV)

Vanessa Scholtes (Nederlandse Orthopaedische Vereniging)

Caroline Terwee (Kenniscentrum Meetinstrumenten, vumc)

Wendy Verhoeven (Stichting Miletus)

Wimjan Vink (Actiz)

Jolyn van Vuuren (cz)

Caroline van Weert (Stichting Miletus)

Philip van der Wees (IQ Healthcare)

Marloes Zuidgeest (Zorginstituut Nederland)

- ***Reacties tijdens bijeenkomsten van expertgroepen***
 Methodologische Expertgroep van Zorginstituut Nederland
 NFU Expertisenetwerk PROs
 Onderzoekersforum
 Wetenschappelijke Ontwikkelgroep PROMs (WOP; Stichting Miletus)

- ***Overige contacten en bijdragen***
 Sandra Beurskens (Zuyd Hogeschool/Maastricht University)
 Xander Koolman (Sirm/vu)
 Caroline Terwee (vumc/Kenniscentrum Meetinstrumenten)
 Wieneke Mokkink (vumc/Kenniscentrum Meetinstrumenten)

Bijlage 2

Onderzoeksopzet en meetmomenten

Het doel van PROMS-metingen is doorgaans om te bepalen of en in hoeverre patiënten verbeteren, verslechteren of geen verandering laten zien. Een belangrijk vraagstuk is of dit in kaart wordt gebracht met herhaalde metingen of dat patiënten op één moment worden gevraagd naar de uitkomst nu en in het verleden. Daarnaast kan het van belang zijn op welk moment in het zorgproces de meting plaatsvindt.

Onderzoeksopzet, response shift & recall bias

Er zijn verschillende designs of methoden van dataverzameling mogelijk met PROMS, elk gekenmerkt door specifieke voor- en nadelen en eventuele voorwaarden voor wat betreft de uitvoering en interpretatie van het onderzoek.

Het in kaart brengen van verandering kan op drie manieren worden gedaan, namelijk met een directe verschilmeting, pre-postmeting of met een then-test (Meyer et al., 2013).

Onder een **directe verschilmeting** vallen bijvoorbeeld ‘*transition ratings*’ met een ‘*global rating of change*’ zoals de *Global Perceived Effect vraag* (GPE) (Kamper et al., 2010). Hiermee wordt de patiënt op een directe wijze gevraagd naar veranderingen in de uitkomst; of deze is verslechterd, verbeterd of gelijk is gebleven ten opzichte van een eerder moment. Deze methode heeft als voordeel dat het relatief snel en gemakkelijk te bevragen, registreren en interpreteren is. Een nadeel is dat het een beroep doet op de herinnering van de patiënt die mogelijk niet geheel accuraat is. Daarnaast zijn dit soort vragen relatief lastig voor patiënten omdat een patiënt in gedachten een inschatting moet maken van de huidige situatie, een vroegere situatie en het verschil daartussen.

Bij de **pre-postmeting** wordt eenzelfde patiëntengroep op verschillende momenten in de tijd gevraagd naar een bepaalde uitkomstindicator. Er wordt dan een baseline-meting gedaan en een meting na de interventie. Door het verschil te berekenen wordt een eventuele verbetering of verslechtering zichtbaar. Hiervoor is het opvolgen van patiënten in de tijd noodzakelijk en is de belasting van patiënten groter vanwege de twee opeenvolgende metingen in de tijd. Daarnaast kan er bij gebruik van deze methode response shift optreden. Response shift is gedefinieerd als een verandering in de evaluatie van iemands kwaliteit

van leven als gevolg van veranderingen in interne standaarden, waarden en de persoonlijke interpretatie/opvattingen van kwaliteit van leven op zich (Sprangers en Schwartz, 1999). Dit is een nadeel omdat het hierdoor niet meer duidelijk is of en in welke mate een verandering in een PROM veroorzaakt is door het effect van een interventie, response shift of beide.

Tot slot is er de **then-test**. Hierbij worden na de interventie/behandeling twee vragen gesteld. Ten eerste wordt retrospectief (terugkijkend) gevraagd naar de status van de patiënt vóór de interventie. Ten tweede wordt gevraagd naar de status op het meetmoment zelf. Er wordt dan voor response shift gecorrigeerd omdat er impliciet vanuit wordt gegaan dat de patiënt retrospectief een inschatting maakt van de twee momenten en daarmee dus dezelfde interne standaard hanteert voor beide momenten. Op die manier wordt dus automatisch gecorrigeerd voor de status van het moment van bevragen. Ook hier geldt echter weer dat de herinnering van de patiënt vertroebeld kan zijn (*recall bias*), wat van invloed is op de retrospectieve voormeting.

De keuze tussen voornoemde strategieën is niet eenvoudig omdat ze allemaal hun voor- en nadelen hebben. Over het algemeen worden de pre-post meting en then-test geprefereerd boven de directe verschilmeting, maar een afweging tussen de pre-post meting en de then-test vereist enige onderbouwing dat response-shift in de betreffende setting een groter of juist een kleiner probleem is dan *recall bias*. Die onderbouwing is in veel gevallen lastig te geven. In tabel B2.1 wordt een overzicht gegeven van verschillende meetmethoden en de bijbehorende voordelen, nadelen en voorwaarden.

Tabel B2.1

Methoden voor het meten van verschillen in uitkomsten van zorg met PROMs

Methoden	Direct	Pre-postmeting	Then-test
	Retrospectieve meting waarbij op een directe manier wordt gevraagd naar de verandering (verbetering/verslechtering of geen verandering) door of sinds de zorg(interventie). Dit kan met een enkele vraag (' <i>global rating of change</i> '), zoals de <i>General Perceived Effect</i> vraag (GPE), of met meerdere vragen naar klachten/symptomen, functioneren of kwaliteit van leven (niveaus).	Prospectief, longitudinaal cohort onderzoek met twee metingen: voor- en nameting (pre- en post-interventie) voor de evaluatie van een specifieke behandeling of ingreep (operatie, medicatie, etc.).	Retrospectieve meting met tenminste twee vragen over de situatie voor en na de interventie (pre- en post-interventie status).
Voordelen	Deze eenmalige meting met één of enkele vragen aan de patiënt betekent een minimale belasting voor de patiënt. Bovendien wordt direct een indruk verkregen van de verandering volgens de patiënt.	Er kan een verschil berekend worden op het niveau van afzonderlijke variabelen of schalen. De absolute scores zijn afzonderlijk te interpreteren en kunnen vergeleken worden met andere doel-/ referentiegroepen en metingen over de tijd.	Twee scores voor verschillende momenten, maar niet de logistieke en administratieve nadelen die bij de pre-post methode horen.
Nadelen	Recall bias en response shift is van invloed op de mate van verandering.	Het prospectief design stelt veel administratieve en logistieke eisen aan de dataverzameling. Door verandering in omstandigheden, is het ingewikkeld om KvL te meten omdat dit een constant referentiepunt vereist. Response shift is van invloed op de mate van verandering. Een pre-meting is niet bij elke patiëntengroep en in elke zorgsetting mogelijk (bijvoorbeeld niet bij chronische patiënten en acute zorg).	Patiënten moeten zich kunnen herinneren hoe hun status vóór de zorginterventie en/of terugkijkend in de tijd was. Deze herinneringen zijn vatbaar voor vertekeningen door de huidige (gezondheids)status van een patiënt. Recall bias is van invloed op de mate van verandering. Onduidelijk hoe response shift zich verhoudt tot retrospectieve metingen.
Voorwaarden	Het pre-post interval moet kort genoeg zijn zodat de patiënt in staat is om de pre-status te herinneren, en lang genoeg om een verandering te kunnen meten. Het pre-moment moet duidelijk en voor iedereen hetzelfde gedefinieerd zijn, bijvoorbeeld: ten opzichte van voor de behandeling/operatie of in vergelijking met ... maanden geleden.	Meetmomenten moeten zo gekozen zijn dat er daadwerkelijk veranderingen kunnen worden gemeten.	Het pre-post interval moet kort genoeg zijn zodat de patiënt in staat is om de pre-status te herinneren, en lang genoeg om een verandering te kunnen meten. Het pre-moment moet duidelijk en voor iedereen hetzelfde gedefinieerd zijn.

Vaststellen meetmomenten

Het is belangrijk om vast te stellen wanneer en hoe vaak patiënten een PROM moeten invullen. De verschillende opties hebben elk voor- en nadelen (Aaronson et al., 2011). Bij een enkele meting is de timing van groot belang voor het gemeten 'effect' en is het wenselijk dat de meting voor alle patiënten op eenzelfde (tijds)periode betrekking heeft. Als er meerdere meetmomenten zijn, is het eveneens van belang dat die op relevante momenten plaatsvinden (wanneer er daadwerkelijk effecten zijn te verwachten). Voor toepassing in de klinische praktijk is het wenselijk om de metingen te laten samenvallen met de contactmomenten van patiënt en behandelaar, zodat de resultaten kunnen worden gebruikt voor het evalueren en zo nodig bijstellen van een behandeling. Voor andere doeleinden – zoals de kwaliteitscontrole/-verbetering door zorgaanbieders, benchmarking en de zorginkoop door zorgverzekeraars – volstaat wellicht één retrospectieve meting voor een evaluatie op groeps-/organisatieniveau of voor een vergelijking van zorgaanbieders.

Dat het meetmoment van invloed kan zijn op de uitkomst, kan aanleiding zijn om dat meetmoment in ieder geval zo constant mogelijk te houden tussen patiënten en tussen zorgaanbieders. Tegelijkertijd heeft enige variatie in meetmomenten ook voordelen, want dit maakt het mogelijk om de eventuele invloed van het meetmoment op de uitkomst te analyseren, bijvoorbeeld met behulp van correlatie- en/of regressie-analyse. Indien blijkt dat het meetmoment van invloed is op de uitkomst kan hier nog rekening mee worden gehouden bij het vergelijken van zorgaanbieders door het meetmoment te betrekken bij de casemix adjustment.



>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<

>±{☑}±<

>±{☒}±<