



# Handreiking

Voor de selectie van  
PROs en PROMs



NEDERLANDSE FEDERATIE VAN UMC'S  
CONSORTIUM KWALITEIT VAN ZORG



# Handreiking

## voor de selectie van PROs en PROMs

Caroline B. Terwee, Philip J. van der Wees en Sandra Beurskens,  
namens het NFU-expertisenetwerk Patient-reported outcomes





## Inhoudsopgave

1.	Inleiding.....	6
1.1	Aanleiding .....	6
1.2	Doelstelling .....	6
1.3	Aanpak .....	7
1.4	Leeswijzer.....	8
2.	PROs en PROMs in het gezondheidszorgbeleid: transparantie van kwaliteit... 10	
2.1	Ontwikkeling van uitkomstindicatoren .....	10
3.	Het stappenplan .....	12
3.1	Het selecteren van PROs .....	12
3.2	Het selecteren van PROMs.....	12
4.	Uitwerking van het stappenplan voor de selectie van PRO en PROMs.....	14
4.1	Stap 1: Overwegingen vooraf: bepaal de context.....	14
4.2	Stap 2: Bepaal op welk niveau men wil meten .....	15
4.2.1	Overwegingen bij de keuzes voor modellen .....	17
4.3	Stap 3: Selecteer relevante PROs per niveau .....	19
4.3.1	Criteria voor het selecteren van PROs .....	22
4.4	Stap 4: Stel de eisen vast aan de PROMs.....	23
4.5	Stap 5: Identificeer alle beschikbare PROMs .....	24
4.6	Stap 6: Maak een voorselectie op basis van inhoud .....	25
4.7	Stap 7: Bepaal de kwaliteit van geselecteerde PROMS .....	26
4.8	Stap 8: Bepaal de hanteerbaarheid van geselecteerde PROMS.....	31
4.9	Stap 9: Selecteer de beste PROMs.....	32
5.	Implementatie van de tools: hoe nu verder?.....	34
	Dankwoord.....	35
	Referenties .....	36
	Bijlage 1 .....	40
	Bijlage 2 .....	43



## 1. Inleiding

### 1.1 Aanleiding

Patiëntgerapporteerde uitkomstmetingen (in het Engels: Patient-Reported Outcome Measures, PROMs) worden steeds vaker en voor steeds meer doelen gebruikt. Ten eerste worden PROMs in wetenschappelijk onderzoek ingezet om effecten van interventies vanuit het perspectief van patiënten in kaart te brengen; veel PROMs zijn oorspronkelijk voor die doelstelling ontwikkeld. Ten tweede worden PROMs in de directe zorgpraktijk gebruikt om meer inzicht te krijgen in het functioneren of de kwaliteit van leven van patiënten. Ze kunnen ondersteunen bij het stellen van een diagnose, het evalueren van een behandeling of om de communicatie te verbeteren. Ten derde, en meer recent, worden PROMs ingezet op geaggregeerd niveau om kwaliteit van zorg te meten en te bevorderen op afdelingen, en te vergelijken tussen afdelingen en tussen organisaties. Zorgverzekeraars willen deze informatie gebruiken om zorginkoop op af te stemmen.

Het gebruik van PROMs voor de verschillende doelen doet een enorm beroep op patiënten om de vragenlijsten in te vullen en op professionals om de resultaten te verwerken en te gebruiken. Men streeft daarom naar een geïntegreerd gebruik van PROMs voor de patiëntenzorg, onderzoek en voor kwaliteit van zorg metingen. Dit is echter niet eenvoudig omdat elk doel vaak andere eisen stelt met betrekking tot de relevante domeinen die men wil meten (ook wel Patient-Reported Outcomes genoemd, PROs), de meeteigenschappen en hanteerbaarheid van de PROMs en de mate van betrokkenheid van de zorgverlener.

De keuze van 'wat' men wil meten (PROs) en 'hoe' men wil meten (met welke vragenlijsten, schalen, PROMs), is daarom een ingewikkeld maar wel heel belangrijk proces. Nog te vaak worden PROMs toegepast zonder stil te staan bij wat ze meten en voor welke doelen ze geschikt zijn. Er is een wildgroei aan vragenlijsten om uit te kiezen en het is vaak niet duidelijk welke het beste is. Daarnaast is er onduidelijkheid over terminologie (bijvoorbeeld over de begrippen construct, domein, concept), waardoor bestaande handreikingen en leidraden voor de selectie van PROs en PROMs soms moeilijk te interpreteren zijn. Bovendien zijn de te nemen stappen vaak nog te summier uitgewerkt en wordt onvoldoende aandacht besteed aan de dilemma's die optreden als PROs en PROMs voor meerdere doelen tegelijkertijd ingezet worden. Er is behoefte aan ondersteuning zodat de keuze van PROs en PROMs gestructureerd en transparant wordt.

### 1.2 Doelstelling

Deze handreiking beschrijft de selectie van PROs en PROMs aan de hand van een stappenplan. De doelgroepen voor deze handreiking zijn inhoudelijke en methodologische deskundigen die PROs en PROMs willen selecteren voor

toepassing in de zorg, voor wetenschappelijk onderzoek, en/of voor het meten, verbeteren en transparant maken van de kwaliteit van zorg.

In deze handreiking worden de methodologische en praktische dilemma's beschreven die een rol spelen bij de selectie van PROs en PROMs voor verschillende doelen: zorgpraktijk, wetenschappelijk onderzoek, en kwaliteit van zorg metingen. Het betreft een eerste versie die in een vervolgtraject uitgewerkt zal worden tot een uitgebreide toolkit.

Deze handreiking is uitgewerkt analoog aan de opzet van de Handleiding en Toolbox Richtlijnontwikkeling in de Nederlandse Gezondheidszorg (HARING):

- [www.ha-ring.nl](http://www.ha-ring.nl)
- [www.zonmw.nl/uploads/tx\\_vipublicaties/De\\_HARING-tools.pdf](http://www.zonmw.nl/uploads/tx_vipublicaties/De_HARING-tools.pdf)

Doelstelling is om tot een vergelijkbare toolbox te komen voor de selectie en toepassing van PROs en PROMs. Vooralnog voorzien we dat de uiteindelijke toolbox uit drie onderdelen zal bestaan: (a) selectie van PROs en PROMs, (b) (door)ontwikkeling van PROMs, en (c) ontwikkeling van kwaliteitsindicatoren op basis van PROMs metingen. Deze handreiking richt zich op deel (a). Voor een verdere verdieping en ontwikkeling van de toolbox dienen we een subsidieaanvraag in bij relevante instanties.

### 1.3 Aanpak

De methoden die gebruikt zijn in deze handreiking zijn gebaseerd op literatuuronderzoek en expert opinion van de leden van het NFU-expertisenetwerk Patient-Reported Outcomes (PRO). Het NFU-expertisenetwerk PRO streeft ernaar om bestaande documenten bij elkaar te brengen tot één geïntegreerde handreiking. Het literatuuronderzoek is gebaseerd op bestaande zoekstrategieën, richtlijnen en boeken. De hieronder beschreven bronnen zijn hierbij gebruikt.

Bestaande zoekstrategieën:

Zoekstrategie die gebruikt is voor het ontwikkelen van een richtlijn voor de selectie van 'outcome measurement instruments for Core Outcome Sets' [1].

Zoekstrategie die gebruikt is door de International Society for Quality Of Life Research (ISOQOL) voor de ontwikkeling van 'minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research' [2].

Richtlijnen van professionele organisaties:

- ISOQOL standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research' (criteria voor goede meeteigenschappen) [2].
- The COMET guidelines for the development of Core Outcome Sets (richtlijnen voor de ontwikkeling van Core Outcome Sets) [3].



- The Outcome Measures in Rheumatology (OMERACT) handbook (richtlijnen voor de ontwikkeling van Core Outcome Sets) [4,5].
- Richtlijn voor de ontwikkeling van ICF (International Classification of Functioning, Disability and Health) Coresets [6].
- Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures. NIVEL & Zorginstituut [7].
- American Medical Association-convened Physician Consortium for Performance Improvement (PCPI): Standards for developing and evaluating patient-reported outcome (PRO) performance measures [8].
- National Quality Forum (NQF): Methodological Issues In The Selection, Administration And Use Of Patient-Reported Outcomes In Performance Measurement In Health Care Settings [9].
- National Quality Forum (NQF): Patient-Reported Outcomes in Performance Measurement. Commissioned Paper on PRO-Based Performance Measures for Healthcare Accountable Entities [10].
- National Quality Forum (NQF): Patient Reported Outcomes (PROs) in Performance Measurement [11].

Adviezen van beroepsverenigingen:

- Advies Nederlandse Orthopaedische Vereniging [12].

Boeken:

- De Vet et al. Measurement in medicine (methodologie vragenlijstontwikkeling en validatieonderzoek) [13].
- Beurskens et al. Meten in de praktijk (stappenplan voor het gebruik van meetinstrumenten in de gezondheidszorg) [14].
- Streiner DL & Norman GR. Health measurement scales. A practical guide to their development and use (methodologie vragenlijstontwikkeling en validatieonderzoek) [15].

#### 1.4 Leeswijzer

Hoofdstuk 2 schetst de inbedding van PROs en PROMs in het gezondheidszorgbeleid voor het meten, verbeteren en transparant maken van kwaliteit. In Hoofdstuk 3 wordt het stappenplan beschreven waarbij een onderscheid gemaakt wordt in enerzijds het selecteren van PROs en anderzijds het selecteren van PROMs. De verschillende stappen grijpen op elkaar in en beschrijven een iteratief proces waarbij de verschillende stappen elkaar beïnvloeden en men ‘heen en weer’ kan gaan tussen de stappen tijdens de selectie van PROs en PROMs. In Hoofdstuk 4 volgt per stap een uitwerking van de aanpak en dilemma’s bij het ontwikkelen van tools. Deze handreiking is gebaseerd op huidige methodologische en praktische inzichten. Niet alle stappen zijn echter voldoende wetenschappelijk onderbouwd. Daarom volgen in Hoofdstuk 5 aanbevelingen voor een onderzoeksagenda met prioriteiten voor toekomstig onderzoek



## 2. PROs en PROMs in het gezondheidszorgbeleid: transparantie van kwaliteit

In het huidige (inter)nationale beleid in de zorg neemt het meten, verbeteren en transparant maken van kwaliteit op basis van patiëntgerapporteerde uitkomsten een belangrijke plaats in. Hiervoor kunnen uitkomstindicatoren worden ontwikkeld die kwaliteitsverschillen in de zorg kunnen signaleren. Het transparant maken van de zorg kan dus één van de doelen van het gebruik van PROMs zijn.

Idealiter vindt de dataverzameling met behulp van PROMs plaats in de zorg. Op die manier kunnen PROMs worden ingezet binnen het primaire proces van de zorg én kunnen de verzamelde data geaggregeerd worden op groepsniveau voor het meten, verbeteren en transparant maken van kwaliteit. Het voordeel daarvan is dat slechts één keer gemeten hoeft te worden voor meerdere doelen. Daarnaast heeft de feedback van uitkomsten op basis van metingen met PROMs meer betekenis als de data uit het primaire proces zelf voortkomen [18–20]. De NFU stimuleert de dataverzameling in de zorg via haar beleid ‘registratie aan de bron’.

Er zijn echter ook nadelen verbonden aan het integreren van verschillende doelen in het meten van PROs. De eisen voor betrouwbaarheid en validiteit van data om de kwaliteit van zorgaanbieders met elkaar te vergelijken zijn hoog. Het openbaar publiceren van uitkomsten kan consequenties hebben voor aanbieders van zorg, waardoor vergelijkbaarheid van uitkomsten van groot belang is. Om de kwaliteit van zorgaanbieders met elkaar te kunnen vergelijken is de beschikbaarheid van betrouwbare en valide data van groot belang. Data die in het primaire proces worden verzameld zijn gevoelig voor vertekening als gevolg van verschillen in werkprocessen en routines. Dergelijke verschillen zullen beperkt moeten blijven tot een bepaalde bandbreedte om uitkomsten daadwerkelijk vergelijkbaar te maken.

### 2.1 Ontwikkeling van uitkomstindicatoren

Uitkomstindicatoren hebben als doel om de kwaliteit van de zorg inzichtelijk te maken op basis van uitkomsten van de zorg die door aanbieders wordt geleverd. Daarmee kunnen eventuele verschillen in kwaliteit worden gesignaleerd. Voor aanbieders zelf kunnen uitkomstindicatoren worden gebruikt om hun kwaliteit te verbeteren. Indien de uitkomsten openbaar worden gepubliceerd kunnen externe partijen de uitkomsten gebruiken. Patiënten kunnen de uitkomsten gebruiken om aanbieders van zorg te vergelijken. Zorgverzekeraars kunnen uitkomsten gebruiken voor zorginkoop. PROMs die gebruikt worden als uitkomstindicatoren dienen betrouwbaar en valide te zijn en moeten grofweg aan drie criteria voldoen: inhoudsvaliditeit, vergelijkbaarheid en statistisch betrouwbaar onderscheiden. Voor een uitgebreide toelichting op deze criteria verwijzen we

naar de Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures en naar de Indicatorstandaard [7,21].

### *Inhoudsvaliditeit*

Inhoudsvaliditeit heeft betrekking op de relevantie en volledigheid van de PROM als uitkomstindicator. Eén van de aspecten van inhoudsvaliditeit is de relevantie van de items voor het doel van de meting. Dit betekent dat moet zijn aangetoond dat de gemeten uitkomst beïnvloedbaar is door de zorgaanbieder(s) waar de indicator betrekking op heeft.

### *Vergelijkbaarheid*

Het is essentieel dat de indicatorwaarden van verschillende zorgaanbieders op een bepaalde indicator vergelijkbaar zijn. Er zijn drie vormen van vergelijkbaarheid te onderscheiden die kunnen leiden tot vertekening (bias) wanneer hier niet aan wordt voldaan:

- Registratievergelijkbaarheid: Het proces van meten, registreren en aanleveren van de voor de indicator benodigde gegevens dient juist, volledig en uniform en zonder systematische meetfouten te gebeuren.
- Populatievergelijkbaarheid: De vergelijkbaarheid van indicatorwaarden wordt niet beïnvloed door verschillen in populatiekenmerken.
- Steekproef- en responsvergelijkbaarheid: De vergelijkbaarheid van indicatorwaarden wordt niet beïnvloed door verschillen tussen de steekproef en de totale populatie.

### *Statistisch betrouwbaar onderscheiden*

Een indicator dient het vermogen te hebben om zorgaanbieders met bovengemiddelde en ondergemiddelde indicatorwaarden te onderscheiden van gemiddeld scorende aanbieders. Bij dit criterium speelt het aantal waarnemingen dat een zorgaanbieder kan aanleveren een essentiële rol: bij een te laag aantal waarnemingen wordt de rol van toeval veelal te groot om betrouwbaar verschillen in prestaties te kunnen detecteren. Dit is een belangrijk punt bij het interpreteren van de resultaten op instelling c.q. behandelaar niveau. Een vuistregel uit de handreiking PROMs is dat er minimaal 30 waarnemingen per praktijk nodig zijn om statistisch betrouwbaar te onderscheiden [7].

### 3. Het stappenplan

Het totale stappenplan bestaat uit negen stappen (zie Figuur 1).

#### 3.1 Het selecteren van PROs

Patient-Reported Outcomes (PROs) zijn uitkomsten van zorg die betrekking hebben op de gezondheid van de patiënt en die de patiënt zelf rapporteert zonder tussenkomst van een zorgverlener [16]. Voorbeelden van PROs zijn pijn, vermoeidheid, beperkingen in dagelijks functioneren, angst, depressie, kwaliteit van leven. Het betreft allemaal concepten die niet 'objectief' waarneembaar zijn en daarom alleen gemeten kunnen worden door het aan de patiënt te vragen. Het selecteren van PROs kan het beste gebeuren op basis van de volgende stappen:

Stap 1: Overwegingen vooraf: bepaal de context

Stap 2: Bepaal op welk niveau men wil meten

Stap 3: Selecteer relevante PROs per niveau<sup>1</sup>

#### 3.2 Het selecteren van PROMs

Patient-reported outcome measures (PROMs) zijn de vragenlijsten die gebruikt worden om PROs te meten. PROMs kunnen generiek of specifiek van aard zijn. Generieke vragenlijsten richten zich op gezondheidserelateerde kwaliteit van leven los van de onderliggende aandoening, zoals de Short Form Health Survey (SF-36) of de EuroQol (EQ-5D). Specifieke vragenlijsten richten zich op een bepaalde ziekte of lokalisatie. Voorbeelden daarvan zijn de Quebec Pain Disability Scale (QBPDS) voor mensen met rugklachten, de Parkinson Disability Questionnaire (PDQ) voor patiënten met Parkinson, of de Disabilities of the Arm, Shoulder and Hand (DASH) voor mensen met klachten aan de bovenste extremiteit. Individuele PROMs richten zich op het meten van specifieke, door de patiënt zelf in te vullen, problemen. Voorbeelden daarvan zijn de Patiënt Specifieke Klachten lijst en de Schedule for the Evaluation of Individual Quality of Life (SEIQoL) [17].

De selectie van PROMs vindt plaats op basis van de beschikbare PROMs, hun meeteigenschappen en een aantal haalbaarheidsfactoren. Het selecteren van PROMs kan het beste gebeuren op basis van de volgende stappen:

Stap 4: Stel de eisen vast aan de PROMs

Stap 5: Identificeer alle beschikbare PROMs

Stap 6: Maak een voorselectie op basis van inhoud

Stap 7: Bepaal de kwaliteit van geselecteerde PROMs

Stap 8: Bepaal de hanteerbaarheid van geselecteerde PROMs

Stap 9: Selecteer de beste PROMs.

---

<sup>1</sup> De leden van het NFU-expertisenetwerk PRO vonden het woord 'niveau' niet de optimale term, maar er was geen consensus over een betere term.

Figuur 1: Stappenplan voor de selectie van een PRO of een PROM



## 4. Uitwerking van het stappenplan voor de selectie van PRO en PROMs

### 4.1 Stap 1: Overwegingen vooraf: bepaal de context

Om tot een goede selectie van relevante PROs te kunnen komen, is het belangrijk om de context waarbinnen de metingen plaats zullen vinden van te voren te bepalen en vast te leggen (Tabel 1). Men dient hierbij zo specifiek mogelijk te zijn. De beantwoording van de volgende vragen biedt ondersteuning bij de definiëring van de context:

#### *Wat is het doel van het meten van de PRO?*

Leg vast voor welk doel de PRO gemeten zal gaan worden: voor de zorg, voor onderzoek, voor het evalueren van kwaliteit van zorg, of voor een combinatie. Denk daarbij aan waar de metingen voor zullen worden gebruikt. Bijvoorbeeld: standaardisatie van het consult, (ondersteunen bij) het stellen van een diagnose of prognose, ondersteunen bij het opstellen van een individueel behandeldoel en behandelplan (shared decision-making), evalueren van een behandeling bij individuele patiënten in de zorg, evalueren van een behandeling bij een groep patiënten om inzicht te verkrijgen in het eigen handelen van de zorgverlener, of verkrijgen van inzicht in het effect van zorg van een afdeling of instelling, en evalueren van een diagnostische strategie of behandeling bij een groep patiënten in onderzoek. Er kunnen ook meerdere doelen zijn. Leg hierbij ook vast of de uitkomsten op individueel niveau zullen worden gebruikt (bijvoorbeeld in de zorg) of op groepsniveau (bijvoorbeeld in onderzoek, of bij benchmarking) of allebei.

#### *Wie wordt gemeten?*

Leg de doelpopulatie vast, inclusief een duidelijke beschrijving van de ziekte of de klacht. Bijvoorbeeld: patiënten met heup of knie artrose, vastgesteld volgens de ACR criteria (classificatie criteria voor reumatische aandoeningen), die in aanmerking komen voor een nieuwe heup of knie.

#### *In welke setting vinden de metingen plaats?*

Leg vast waar de metingen zullen plaatsvinden. Bijvoorbeeld in het ziekenhuis of in een eerstelijnspraktijk. In een ziekenhuis ziet men vaker patiënten met ernstiger problematiek dan in de eerste lijn. Hiermee moet rekening worden gehouden bij de keuze van de uitkomstmaat.

Tabel 1: Checklist voor het beschrijven van de context van de PRO metingen

Context van PRO metingen	
Wat is het doel van de meting?	Wetenschappelijk onderzoek, zorgverlening (primaire proces), (interne) kwaliteitsverbetering, transparantie van zorg, zorginkoop. Metingen zullen op individueel of groepsniveau worden gebruikt
Wie wordt gemeten?	Specificatie van populatie en eventuele subcategorieën daarin
In welke setting vinden de metingen plaats?	Eerste lijn, tweede lijn, derde lijn Specificatie zorgverleners, monodisciplinair, multidisciplinair

#### 4.2 Stap 2: Bepaal op welk niveau men wil meten

Voor het selecteren van PROs wordt aanbevolen om een conceptueel model als uitgangspunt te nemen, waarin de belangrijkste domeinen van gezondheid of kwaliteit van leven zijn weergegeven op verschillende niveaus. Zo'n model biedt houvast bij het selecteren van PROs omdat ze laten zien dat je op verschillende niveaus kunt meten. Er bestaan verschillende modellen in de gezondheidszorg waarmee domeinen van gezondheid en kwaliteit van leven in kaart kunnen worden gebracht. Deze modellen zijn vaak ontwikkeld binnen een specifiek vakgebied met een eigen jargon en begrippenkader. De terminologie die in de verschillende modellen en schema's wordt gehanteerd verschilt sterk, mede als gevolg van verschillen in de context, (para)medische discipline en andere achtergrondkenmerken van de ontwikkelaars van modellen. We maken in deze handreiking geen expliciete keuze voor een van de modellen, maar proberen een verbinding te leggen in de begrippenkaders van de verschillende modellen. Voor de selectie van PROs is het van belang dat gebruikers vanuit verschillende achtergronden inzicht hebben in de verschillen in terminologie en elkaars taal begrijpen.

In de Bijlage 1 hebben we een aantal relevante modellen voor gezondheid en kwaliteit van leven op een rij gezet. Wij hebben gekozen voor modellen op het gebied van functioneren en niet op het gebied van ziekte, zoals de ICD, omdat het bij PROs niet gaat om het beschrijven van de ziekte zelf, maar om de impact van ziekte op het (ervaren) functioneren van patiënten.

Voorbeelden van relevante modellen zijn het schema van de International Classification of Functioning, Disability and Health (ICF) van de World Health Organization (Figuur 3) [22], het model van Wilson & Cleary (Figuur 4) [23], en het model van Spilker (Figuur 5) [24]. Het model van Spilker was één van de eerste modellen voor kwaliteit van leven onderzoek en werd in de jaren '90 van de



vorige eeuw veel geciteerd. Tegenwoordig wordt dit model niet veel meer gebruikt. Het Patient-Reported Outcome Measurement Information System (PROMIS) is een nieuw, betrouwbaar en efficiënt meetsysteem voor PROs, dat naar verwachting wereldwijd steeds vaker gebruikt zal gaan worden. PROMIS heeft een eigen conceptueel model van PROs ontwikkeld, waarin onderscheid gemaakt wordt tussen fysieke, mentale en sociale aspecten van gezondheid (Figuur 6) ([www.nihpromis.org](http://www.nihpromis.org)). De internationale OMERACT groep (Outcome Measures in Rheumatology) heeft een model ontwikkeld voor het selecteren van uitkomsten voor Core Outcome Sets (COS, een afgesproken set uitkomsten die in elke clinical trial in een bepaalde doelpopulatie wordt gemeten en gerapporteerd) [4,5]. Dit model is mede gebaseerd op het ICF schema en het model van Wilson & Cleary (Figuur 7). In een systematische review van Bakas et al. zijn verschillende modellen met elkaar vergeleken waaruit blijkt dat er twee modellen zijn waarin het model van Wilson & Cleary en de ICF zijn geïntegreerd [25]. Valderas en Alonso hebben geprobeerd het model van Wilson & Cleary en het ICF schema samen te voegen tot één geïntegreerd model voor PROs (Figuur 8) [26].

Idealiter wordt een gezamenlijk model en terminologie voor de indeling van PROs gebruikt. Het is echter lastig om modellen die vanuit verschillende vakgebieden zijn ontwikkeld bij elkaar te brengen in een gezamenlijk model. Elk model heeft een eigen context en doelstelling. Daarom beperken we ons tot het geven van inzicht in de verschillen en overeenkomsten tussen de modellen. In Paragraaf 4.2.1 bespreken we voor- en nadelen van de modellen ter overweging van te maken keuzes. In Tabel 2 wordt een overzicht gegeven van de verschillende niveaus op basis van alle bovenstaande modellen met voorbeelden van relevante PROs per niveau. Dit overzicht kan als uitgangspunt worden genomen in het selectieproces van relevante PROs. Het geeft de 'grootste gemene deler' weer tussen gehanteerde begrippen in de verschillende modellen, zonder dat we de pretentie hebben om volledig en sluitend te zijn.

Tabel 2: Niveaus waarop PROs gemeten kunnen worden en voorbeelden van PROs

Niveaus	Voorbeelden van PROs
Symptomen	Pijn Vermoeidheid Slaapproblemen Ziekte-specifieke symptomen Cognitie Depressie Angst
Functionele status	Uitvoeren van dagelijkse activiteiten Problemen oplossen Uitvoeren van sociale activiteiten (sport, hobby's) Uitvoeren van sociale rollen (werk, gezin, familie) Seksueel functioneren
Ervaren gezondheid	Ervaren gezondheid
Kwaliteit van leven	Ervaren kwaliteit van leven

#### 4.2.1 Overwegingen bij de keuzes voor modellen

Kwaliteit van leven bevat dimensies van gezondheid op verschillende niveaus, gerelateerd aan het functioneren van mensen op fysiek, mentaal en sociaal gebied en de subjectieve evaluatie daarvan. Alle modellen laten zien dat PROs gemeten kunnen worden op verschillende niveaus. Ook laten de meeste modellen zien dat er een tijdvolgorde zit in de niveaus. Bijvoorbeeld, symptomen kunnen leiden tot beperkingen in het uitvoeren van dagelijkse activiteiten. Indelingen in niveaus verschillen tussen de modellen. Het model van Wilson & Cleary maakt onderscheid tussen het niveau van symptomen en functioneren. Symptomen zijn klachten die patiënten rapporteren als gevolg van een ziekte of behandeling, zoals pijn, misselijkheid, vermoeidheid, angst. Deze symptomen kunnen invloed hebben op het functioneren, dat wil zeggen het kunnen uitvoeren van dagelijkse activiteiten. In het ICF schema wordt onderscheid gemaakt tussen (stoornissen in) functies/anatomische eigenschappen, (beperkingen in het uitvoeren van) dagelijkse activiteiten en participatie(problemen) (kunnen werken, kunnen sporten etc). Functioneren is in de ICF een paraplueterm voor functies / anatomische eigenschappen, activiteiten en participatie; anders dan het model van Wilson & Cleary waar functioneren beperkt is tot het uitvoeren van activiteiten.

De modellen van Wilson & Cleary en Spilker laten zien dat beperkingen in functioneren/activiteiten van invloed zijn op de algemene ervaren gezondheid en kwaliteit van leven. Ten slotte wordt in alle kwaliteit van leven modellen en definities onderscheid gemaakt tussen fysieke, mentale en sociale aspecten (het duidelijkste te zien in het model van PROMIS).

Het model van de OMERACT groep wordt gebruikt bij het selecteren van uitkomsten voor een Core Outcome Set. De aanbeveling hierbij is om uit elke Core Area ten minste één uitkomst (domain) te kiezen. PROs vallen in dit model onder de Core Area 'Life Impact'. Deze Core Area wordt echter verder niet nader uitgewerkt in niveaus, en daarom lijkt het model niet zo behulpzaam bij het selecteren van PROs.

Er is geen consensus welke conceptueel model het beste gebruikt kan worden en geen van de modellen lijkt ideaal. Voor PROs lijkt het model van Valderas en Alonso het meest geschikt omdat deze het meest uitgewerkt is. Echter, er is geen duidelijk onderscheid tussen fysieke, mentale en sociale aspecten van gezondheid en er wordt geen duidelijk onderscheid gemaakt tussen fysiek functioneren en participatie. Het model van OMERACT zal naar verwachting veel gebruikt gaan worden bij de ontwikkeling van COS, maar is op het gebied van PROs niet uitgewerkt. Het model van PROMIS is handig omdat het direct gerelateerd is aan de beschikbare meetinstrumenten, maar maakt slechts een globaal onderscheid in fysieke, mentale en sociale aspecten van gezondheid en is hiermee te beperkt. Het voordeel van de ICF is het bestaan van ICF Core Sets waarin de belangrijkste items voor een bepaalde patiëntengroep zijn gedefinieerd. Via de ICF linking rules kunnen deze ICF items gekoppeld kunnen worden aan PROs en PROM items om de inhoudsvaliditeit van PROMs te bepalen. Nadeel is dat de ICF nog niet bij alle beroepsgroepen in de zorg is ingeburgerd en dat het onderscheid tussen activiteiten en participatie erg lastig te maken is. Er is dus behoefte aan een meer compleet conceptueel model voor PROs.

Bij de keuze van PROs is het belangrijk om na te denken op welk niveau men wil meten, functies/symptomen, activiteiten/functioneren, participatie, ervaren gezondheid, en/of kwaliteit van leven. Een aantal dilemma's moeten hierbij in overweging genomen worden:

Indien PROs worden gebruikt om een behandeling te evalueren, dienen ze aan te sluiten bij het beoogde doel van de behandeling. Het model van Wilson & Cleary laat zien dat hoe 'hoger' het niveau (d.w.z. meer naar rechts in het model), hoe groter de invloed van persoonlijkheidsfactoren en omgevingsfactoren (de invloed op verschillende niveaus telt op). Indien de behandeling plaatsvindt op het niveau van 'body functions and structures' of biologische en fysiologische variabelen (wat geen PROs zijn) dan is de directe relatie met ervaren gezondheid of kwaliteit van leven vaak niet zo groot. Een bloeddruk verlagend medicijn bijvoorbeeld kan bij de ene patiënt een minimale impact op de ervaren kwaliteit van leven hebben, maar bij de andere patiënt juist een grote impact.

Het effect van een behandeling wordt daarom in de kliniek vaak gemeten op uitkomsten links in het model (biologische en fysiologische variabelen en symptomen/functies) omdat deze uitkomsten het meest aansluiten bij het beoogde doel van de behandeling vanuit klinisch perspectief. Echter, patiënten vinden de PROs rechts in het model meestal het meest relevant (bijvoorbeeld

activiteiten of participatie). Zo kan het bijvoorbeeld zijn dat het voor een orthopeed of fysiotherapeut belangrijk is dat een patiënt na een nieuwe heup operatie geen pijn meer heeft en 100 meter kan lopen, maar de einddoelstelling voor de patiënt is dat hij zijn werk weer kan doen. Daarom is het vaak aan te bevelen om PROs op meerdere niveaus te meten.

Een ander dilemma is dat symptomen vaak ziekte-specifiek zijn en dus met ziekte-specifieke vragenlijsten moeten worden gemeten, terwijl beperkingen in het uitvoeren van dagelijkse activiteiten vaak veel minder specifiek zijn en dus met generieke vragenlijsten gemeten kunnen worden. Dit heeft gevolgen voor de vergelijkbaarheid van PRO scores.

#### 4.3 Stap 3: Selecteer relevante PROs per niveau

De keuze voor niveaus waarop men PROs wil meten is een eerste stap in de selectie van relevante PROs. Op sommige niveaus kunnen bestaande classificaties gebruikt worden om de verschillende PROs per niveau nog verder in te delen. Valderas en Alonso gebruiken bijvoorbeeld voor de categorie ‘symptoms’ de ICD-10 om symptomen verder te klasseren naar aandoening. Om de categorie ‘functional status’ verder te klasseren, gebruiken ze de indeling van de ICF, waarin bijvoorbeeld onderscheid gemaakt wordt tussen mobiliteit, huishoudelijke taken, zelfverzorging, etc. (zie Tabel 3 hieronder) [25].

Tabel 3: Categorieën voor het construct “ functionele status”

International classification of functioning, disability, and health chapter	
	Code
(a) Learning and applying knowledge	d110–d199
(b) General tasks and demands	d210–d299
(c) Communication	d310–d399
(d) Mobility	d410–d499
(e) Self-care	d510–d599
(f) Domestic life	d610–d699
(g) Interpersonal interactions and relationships	d710–d799
(h) Major life areas	d810–d899
(i) Community, social and civic life	d910–d999

Voor het selecteren van relevante PROs wordt aanbevolen om eerst een zo uitgebreid mogelijke lijst op te stellen van mogelijk relevante PROs. Deze potentiële lijst van PROs wordt dan vervolgens op basis van consensus met stakeholders ingekort tot een hanteerbare lijst.

Voor het identificeren van mogelijk relevante PROs kunnen verschillende methodieken gebruikt worden. Te onderscheiden methoden zijn:

- Gebruik maken van ziekte-specifieke conceptuele modellen
- Gebruik maken van ICF coresets

- Gebruik maken van een bestaande systematische reviews van meetinstrumenten of er zelf een uitvoeren
- Kwalitatief onderzoek onder patiënten en behandelaars (focus groepen, interviews, concept mapping)
- Schriftelijke Delphi studie
- Stakeholder bijeenkomst

Hieronder geven we een korte beschrijving van elk van de te onderscheiden methoden. Onderliggende modellen en classificaties die de selectie van PROs ondersteunen zijn in Bijlage 2 weergegeven.

Ad 1. Voor bepaalde aandoeningen bestaan ziekte-specifieke conceptuele modellen, die relevante uitkomsten voor een bepaalde doelgroep weergeven. Een voorbeeld is het model voor borstkanker van Dow & Ferrell, zie Figuur 9 (Bijlage 2) [27]. Dit model onderscheidt de niveaus fysiek, psychologisch, sociaal en spiritueel welzijn, en daarbinnen worden borstkanker specifieke PROs geklasseerd. Anders dan bij generieke conceptuele modellen, zijn ziekte-specifieke modellen vaak ontwikkeld op basis van input van patiënten en bevatten daarom vrijwel altijd PROs die relevant zijn voor patiënten. Wanneer voor een bepaalde doelgroep geen ziekte-specifiek model bestaat, kan het nuttig zijn om een dergelijk conceptueel model te ontwikkelen. Hiervoor gebruikt men tegenwoordig vaak de ‘concept mapping’ methode. Dit is een methode waarbij op systematische wijze een conceptueel model wordt ontwikkeld op basis van focus groepen met patiënten en behandelaars. Een voorbeeld is het conceptueel model ontwikkeld door Buchbinder et al. voor lage rugpijn, zie Figuur 10 in (Bijlage 2) [28].

Ad 2. Er bestaat een groot aantal ICF Core Sets. Dit zijn selecties uit de ICF voor een bepaalde aandoening. Zo zijn er ICF Core Sets voor RA, voor lage rugpijn, voor MS, voor autisme, voor borstkanker. Een Core Set biedt een overzicht van de ICF-klassen die relevant zijn voor mensen met die aandoening. Er zijn strikte regels voor het opstellen van een Core Set met verschillende typen onderzoek, waaronder focus groepen met patiënten/cliënten. De Core Sets bestaan uit een uitgebreide (comprehensive) set van selecties en een beperkte (brief) set van selecties. In Figuur 11 (Bijlage 2) is een voorbeeld weergegeven van de ICF coreset voor coronaire hartziekten (bron: WHO; <http://www.icf-research-branch.org/icf-core-sets-projects-sp-1641024398>).

Naast de ICF core sets zijn er ICF linking rules ontwikkeld die een verbinding (link) leggen tussen de gezondheidstatus zoals in PROMs gemeten en de ICF. Het doel hiervan is om meetinstrumenten (PROMs) te kunnen kiezen voor het evalueren van interventies [29]. Figuur 12 (Bijlage 2) geeft een voorbeeld van linking van de SF-12 aan de ICF.

Ad 3. Systematische reviews van meetinstrumenten geven een overzicht van alle bestaande meetinstrumenten om een bepaald concept te meten en hun

meeteigenschappen. Voor bepaalde aandoeningen bestaan systematische reviews van alle PROMs. Uit deze PROMs kan een overzicht gemaakt worden van de gemeten PROs. Voor andere aandoeningen bestaan alleen systematische reviews van meetinstrumenten die één bepaald concept meten, bijvoorbeeld vermoeidheid. Deze reviews zijn wel bruikbaar om geschikte PROMs te vinden maar zijn minder nuttig voor het identificeren van relevante PROs. De COSMIN website ([www.cosmin.nl](http://www.cosmin.nl)) geeft een overzicht van alle systematische reviews van meetinstrumenten die aspecten van gezondheid (inclusief PROs) meten. Dit overzicht wordt jaarlijks ge-update. Op dit moment wordt gewerkt aan een doorzoekbare online database waarin deze reviews zullen worden opgenomen. In deze database kan gezocht worden op concept, op doelpopulatie, maar ook op PROs in het algemeen.

Ad 4. Kwalitatief onderzoek onder patiënten en behandelaars is onmisbaar om relevante PROs te identificeren. Dit kwalitatief onderzoek kan bestaan uit individuele interviews of focus groepen (bijvoorbeeld concept mapping). Als dergelijke kwalitatief onderzoek niet al gepubliceerd is, verdient het aanbeveling zelf een dergelijk onderzoek op te zetten. Er zijn diverse richtlijnen voor de opzet en uitvoering van dergelijk kwalitatief onderzoek (zie bijvoorbeeld [30–34]).

Ad 5. Bij de ontwikkeling van Core Outcome Sets wordt vaak een schriftelijke Delphi studie gebruikt om relevante uitkomsten te identificeren en te selecteren. Het voordeel van deze methode is dat het uitgevoerd kan worden onder een grote groep patiënten, behandelaars, onderzoekers en/of andere stakeholders. Respondenten kunnen anoniem hun mening geven en staan hierbij niet onder invloed van het groepsproces, wat bij andere consensusmethoden wel het geval is. Een COS wordt vaak in een internationale setting ontwikkeld, daarom wordt deze methode vaak gebruikt. Bij een schriftelijke Delphi studie wordt eerst een lijst met potentieel relevante PROs vastgesteld op basis van literatuur. Vervolgens wordt deze lijst voorgelegd aan een panel van patiënten, behandelaars, onderzoekers en/of andere stakeholders. Er worden verschillende methoden gebruikt om de mening van panelleden te vragen en om consensus te bereiken in dergelijke studies. Panelleden worden bijvoorbeeld gevraagd om potentiële PROs te rangordenen op belangrijkheid. Een andere methode is om panelleden 100 punten te laten verdelen over de voor hen belangrijkste PROs (de belangrijkste PRO krijgt de meeste punten, de rater kiest zelf hoeveel punten hij geeft aan een zelf gekozen aantal PROs). Deze methode biedt meer mogelijkheden om het relatieve belang van PROs aan te geven. Bijna altijd wordt ook de gelegenheid gegeven om nieuwe PROs toe te voegen. In een tweede ronde wordt de gemiddelde rangorde van het panel teruggelinkt en worden panelleden opnieuw gevraagd PROs te rangordenen of een maximaal aantal PROs te selecteren (zie bijvoorbeeld [35]).

Ad 6. Het is vrijwel altijd nuttig om een stakeholder bijeenkomst te organiseren, bijvoorbeeld na een literatuurstudie, na kwalitatief onderzoek of na een

schriftelijke Delphi studie. Deze bijeenkomst kan gebruikt worden om een definitieve keuze te maken voor de selectie van PROs (en eventueel PROMs), maar is vooral belangrijk om draagvlak voor de gekozen PROs (en PROMs) te verkrijgen. Draagvlak is nodig om de implementatie van PROMs te faciliteren.

Er is geen consensus over de beste methodologie voor het selecteren van PROs. Er worden verschillende strategieën gebruikt. Ontwikkelaars van PROMs starten vaak met literatuuronderzoek om relevante PROs te identificeren op basis van bestaande vragenlijsten en ziekte-specifieke conceptuele modellen. Anderen starten juist met kwalitatief onderzoek met patiënten, bijvoorbeeld met focus groepen met patiënten om relevante PROs te identificeren. Bij de ontwikkeling van Core Outcome Sets worden vaak schriftelijke Delphi studies gebruikt omdat dit internationaal beter toepasbaar is, vaak aangevuld met consensus bijeenkomsten. Vrijwel altijd wordt een stakeholder bijeenkomst georganiseerd om draagvlak voor de gekozen PROs (en PROMs) te verkrijgen. De meest optimale methode is waarschijnlijk een combinatie van literatuuronderzoek, kwalitatief onderzoek met patiënten en behandelaars en een stakeholder bijeenkomst.

*Voorbeeld:*

In een lopend project naar de selectie van PROs en PROMs bij coronaire hartziekten worden onderdelen van bovenstaande stappen gecombineerd. Naast een literatuur review wordt een aangepaste Delphi procedure uitgevoerd om tot consensus te komen. Voor een prioritering van domeinen en meetinstrumenten wordt de IQ consensustool gebruikt [36]. IQ healthcare heeft deze tool op basis van veelvuldig onderzoek ontworpen.

De consensustool bestaat uit twee stappen:

1. online enquête voor het inventariseren van de relevantie van domeinen en/of meetinstrumenten;
2. focusgroep bijeenkomst om tot consensusbijeenkomst te komen.

In het project wordt de IQ consensustool zowel in een groep patiënten als experts uitgevoerd. De toepassing beslaat dan in totaal vier stappen. Tot slot vindt er een vijfde stap plaats waarin de begeleidingsgroep van het project, met daarin experts/behandelaars, patiënt vertegenwoordigers en verzekeraars, de definitieve selectie voor domeinen en meetinstrumenten maakt op basis van het advies van de expertgroep en patiëntengroep.

#### 4.3.1 Criteria voor het selecteren van PROs

Het is belangrijk om duidelijke criteria op te stellen voor het selecteren van PROs. Hierbij speelt een aantal dilemma's:

Het eerste dilemma betreft de vraag wiens mening (het meeste) telt. Sommige onderzoekers zijn van mening dat de selectie van relevante PROs uitsluitend (of voornamelijk) gebaseerd moet zijn op de mening van patiënten omdat het om

patiënt–gerapporteerde uitkomsten van zorg gaat. Anderen zijn van mening dat behandelaars en patiënten een even belangrijke stem hebben. Maar er zijn ook mensen die vinden dat de mening van behandelaars doorslaggevend moet zijn omdat de PROs moeten aansluiten op de verwachte effecten van de behandeling (bijvoorbeeld op het werkingsmechanisme en verwachte bijwerkingen van geneesmiddelen). Patiënten kunnen bepaalde uitkomsten heel belangrijk vinden, maar als de behandelaar hier geen invloed op heeft, is het de vraag of deze uitkomsten meegenomen moeten worden om de effectiviteit van een behandeling te evalueren.

Een tweede dilemma betreft de vraag hoeveel PROs men wil selecteren en hoe hierin een prioritering kan worden aangebracht. Het aantal te selecteren PROs hangt af van het doel van de PRO metingen, de lengte van de vragenlijsten en de praktische mogelijkheden, bijvoorbeeld de duur van het consult. De lengte van de vragenlijst moet worden afgewogen tegen de betrouwbaarheid.

#### 4.4 Stap 4: Stel de eisen vast aan de PROMs

Bij het selecteren van PROMs moet een aantal zaken in overweging worden genomen. Het is aan te bevelen om hierover na te denken voordat men start met het zoeken naar PROMs. Deze overwegingen betreffen de volgende:

Type meetinstrument: bijvoorbeeld generiek of ziekte–specifiek, of individueel. Voor gebruik in de dagelijkse praktijk zijn ziekte–specifieke instrumenten op individueel niveau (zoals SeiQol [17]) vaak gewenst, terwijl voor het meten van verschillen in kwaliteit van zorg tussen instellingen generieke instrumenten handiger kunnen zijn. In onderzoek worden vaak beide soorten PROMs gebruikt. Voor het meten van symptomen kan het best een ziekte–specifiek instrument gebruikt worden. Voor het meten op hogere niveaus (activiteiten, participatie, ervaren gezondheid) kunnen goede kwaliteit generieke instrumenten worden gebruikt, mits ze voldoende betrouwbaar zijn. PROMIS biedt een hele serie meetinstrumenten die hiervoor gebruikt kunnen worden ([www.dutchflemishpromis.nl](http://www.dutchflemishpromis.nl)). Er moet echter nog onderzocht worden of PROMIS instrumenten in staat zijn om te discrimineren tussen zorgaanbieders en instellingen.

Manier van vragenlijst afnemen: bijvoorbeeld op papier, in een interview, web–based, via een Computer Adaptieve Test (CAT), of met behulp van een proxy. Een interview kan geschikter zijn dan een papieren vragenlijst voor een bepaalde doelpopulatie, bijvoorbeeld slechtzienden. CAT is een methode waarbij de computer vragen selecteert uit een item bank, op basis van het antwoord op eerdere vragen. Deze methode is zeer efficiënt omdat patiënten gemiddeld maar 3 tot 7 vragen hoeven in te vullen om een betrouwbare score te krijgen. Deze methode kan echter alleen gebruikt worden bij item banken die op basis van Item Response Theorie methoden ontwikkeld zijn, zoals bijvoorbeeld de PROMIS item banken. Daarnaast ontbreekt nog de infrastructuur om CATs af te nemen.



Een proxy meting is een meting waarbij de vragenlijst door iemand anders, namens of over de patiënt, wordt ingevuld, meestal omdat de patiënt zelf niet in staat is om vragenlijsten in te vullen, bijvoorbeeld kinderen, ernstig zieke patiënten, of patiënten met cognitieve stoornissen of beperkingen.

Criteria voor klinimetrische eigenschappen en hanteerbaarheid: Het is aan te raden om van te voren na te denken welke klinimetrische eigenschappen het meest belangrijk zijn voor de beoogde toepassing. Voor het stellen van een diagnose is bijvoorbeeld de betrouwbaarheid en diagnostische waarde van een instrument heel belangrijk, terwijl voor het meten van de effecten van behandeling over de tijd de responsiviteit van een meetinstrument heel belangrijk is. Voor het gebruik van PROMs in de dagelijkse zorg is de absolute meetfout heel belangrijk. Van veel PROMs is de meetfout echter niet bekend. Ook is het aan te raden om van te voren na te denken over aspecten van hanteerbaarheid die een rol zouden kunnen spelen bij de keuze van een PROM, zoals de lengte van de vragenlijst en beschikbaarheid van een Nederlandse versie (zie ook stap 8).

#### 4.5 Stap 5: Identificeer alle beschikbare PROMs

Het verdient aanbeveling om per PRO alle beschikbare PROMs te identificeren om hieruit vervolgens de meeste geschikte PROM per PRO te kunnen selecteren. Sommige onderzoekers zijn van mening dat het genoeg is om één of meerdere PROMs te identificeren die aan de gestelde criteria voldoen (dit zijn vaak PROMs die veel gebruikt worden). Echter, in een recente Delphi studie werd consensus bereikt voor de aanbeveling om alle beschikbare PROMs te identificeren alvorens een keuze te maken [1]. Een belangrijke reden hiervoor is dat veelgebruikte PROMs niet per se de beste PROMs zijn. Veel van de populaire PROMs zijn ontwikkeld in de jaren '90 van de vorige eeuw en zijn soms verouderd (bevatten bijvoorbeeld geen vragen over problemen met computer gebruik). Ook zijn ze vaak niet ontwikkeld volgens de huidige geldende standaarden voor PROM ontwikkeling (met name de bijdrage van patiënten bij de ontwikkeling van de PROM is vaak beperkt geweest), waardoor de relevantie en formulering van vragen soms te wensen over laten. Recent ontwikkelde PROMs kunnen veelbelovend zijn, maar zijn misschien nog niet voldoende gevalideerd. Een afweging op basis van alle bestaande PROMs is dus van belang.

Er zijn diverse bronnen die gebruikt kunnen worden om beschikbare PROMs te identificeren:

- Bestaande systematische reviews van meetinstrumenten ([www.cosmin.nl](http://www.cosmin.nl)). Bij het gebruik van dit soort reviews moet men wel beseffen dat de bestaande systematische reviews niet altijd van goede kwaliteit zijn. Er zijn sterke aanwijzingen dat veel systematische reviews van meetinstrumenten niet compleet zijn [37].

- Ook zijn reviews gauw verouderd en ontbreekt vaak informatie over recent ontwikkelde/gevalideerde PROMs, die juist heel interessant zijn. Aanvullende zoekstrategieën zijn daarom nodig.
- Systematische literatuur search. Als er geen systematische review van de geselecteerde PROs bestaat of als deze verouderd of van minder goede kwaliteit is, wordt aangeraden een systematische literatuur search te doen. Een onderzoeksgroep uit Oxford heeft voor dit doel een speciaal PROM zoekfilter voor Pubmed ontwikkeld (<http://phi.uhce.ox.ac.uk/about.php>). Dit filter kan gecombineerd worden met een zoekfilter voor het vinden van validatiestudies [38] wanneer men alleen PROMs wil vinden waarvan ten minste iets bekend is over de klinimetrische eigenschappen.
- Databases van meetinstrumenten. Op internet zijn vele databases van meetinstrumenten te vinden. Voor PROMs zijn met name de database Meetinstrumenten in de zorg ([www.meetinstrumentenzorg.nl](http://www.meetinstrumentenzorg.nl)) en de PROQOLID database ([www.proqolid.org](http://www.proqolid.org)) interessant. Een overzicht van relevante websites is te vinden op de website van het Kenniscentrum Meetinstrumenten VUmc ([www.kmin-vumc.nl](http://www.kmin-vumc.nl)).
- Veelgebruikte PROMs. Het kan wenselijk zijn om aan te sluiten bij PROMs die veel gebruikt worden (of die in het buitenland gebruikt worden), maar er kunnen ook goede redenen zijn om dat niet te doen. Het is in ieder geval handig om in kaart te brengen welke PROMs veel gebruikt worden en wat de positieve en negatieve ervaringen met het gebruik van die PROMs zijn.
- Aanschrijven experts. Een goede (aanvullende) manier om PROMs te vinden is het aanschrijven van experts in een bepaald vakgebied. Hiermee kan bijvoorbeeld informatie verkregen worden over PROMs die nog in ontwikkeling zijn.

#### 4.6 Stap 6: Maak een voorselectie op basis van inhoud

De volgende stap is om de inhoud van de geïdentificeerde PROMs te vergelijken met de geselecteerde PROs en het gebruikte conceptueel model. Belangrijk om te realiseren is dat de naam van een PROM vaak geen goede beschrijving is van wat het instrument daadwerkelijk meet. Hier zijn talloze voorbeelden van. Zo meten veel 'kwaliteit van leven' instrumenten (bijvoorbeeld SF-36, EQ-5D) feitelijk geen kwaliteit van leven, maar aspecten van functioneren en ervaren gezondheid. Er zijn veel vragenlijsten die 'disability' beogen te meten, waarbij verschillende definities van disability worden gebruikt. In de ICF is disability de negatieve paraplu-term (stoornissen, beperkingen, participatieproblemen) en in deze de negatieve evenknie van functioneren. Veel vragenlijsten die 'participatie' beogen te meten, meten feitelijk voornamelijk aspecten van fysieke functioneren [39-41]. En vragenlijsten die 'pijn' beogen te meten, operationaliseren dit op heel verschillende manieren (bijvoorbeeld gemiddelde pijn intensiteit in de afgelopen week, pijn tijdens het uitvoeren van activiteiten). Zoals in Paragraaf 4.2.1 is toegelicht, worden PROMs vaak op item niveau gelinkt aan de ICF om inzicht te krijgen in wat PROMs precies meten [29]. Al in 2008 waren er 57 artikelen verschenen die gaan over het linken van PROMs aan de ICF [42]. Als een dergelijk

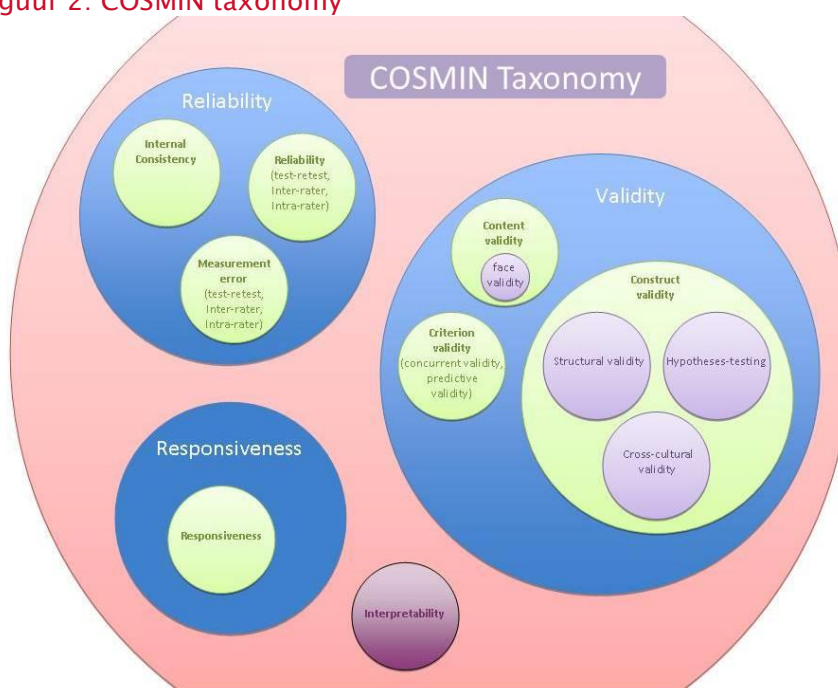
artikel niet bestaat, is het aan te raden om zelf een kritische beschouwing te maken van de inhoud van de PROMs.

Op basis van deze inhoudelijke beschouwing kan een eerste selectie van PROMs worden gemaakt voor een nadere beschouwing in de volgende stap.

#### 4.7 Stap 7: Bepaal de kwaliteit van geselecteerde PROMs

De volgende stap is het in kaart brengen van de kwaliteit van de geselecteerde PROMs. Daarbij gaat het om de klinimetrische eigenschappen van de PROMs. De relevante klinimetrische eigenschappen voor PROMs zijn weergegeven in de COSMIN taxonomie, die ontwikkeld is in een internationale Delphi studie [43] (Figuur 2).

Figuur 2: COSMIN taxonomy



Aanbevolen wordt om de klinimetrische eigenschappen in de volgende volgorde te bekijken:

1. Face validity. Dit is de allerbelangrijkste eigenschap van een PROM. Het is belangrijk om alle vragen van de geselecteerde PROMs kritisch te lezen en te beoordelen. Criteria voor face validity zijn weergegeven in Tabel 4.
2. Content validity (inhoudsvaliditeit). Bij content validity gaat het om de vraag of alle vragen in een PROM relevant zijn voor de PRO die men beoogt de meten, voor de doelpopulatie en voor het doel van de meting. Daarnaast is het belangrijk dat alle relevante aspecten van een PRO gedekt worden door de vragenlijst. Content validity kan het beste onderzocht worden aan de hand van informatie over de ontwikkeling van de PROM. Om een goede content validity te krijgen, is kwalitatief onderzoek met patiënten (representatief voor de doelpopulatie) onontbeerlijk bij de ontwikkeling van een PROM. Het

kwalitatief onderzoek heeft tot doel om relevante (aspecten van) PROs te identificeren, maar ook om na te gaan of patiënten de vragen goed begrijpen en op de juiste manier interpreteren.

Tabel 4. Criteria voor face validity [47] - in ontwikkeling

1	Is/are the construct(s) of interest clearly described?
2	Do all items seem to measure the construct of interest?
3	Is the formulation of questions appropriate, considering the construct and target population of interest?
4	Are the response options appropriate, considering the construct and target population of interest?
5	Are the response options consistent with the stem of the questions?
6	Is the recall period appropriate, considering the construct and target population of interest?
7	Are cultural issues avoided?
8	Are the items relevant for the purpose of the instrument (e.g. for evaluation, the items should potentially be changeable)
9	Is the mode of administration appropriate, considering the construct and target population of interest?
10	Does the scoring algorithm make sense?

Als dergelijk onderzoek niet of onvoldoende goed gedaan is, verdient het aanbeveling om de inhoud van de PROM met een selectie van patiënten en behandelaars te bespreken om zelf de content validity te kunnen beoordelen. Criteria voor content validity zijn weergegeven in Tabel 5.

3. Structurele validiteit en interne consistentie. Structurele validiteit is de mate waarin de dimensionaliteit van een PROM (de verdeling van items over subschalen) overeenkomt met wat je zou verwachten op basis het te meten construct en conceptuele model. Bijvoorbeeld, als een instrument pijn, fysiek functioneren en participatie beoogt te meten, dan verwacht je 3 subschalen met vragen over respectievelijk pijn, fysiek functioneren en participatie (en liever niet 1 schaal waarin alle vragen door elkaar in zitten, want dan weet je niet goed wat een score of verandering in score precies betekent). Vragen over hetzelfde construct horen dus bij elkaar in 1 subschaal en vragen over verschillende constructen horen in verschillende subschalen. De wordt statistisch onderzocht met behulp van factor analyse. Vervolgens wordt per subschaal de Cronbach's alpha berekend, als maat voor interne consistentie (de gemiddelde correlatie tussen de vragen in een subschaal). De Cronbach's alpha moet groter zijn dan 0.70 per subschaal [44].

Tabel 5. Criteria voor content validity [47] – in ontwikkeling

1	Is there theoretical support for the chosen domains (conceptual model)?
2	Are the included domains relevant for the construct of interest?
3	Are the included items relevant for the construct of interest?
4	Are the included domains relevant for the target population?
5	Are the included items relevant for the target population?
6	Are the included domains relevant for the purpose of the measurement?
7	Are the included items relevant for the purpose of the measurement?
8	Are no important domains missing?
9	Are no important items missing?
10	Are the items and instructions correctly understood by the target population?

4. Test–hertest betrouwbaarheid en meetfout. De test–hertest betrouwbaarheid geeft aan hoe goed je patiënten van elkaar kunt onderscheiden ondanks meetfouten. Als de variatie tussen patiënten (of veranderingen binnen patiënten) groot is, kun je verschillen tussen patiënten (of veranderingen binnen patiënten) wel onderscheiden van meetfouten, maar als de variatie (of de verandering) klein is wordt dit moeilijker. De betrouwbaarheid wordt uitgedrukt in de Intraclass Correlatie Coefficient (ICC). Deze moet groter zijn dan 0.70 voor het gebruik van PROMs op groepsniveau. Voor het gebruik van PROMs op individueel niveau (in de dagelijkse klinische praktijk) wordt gezegd dat de ICC ten minste 0.90 moet zijn [45]. Echter, voor het gebruik van PROMs op individueel niveau (in de dagelijkse klinische praktijk), is de absolute meetfout veel interessanter. Deze geeft namelijk de meetfout van een score weer, uitgedrukt in de eenheid van de schaal. De Smallest Detectable Change (SDC) geeft de kleinste verandering in score aan die je kunt onderscheiden van meetfouten in een individuele patiënt. Als de SDC bijvoorbeeld 25 punten is op een vragenlijsten die gescoord wordt van 0 tot 100, dan betekent dit dat een patiënt minimaal 25 punten veranderd moet zijn (beter of slechter) om met 95% zekerheid te kunnen zeggen dat de patiënt echt veranderd is. Bij een kleinere verandering zou dit dus best meetfout kunnen zijn. Veel PROMs hebben een grote meetfout (omdat ze eigenlijk voor onderzoek zijn ontwikkeld). Bij PROMs met een schaal van 0 tot 100 is een SDC van 25 tot 30 punten niet ongebruikelijk. Men moet er dan dus rekening mee houden dat alleen vrij grote veranderingen met grote zekerheid van meetfouten kunnen worden onderscheiden. Het is dus altijd verstandig om met de patiënt in gesprek te gaan over veranderingen in scores, en deze niet al te absoluut te nemen. Onderzoek heeft aangetoond dat meetinstrumenten die op basis van IRT methoden ontwikkeld zijn, zoals bijvoorbeeld de PROMIS meetinstrumenten, een kleinere meetfout hebben. Deze instrumenten zijn dus beter geschikt om in de klinische praktijk te gebruiken. Wanneer PROM scores op geaggregeerd niveau worden gebruikt

(bijvoorbeeld voor kwaliteit van zorg metingen) heeft men niet zo'n last van die meetfout omdat meetfouten op groepsniveau uitmiddelen. Een ander probleem is dat van veel PROMs de meetfout niet bekend is omdat de meeste betrouwbaarheidsstudies alleen ICCs rapporteren. Dit is jammer omdat met dezelfde gegevens ook de SDC berekend kan worden, maar meestal staat er niet voldoende informatie in de artikelen om dit achteraf nog zelf te kunnen doen.

5. Construct validiteit. Construct validiteit is de mate waarin correlaties tussen de PROM en andere meetinstrumenten die vergelijkbare of juiste andere constructen meten, of verschillen in scores tussen (sub)groepen, overeenkomen met wat je zou verwachten op basis van het construct en het conceptueel model. Bijvoorbeeld, als een instrument pijn, fysieke activiteiten en participatie beoogt te meten, dan zou je verwachten dat de subschaal pijn hoog correleert met bijvoorbeeld een VAS pijn. Je zou ook verwachten dat de subschaal fysiek functioneren hoog correleert met een performance-based test waarin patiënten gevraagd wordt dezelfde activiteiten uit te voeren. En je zou verwachten dat de subschaal participatie hoog correleert met bijvoorbeeld een vragenlijst over rol functioneren of met een maat voor werkverzuim. Of je zou de hypothese kunnen toetsen dat patiënten die in de ziektewet zitten lager scores op de participatieschaal dan patiënten die nog wel werken. Ten slotte is het ook belangrijk om aan te tonen dat de schalen verschillende constructen meten, door bijvoorbeeld te laten zien dat de subschaal fysieke activiteiten hoger met de performance test correleert dan de subschaal participatie. Het probleem met construct validiteit is dat dit niet te bewijzen valt. Construct validiteit is niet in één getal uit te drukken, zoals betrouwbaarheid. Hoe meer hypothesen bevestigd worden, hoe meer vertrouwen men krijgt in de validiteit van de PROM, maar deze kan nooit met zekerheid worden aangetoond. Er is geen consensus over hoeveel hypothesen men zou moeten toetsen of met hoeveel andere instrumenten men een PROM zou moeten vergelijken. Vaak wordt de vuistregel gehanteerd die zegt dat ten minste 75% van de resultaten van een validatiestudie in overeenstemming moet zijn met de hypothesen [46].
6. Criterion validiteit (de overeenstemming met een gouden standaard) wordt hier niet besproken omdat er voor PROMs geen gouden standaarden bestaan. De enige situatie waarin je van een gouden standaard zou kunnen spreken is wanneer een korte versie van een PROM vergeleken wordt met een lange versie. De verwachting is dan dat de correlatie tussen de korte en lange versie heel hoog is ( $>0.70$ ).
7. Responsiviteit. Responsiviteit is een vorm van validiteit, maar dan in een longitudinale context. Bij validiteit gaat het om de validiteit van één meting, bij responsiviteit gaat het om de validiteit van een verschilscore van twee metingen (bijvoorbeeld voor en na een behandeling). Responsiviteit kan het beste op dezelfde manier worden onderzocht als validiteit: door hypothesen

te toetsen over verwachte correlaties. In dit geval gaat het dan om correlaties tussen de verschillcores. Je zou dus bijvoorbeeld verwachten dat veranderingen in de subschaal fysiek functioneren hoog correleren met veranderingen in de performance-based test. Hierbij kan dezelfde vuistregel voor het interpreteren van de resultaten worden gebruikt als bij validiteit [47]. Responsiviteit is alleen belangrijk voor PROMs die op meerdere momenten worden afgenomen waarbij verschillcores worden berekend. Van veel PROMs die in dit soort situaties worden gebruikt is echter de responsiviteit niet onderzocht.

8. Cross-culturele validiteit. Cross-culturele validiteit is van belang wanneer PROM scores uit verschillende landen met elkaar vergeleken worden, of als je doelgroep bestaat uit mensen met een verschillende culturele achtergrond. De scores uit de verschillende landen moeten dan op dezelfde manier geïnterpreteerd kunnen worden. Dit is niet altijd het geval omdat door vertaling en door culturele verschillen vragen in verschillende landen soms verschillend geïnterpreteerd worden. Zo wordt het horen van stemmen in je hoofd in sommige culturen opgevat als indicatie van psychose, terwijl dit in andere culturen juist een religieuze betekenis heeft. Een ander voorbeeld; omdat in Canada de toiletten hoger zijn dan in Nederland hebben Canadezen minder moeite met opstaan van het toilet. Dat wil echter niet zeggen dat hun fysiek functioneren beter is dan die van Nederlanders. Dergelijke aspecten moeten zijn onderzocht wanneer men scores uit verschillende landen direct met elkaar wil vergelijken. Statistische methoden die hiervoor gebruikt worden zijn multiple-group factor analyse en Differential Item Functioning. Van veel PROMs is de cross-culturele validiteit niet onderzocht. In systematische reviews wordt dan meestal alleen naar de kwaliteit van de vertaling gekeken. Een slechte vertaling kan een indicatie zijn voor een slechte content validity of meetfout (omdat de vraag niet goed wordt begrepen), of een slechte cross-culturele validiteit.
9. Interpreteerbaarheid. Interpreteerbaarheid is eigenlijk geen klinimetrische eigenschap van een PROM, maar is wel zodanig belangrijk dat het vaak in het rijtje wordt meegenomen. Om PROMs in de klinische praktijk te kunnen gebruiken moet duidelijk zijn wat een score van bijvoorbeeld 37 punten betekent, is dit hoog of laag ten opzichte van bijvoorbeeld vergelijkbare patiënten of ten opzichte van een referentiegroep (zoals de algemene bevolking). De Smallest Detectable Change (SDC) geeft bruikbare informatie om veranderingen in scores te kunnen interpreteren. Daarnaast is het belangrijk om te weten wat de kleinste verandering is in score die patiënten relevant vinden, de Minimal Important Change (MIC). De MIC wordt vaak bepaald door in een longitudinaal onderzoek patiënten te vragen om op een 5- of 7-puntsschaal aan te geven in hoeverre zij zijn veranderd ten opzichte van de uitgangssituatie (bijvoorbeeld gelijk gebleven, een beetje verbeterd, veel verbeterd). Wat patiënten hierop antwoorden wordt vergeleken met de verandering in scores op de PROM, om zo de MIC te bepalen. Een probleem is

dat van veel PROMs de MIC niet bepaald is. Een ander probleem is dat de MIC waardes die in onderzoek worden gevonden vaak erg onbetrouwbaar zijn. In verschillende studies worden verschillende waardes voor de MIC van dezelfde PROM gevonden. Dit heeft deels te maken met het feit dat de MIC vaak op een te kleine groep patiënten wordt berekend, maar er zijn ook andere methodologische problemen. Er is nog geen consensus over de beste methode om de MIC te bepalen.

Bij de beoordeling van de klinimetrische eigenschappen moet ook de kwaliteit van de validatiestudies in aanmerking genomen worden. Als de kwaliteit van de studie niet goed is, kunnen de gerapporteerde resultaten onbetrouwbaar of vertekend zijn. De methodologische kwaliteit van klinimetrische studies kan worden beoordeeld met de COSMIN checklist [47]. De COSMIN checklist bevat voor elke klinimetrische eigenschap standaarden waar een goede studie aan moet voldoen. In een systematische review van goede kwaliteit wordt de methodologische kwaliteit van de geïncludeerde studies meegewogen bij het beoordelen van de kwaliteit van de PROMs. Ook wordt het aantal beschikbare studies en de consistentie tussen de resultaten van verschillende studies meegewogen (zie bv [48]).

Om de kwaliteit van de geselecteerde PROMs op al deze klinimetrische eigenschappen te kunnen beoordelen, rekening houdend met de kwaliteit van de studies, is behoorlijk wat expertise nodig. Het wordt aanbevolen om eerst te zoeken naar bestaande goede kwaliteit systematische reviews van PROMs, waarin alle klinimetrische eigenschappen door de auteurs van de reviews zijn beoordeeld (daarbij de beperkingen van deze reviews in acht nemende). Daarnaast is het verstandig om een onderzoeker met expertise op het gebied van klinimetrie te betrekken bij de beoordeling van de kwaliteit van de PROMs. Meer achtergrond informatie over de klinimetrische eigenschappen is te vinden in het boek 'Measurement in Medicine' [13].

#### 4.8 Stap 8: Bepaal de hanteerbaarheid van geselecteerde PROMS

Ten slotte is het belangrijk om de hanteerbaarheid van de geselecteerde PROMs te beoordelen. Een PROM kan hele goede klinimetrische eigenschappen hebben, maar als het in de praktijk te lastig is om de PROM af te nemen, is het niet aan te bevelen deze PROM te gebruiken. Bij hanteerbaarheid moet onderscheid gemaakt worden tussen hanteerbaarheid voor de patiënt (zoals leesbaarheid, aard en belasting van vragen, invultijd) en hanteerbaarheid voor de hulpverlener (zoals benodigde expertise om de scores te kunnen berekenen en interpreteren). Het moet duidelijk zijn hoe de scores moeten worden berekend en hoe moet worden omgegaan met ontbrekende antwoorden zodat scores altijd op dezelfde manier berekend worden. Wanneer scores moeten worden teruggekoppeld naar patiënten en behandelaars is het belangrijk dat duidelijk is wat scores betekenen (zie onder interpreteerbaarheid). Daarnaast zijn er nog praktische aspecten om



aan te denken zoals kosten (copyright) en beschikbaarheid van PROMs in het Nederlands.

#### 4.9 Stap 9: Selecteer de beste PROMs

Om uiteindelijk per PRO de beste PROM te selecteren moeten alle bovenstaande criteria tegen elkaar worden afgewogen: Welke PROM past het beste bij het construct dat je beoogt te meten? Welke PROM heeft de beste klinimetrische eigenschappen? En welke PROM is het best hanteerbaar? Geen enkele PROM zal aan alle criteria voldoen, dus er moeten vrijwel altijd concessies gedaan worden. Daarnaast is misschien ook wel geen enkele PROM geschikt voor alle doelen. Als men een PROM voor meerdere doelen wil gebruiken, zullen dus ook altijd concessies gedaan moeten worden. Het is belangrijk om deze concessies expliciet te maken.

Om een PROM überhaupt te kunnen inzetten moet in ieder geval de face validity goed zijn en moet de PROM voldoende hanteerbaar zijn. Daarnaast moeten er geen aanwijzingen zijn dat een PROM op bepaalde klinimetrische eigenschappen slecht scoort. Als bijvoorbeeld aangetoond is dat een PROM een slechte responsiviteit heeft is het niet aan te bevelen deze PROM in te zetten om patiënten te monitoren over de tijd. Als de responsiviteit van een PROM echter niet bekend is kan deze in principe wel ingezet worden onder de voorwaarde dat de face validity goed is. Echter, het aan te raden om extra gegevens te verzamelen (bijvoorbeeld extra vragenlijsten af te nemen) om de responsiviteit alsnog te kunnen onderzoeken.

Voor de selectie kunnen ook subschalen van PROMs gebruikt worden. Het is dan wel van belang dat de validiteit op subschaal niveau is onderzocht en dat de score op de subschaal onafhankelijk gebruikt kan worden. Op die manier kan een 'nieuwe' PROM ontwikkeld worden die bestaat uit subschalen van verschillende bestaande PROMs. Die scores van die subschalen kunnen dan uiteraard niet bij elkaar opgeteld worden tot een totaalscore.

Indien PROMs van goede kwaliteit beschikbaar zijn, maar niet beschikbaar zijn in het Nederlands dan kan overwogen worden om de PROM te vertalen. Die vertaling dient zorgvuldig te gebeuren en het verdient aanbeveling om daarbij een cognitieve validatie te doen van de vertaalde vragenlijst bij de beoogde patiëntengroep [47]. Daarnaast is de vertaalde PROM feitelijk niet gebruikt bij specifieke doelgroep in Nederland en is verdere validatie wenselijk.



## 5. Implementatie van de tools: hoe nu verder?

Deze handreiking beschrijft een stappenplan voor de selectie van PROs en PROMs, inclusief per stap de beschrijving van de methodologische en praktische dilemma's die een rol spelen bij de selectie van PROs en PROMs voor verschillende doelen: zorg, onderzoek, en kwaliteit van zorgmetingen. Het proces van het selecteren van PROs en PROMs moet gebeuren door een multidisciplinair team van stakeholders, zoals zorgverleners, patiënten (vertegenwoordigers), methodologen, en zorgverzekeraars.

Niet alle stappen zijn even concreet uitgewerkt. Voor een verdere verdieping en uitwerking van deze handreiking is raadpleging van en samenwerking met de verschillende stakeholders gewenst. Naar sommige stappen is meer onderzoek nodig. Onze intentie is om deze handreiking uit te werken tot een Toolbox voor de selectie, doorontwikkeling en toepassing van PROs en PROMs in de Nederlandse Gezondheidszorg, vergelijkbaar met de HARING toolbox voor richtlijnontwikkeling.

Voor de ontwikkeling van uitkomstindicatoren is het van belang dat aanbieders patiënten en zorgverzekeraars een gezamenlijke visie ontwikkelen [20]. Onderling vertrouwen is nodig om kennis en expertise te ontwikkelen om uitkomstindicatoren te ontwikkelen op basis van patiënt-gerapporteerde uitkomsten zonder misbruik van data. Een van de aspecten van een gezamenlijke visie is dat de uitkomstindicatoren als signaal worden gebruikt voor nader onderzoek of gesprek, en niet om een absoluut oordeel over kwaliteit te vellen.

## Dankwoord

De auteurs danken dr. Yvonne Heerkens (lector Arbeid & Gezondheid Hogeschool Arnhem en Nijmegen en senior onderzoeker Nederlands Paramedisch Instituut) voor haar ondersteuning bij het ordenen van de terminologie om de domeinen van gezondheid op verschillende niveaus weer te kunnen geven.

Dit document is mede tot stand gekomen dankzij financiële ondersteuning van de Stichting Miletus.

## Referenties

- [1] Prinsen CA, Vohra S, Rose MR, King-Jones S, Ishaque S, Bhaloo Z, et al. Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set'. *Trials* 2014;15:247.
- [2] Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013.
- [3] Macefield RC, Jacobs M, Korfage IJ, Nicklin J, Whistance RN, Brookes ST, et al. Developing core outcomes sets: methods for identifying and including patient-reported outcomes (PROs). *Trials* 2014;15:49.
- [4] Boers M, Kirwan JR, Gossec L, Conaghan PG, D'Agostino MA, Bingham CO, III, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. *J Rheumatol* 2014;41:1025-30.
- [5] Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, D'Agostino MA, et al. Developing Core Outcome Measurement Sets for Clinical Trials: OMERACT Filter 2.0. *J Clin Epidemiol* 2014;67:745-53.
- [6] Selb M, Escorpizo R, Kostanjsek N, Stucki G, Ustun B, Cieza A. A guide on how to develop an international classification of functioning, disability and health core set. *Eur J Phys Rehabil Med* 2014.
- [7] van Kessel P, Triemstra M, de Boer D. Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures. Nivel & Zorginstituut, 2014.
- [8] Basch E, Torda P, Adams K. Standards for patient-reported outcome-based performance measures. *JAMA* 2013;310:139-40.
- [9] Cella D, Hahn EA, Jensen SE, Butt Z, Nowinsky CJ, Rothrock N. Methodological Issues In The Selection, Administration And Use Of Patient-Reported Outcomes In Performance Measurement In Health Care Settings. 2012. National Quality Forum.
- [10] Deutsch A, 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. 2012. National Quality Forum.
- [11] National Quality Forum. Patient Reported Outcomes (PROs) in Performance Measurement. 2013.
- [12] Nederlandse Orthopaedische Vereniging. Patient Reported Outcome Measures. Advies Nederlandse Orthopaedische Vereniging (NOV). 2012.
- [13] de Vet HCW, Terwee CB, Mokkink LB, Knol DJ. Measurement in medicine. A practical guide. Cambridge: Cambridge University Press; 2011.
- [14] Beurskens S, van Peppen R, Stutterheim E, Swinkels R, Wittink H. Meten in de praktijk. Stappenplan voor het gebruik van meetinstrumenten in de gezondheidszorg. Houten: Bohn Stafleu van Loghum; 2012.

- [15] Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. New York: Oxford University Press; 2003.
- [16] US Food & Drug Administration. Guidance for industry – patient reported outcome measures: use in medical product development to support labeling claims. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>. 2014.
- [17] Joyce CR, Hickey A, McGee HM, O'Boyle CA. A theory-based method for the evaluation of individual quality of life: the SEIQoL. *Qual Life Res* 2003;12:275–80.
- [18] Black N. Patient reported outcome measures could help transform healthcare. *BMJ* 2013;346:f167.
- [19] Wu AW, Kharrazi H, Boulware LE, Snyder CF. Measure once, cut twice--adding patient-reported outcome measures to the electronic health record for comparative effectiveness research. *J Clin Epidemiol* 2013;66:S12–S20.
- [20] van der wees PJ, Nijhuis–van der Sanden MWG, Ayanian JZ, Black N, Westert GP. Integrating the Use of Patient–Reported Outcomes for Both Clinical Practice and Performance Measurement: Views of Experts from Three Countries. *The Milbank Quarterly* in press. 2014.
- [21] Koolman X, Zuidgeest M, Visser J, Appelman M. Indicatorstandaard: methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren in de zorg. 2012. Diemen, Kwaliteitsinstituut i.o./CVZ.
- [22] WHO. Towards a common language for functioning, disability and Health – ICF. 2002. Geneva, World Health Organization.
- [23] Wilson IB, Cleary PD. Linking clinical variables with health–related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273:59–65.
- [24] Spilker B. *Quality of Life and Pharmacoeconomics in Clinical Trials*. Philadelphia: Lippincott–Raven; 1996.
- [25] Bakas T, McLennon SM, Carpenter JS, Buelow JM, Otte JL, Hanna KM, et al. Systematic review of health–related quality of life models. *Health Qual Life Outcomes* 2012;10:134.
- [26] Valderas JM, Alonso J, Guyatt GH. Measuring patient–reported outcomes: moving from clinical trials into clinical practice. *Med J Aust* 2008;189:93–4.
- [27] Dow KH, Ferrell BR, Leigh S, Ly J, Gulasekaram P. An evaluation of the quality of life among long–term survivors of breast cancer. *Breast Cancer Res Treat* 1996;39:261–73.
- [28] Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity–driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. *Arthritis Research & Therapy* 2011;13:R152.
- [29] Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. *J Rehabil Med* 2005;37:212–8.

- [30] Hak T, van der Veer K, Jansen H. The three-step Test-Interview (TSTI): An observation-based method for pretesting self-completion questionnaires. *Survey Research Methods* 2, 143-150. 2008.
- [31] Ploughman M, Austin M, Stefanelli M, Godwin M. Applying cognitive debriefing to pre-test patient-reported outcomes in older people with multiple sclerosis. *Qual Life Res* 2010;19:483-7.
- [32] Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res* 2003;12:229-38.
- [33] Kuper A, Reeves S, Levinson W. An introduction to reading and appraising qualitative research. *BMJ* 2008;337:a288.
- [34] Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;19:349-57.
- [35] Wylde V, MacKichan F, Bruce J, Gooberman-Hill R. Assessment of chronic post-surgical pain after knee replacement: Development of a core outcome set. *Eur J Pain* 2014.
- [36] Stienen JOM. Hoe komje tot de kern van de zaak? Methoden voor het bereiken van consensus. *Tijdschrift Sociale Geneeskunde* 89, 405-408. 2011.
- [37] Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res* 2009;18:313-33.
- [38] Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115-23.
- [39] Eysen IC, Steultjens MP, Dekker J, Terwee CB. A systematic review of instruments assessing participation: challenges in defining participation. *Arch Phys Med Rehabil* 2011;92:983-97.
- [40] Stallinga HA, Dijkstra PU, Bos I, Heerkens YF, Roodbol PF. The ambiguity of the concept of participation in measurement instruments: operationalization of participation influences research outcomes. *Clin Rehabil* 2014.
- [41] Wilkie R, Peat G, Thomas E, Croft PR. Measuring the consequences of osteoarthritis and joint pain in population-based studies: can existing health measurement instruments capture levels of participation? *Arthritis Rheum* 2004;51:755-62.
- [42] Fayed N, Cieza A, Bickenbach JE. Linking health and health-related information to the ICF: a systematic review of the literature from 2001 to 2008. *Disabil Rehabil* 2011;33:1941-51.
- [43] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-45.
- [44] Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.

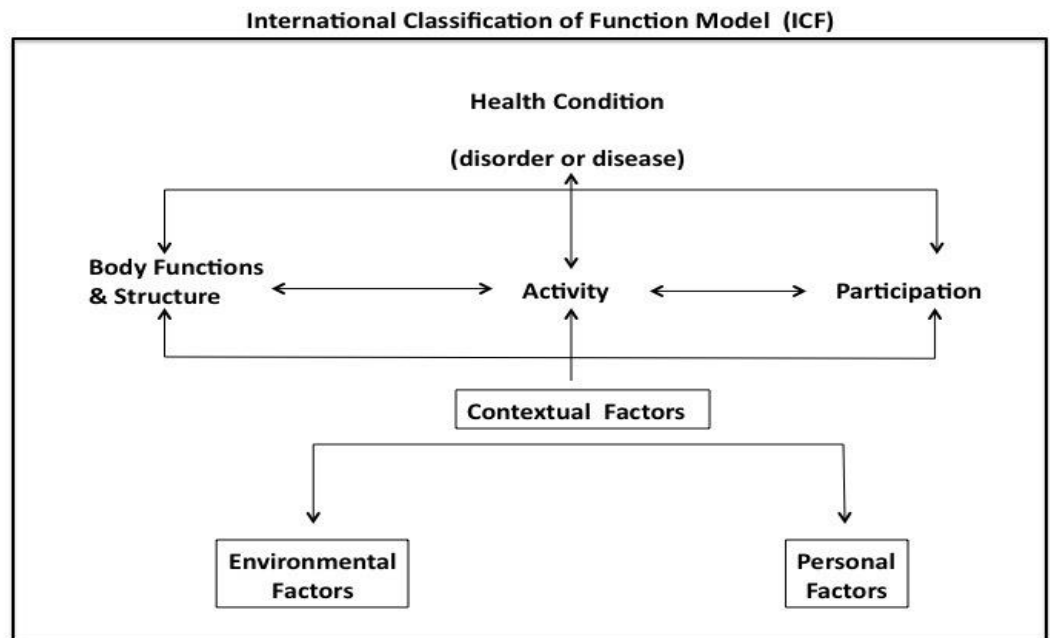
- [45] Streiner DL, Norman GR. Health measurement scales – a practical guide to their development and use. Third edition ed. New York: Oxford University Press; 2003.
- [46] Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- [47] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539–49.
- [48] Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res* 2012;21:659–70.



## Bijlage 1

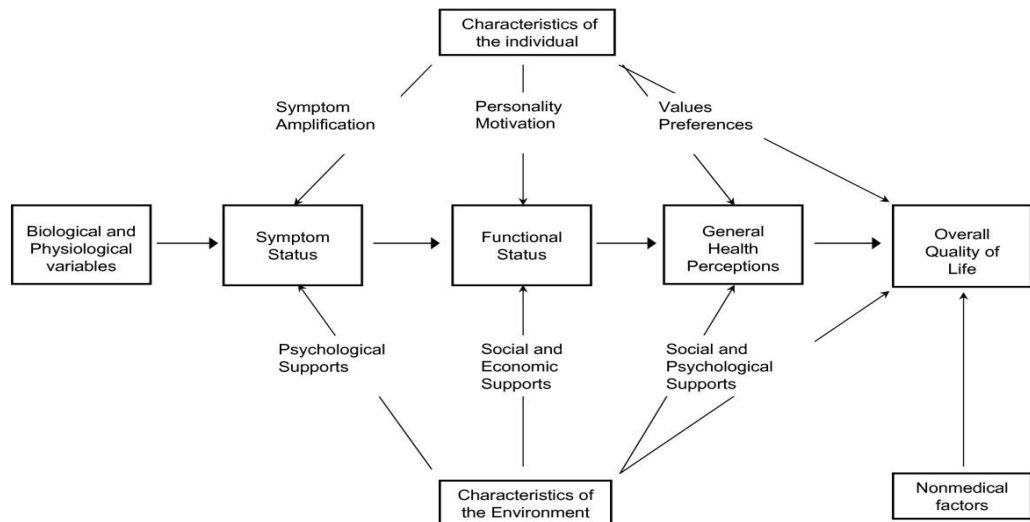
Modellen voor het in kaart brengen van domeinen van gezondheid en kwaliteit van leven

Figuur 3: ICF schema

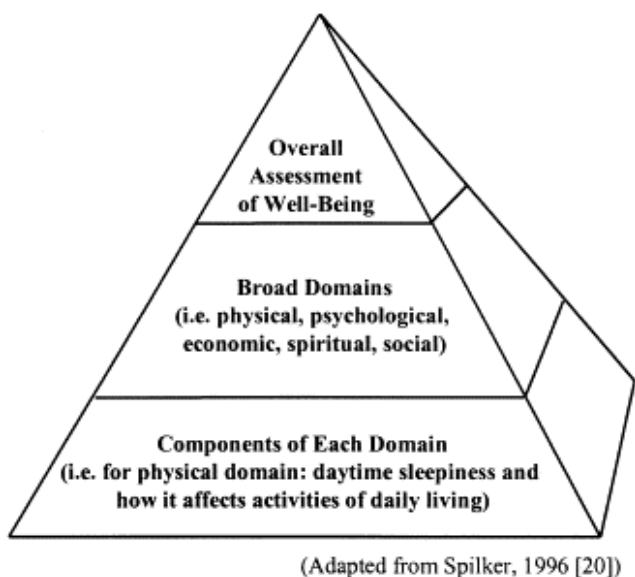


Adapted From: Model of Disability – ICF Model

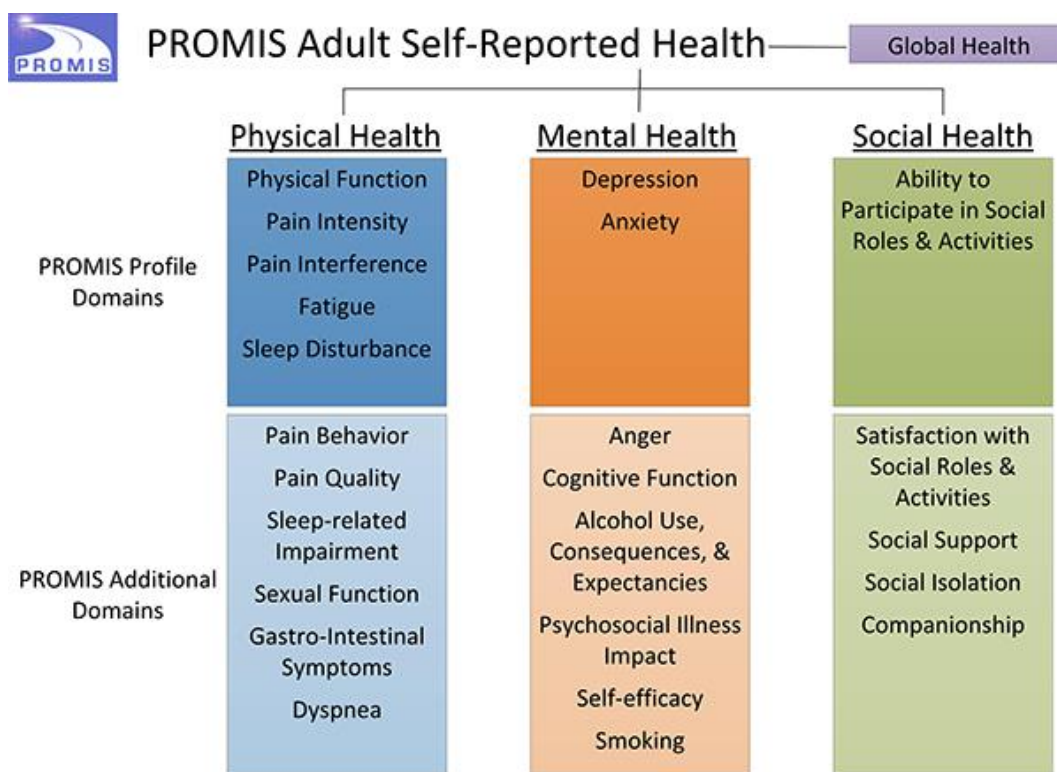
Figuur 4: Model Wilson & Cleary



Figuur 5: Model van Spilker



Figuur 6: Model van PROMIS



03/28/2014

Figuur 7: Model OMERACT

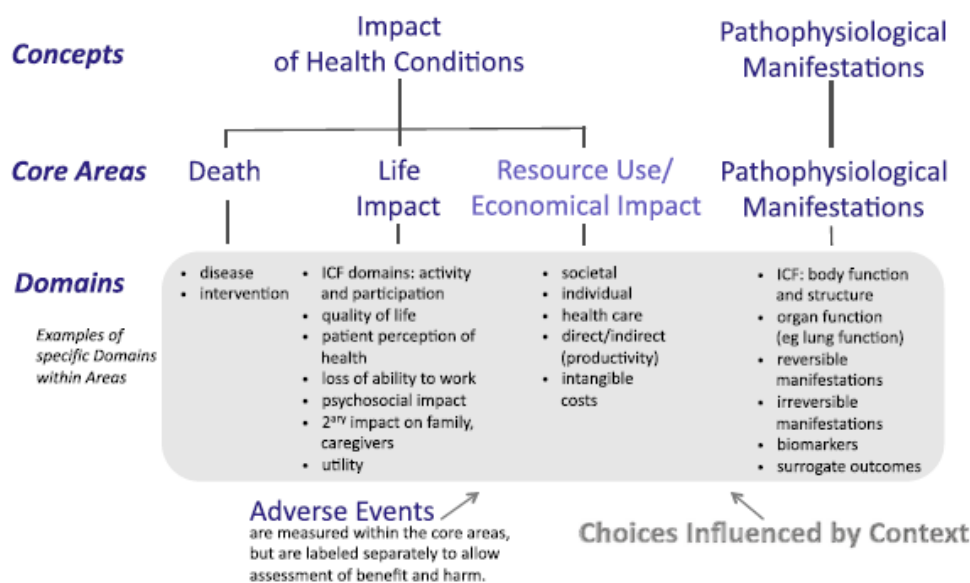
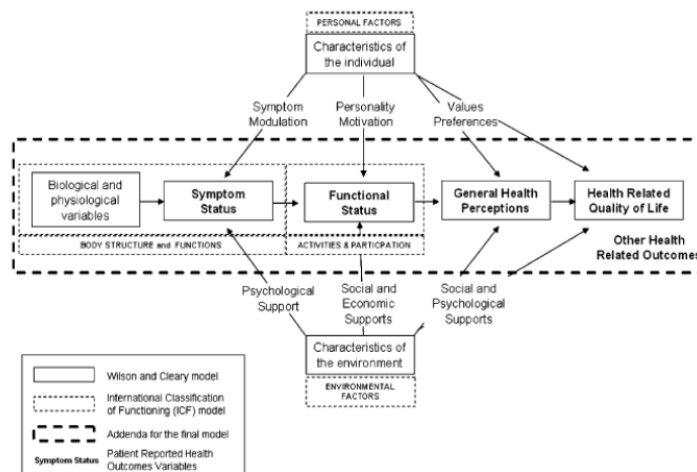


Fig. 1. Conceptual framework of Core Areas for outcome measurement in the setting of health intervention studies. Resource Use has a lighter shade to indicate it is currently strongly recommended, but not mandatory for inclusion. The choice of specific Domains within an Area depends on the context for which the core set is being developed in all areas, domains can be generic or made more specific, for example disease-specific, time-specific (eg, short or long-term), specific for patient preference, and so forth. ICF, International Classification of Functioning, Disability and Health.

Figuur 8: Model van Valderas & Alonso

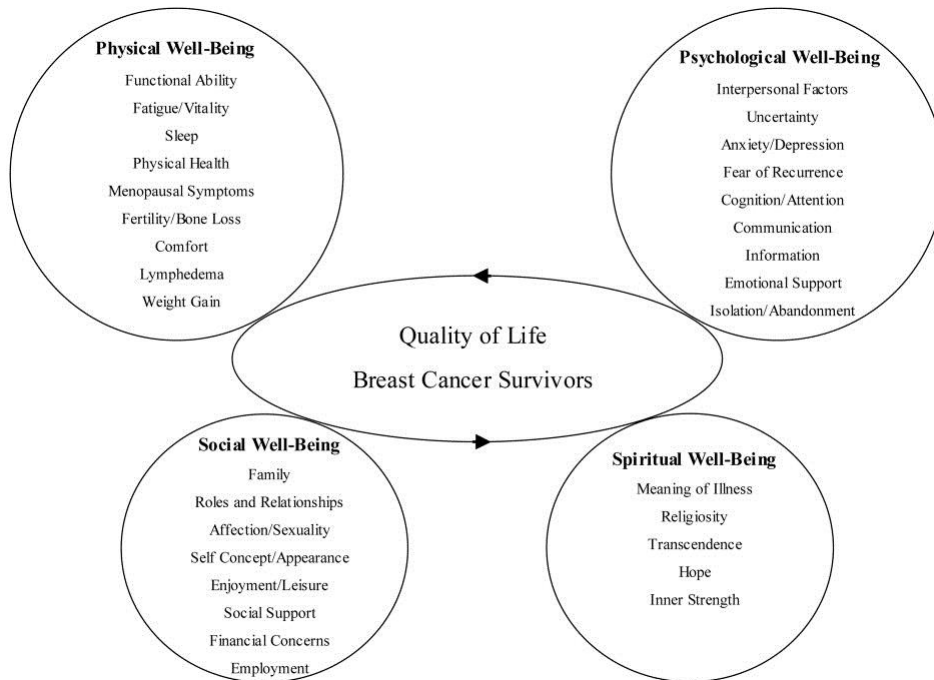
Fig. 1 An integrated model for health outcomes. Modified from [29] and [46]



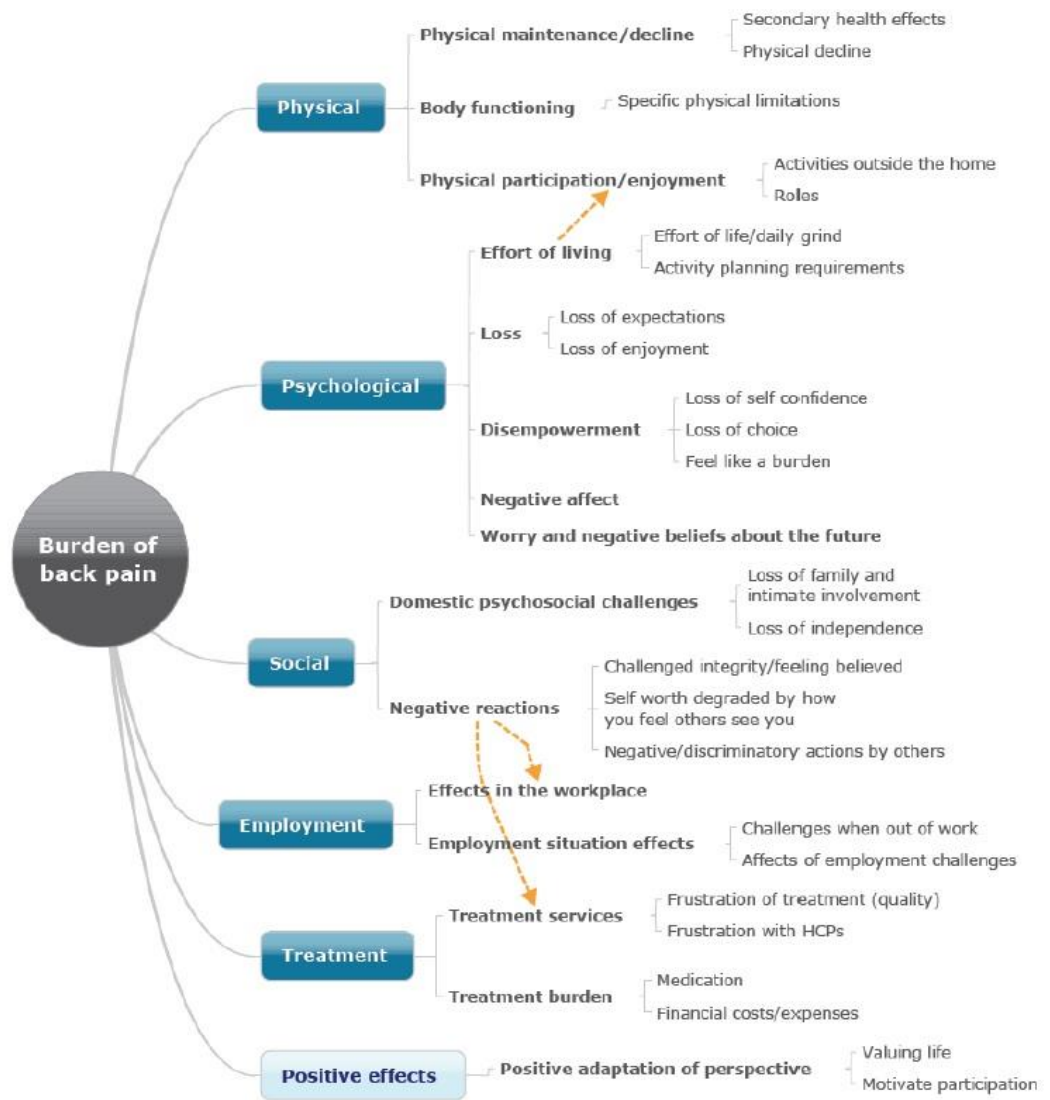
## Bijlage 2

Classificaties en modellen voor het in kaart brengen van relevante PROs

Figuur 9: Model van Dow & Ferrell [27]



Figuur 10: Conceptueel model Buchbinder [28]



Figuur 11: Deel van de ICF Core Set coronaire hartziekten betreffende activiteiten en participatie (bron: WHO)

**Categories of the component 'activities and participation':**

ICF Code	ICF Category Title
d230	Carrying out daily routine
d240	<b>Handling stress and other psychological demands</b>
d430	Lifting and carrying objects
d450	<b>Walking</b>
d455	<b>Moving around</b>
d460	Moving around in different locations
d470	Using transportation
d475	Driving
d480	Riding animals for transportation
d570	<b>Looking after one's health</b>
d620	<b>Acquisition of goods and services</b>
d630	Preparing meals
d640	Doing housework
d760	<b>Family relationships</b>
d770	<b>Intimate relationships</b>
d850	<b>Remunerative employment</b>
d920	Recreation and leisure

*Categories in bold belong to the Brief ICF Core Set for Chronic Ischemic Heart Disease.*

## Figuur 12: Linking van SF-12 en ICF [29]

Table III. *Short-Form-12 linked with the updated rules. (See Tables I and II)*

Item	ICF category	Additional information
1. In general, would you say your health is: (excellent, very good, good, fair, poor)	nd-gh	
2. Does your health now limit you in these activities? a. Moderate activities, such as, moving a table pushing a vacuum cleaner bowling or playing golf b. Climbing several flights of stairs	nd-gh A&P* (d430) Lifting and carrying objects (d4451) Pushing (d9201) Sports (d9201) Sports d4551 Climbing	Moderate A table A vacuum cleaner Bowling Playing golf Several flights of stairs
3. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of your physical health? a. Accomplished less than you would like b. Were limited in the kind of work or other activities	d850 Remunerative employment d230 Carrying out daily routine nd-ph	
4. During the past week, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? a. Accomplished less than you would like b. Didn't do work or other activities as carefully as usual	d850 Remunerative employment d230 Carrying out daily routine b152 Emotional functions (b152) Emotional functions (b152) Emotional functions	Depressed Anxious
5. During the past week, how much did pain interfere with your normal work (including both work outside the home and housework)?	b280 Sensation of pain d850 Remunerative employment d640 Doing housework	
6. These questions are about how you feel and how things have been with you during the past week. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past week: a. Have you felt calm and peaceful? b. Did you have a lot of energy? c. Have you felt downhearted and blue?	b152 Emotional functions b1300 Energy level b152 Emotional functions	Calm and peaceful Downhearted and blue
7. During the past week, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?	nd-ph b152 Emotional functions d9205 Socializing	

\* A&P = Activities and Participation.





## Colofon

Deze handreiking is opgesteld door het expertisenetwerk Patient-Reported Outcomes (PRO) van de Nederlandse Federatie van Universitair Medische Centra (NFU). Dit expertisenetwerk is onderdeel van het NFU-consortium Kwaliteit van Zorg. Binnen het NFU-expertisecentrum Kwaliteitsregistratie bundelt het NFU-consortium kennis op het gebied van kwaliteitsregistratie in expertisenetwerken. Bundeling, ontwikkeling en verspreiding van expertise rond het meten van kwaliteit met behulp van patiënt-gerapporteerde uitkomstmetingen (PROMs) vindt plaats in het NFU-expertisenetwerk PRO. Binnen het netwerk is een onderzoeksagenda opgesteld die drie thema's bevat: (1) opzetten van PRO registraties; (2) methodologie voor de selectie van PROs en PROMs; (3) ontwikkeling van kwaliteitsindicatoren. Deze handreiking voor de selectie van PROs en PROMs is het eerste product van het expertisenetwerk volgend uit de onderzoeksagenda.

Philip van der Wees (Radboudumc) en Caroline Terwee (VUmc) zijn co-voorzitters van het NFU-expertisenetwerk PRO. Andere deelnemers zijn: Neil Aaronson (NKI), Carla Bakker (NFU-consortium Kwaliteit van Zorg), Sandra Beurskens (MUMC+), Dolf de Boer (Nivel), Jozé Braspenning (NFU-consortium Kwaliteit van Zorg), Rianne van den Brink (TNO/CBO), Kalinka van de Camp (Radboudumc), Christine Fransman (DICA), Martha Grootenhuis (AMC), Rosella Hermens (Radboudumc), Xander Koolman (VU), Barbara van Leeuwen (UMCG), Hester Lingsma (Erasmus MC), Perla Marang (LUMC), Jan-Willem Mulder (NPCF), Annet Nugter (GGZ Noord-Holland-Noord), Lydia van Overveld (Radboudumc), Hein Raat (Erasmus MC), Henk van Stel (UMC Utrecht), Jolanda de Vries (UvT), Caroline van Weert (Stichting Miletus).

### Contact NFU-Expertisenetwerk Patient-Reported Outcomes

Dr. C.B. Terwee  
Kenniscentrum Meetinstrumenten  
Afdeling Epidemiologie en Biostatistiek  
VU Medisch Centrum  
Postbus 7057  
1007 MB Amsterdam  
cb.terwee@vumc.nl

Dr. P.J. van der Wees  
Radboud Institute for Health Sciences  
Afdeling IQ healthcare  
Radboudumc  
Postbus 9101  
6500 HB Nijmegen  
philip.vanderwees@radboudumc.nl

### NFU-consortium Kwaliteit van Zorg

p/a LUMC  
Postbus 9600  
2300 RC Leiden  
071 526 37 08  
kwaliteit@nfu.nl  
www.nfukwaliteit.nl



Februari 2015

NFU-consKvZ-1502.03/MV

