

# Tool Evalueren en optimaliseren

van patiëntervaringsvragenlijsten | versie 1.0



Zorginstituut Nederland



## Over deze tool

Deze tool geeft praktische aanwijzingen voor het evalueren en optimaliseren van vragenlijsten om de ervaringen van patiënten met de zorg te meten. De tool is bedoeld als hulpmiddel voor onderzoekers die zogeheten PREMs<sup>1</sup> of PROMs<sup>2</sup> willen evalueren. Doel van deze tool is om onderzoekers concrete handvatten te bieden voor het evalueren en optimaliseren van vragenlijsten waarin het patiëntenperspectief centraal staat.

Deze tool maakt deel uit van een reeks van tools over patiëntervaringsmetingen, over de onderwerpen privacy, dataverzameling, analyses en kwalitatieve methoden. De tools bieden verdiepende en praktische informatie in aanvulling op de Handreiking Ontwikkelen van Patiëntervaringsvragenlijsten om kwaliteit van zorg te meten (Bos et al., 2015). Deze handreiking legt op hoofdlijnen uit hoe partijen in de zorg vragenlijsten op een verantwoorde manier kunnen ontwikkelen en gebruiken om te komen tot valide en betrouwbare kwaliteitsinformatie. De criteria uit het Toetsingskader kwaliteitsstandaarden, informatiestandaarden en meetinstrumenten (Zorginstituut, 2015) zijn het uitgangspunt. Procedures en werkinstructies die eerder zijn ontwikkeld voor de Consumer Quality Index (CQ-index) methodiek waren de basis van de tools.

Zorginstituut Nederland biedt deze ondersteunende materialen aan om het meten, verzamelen en analyseren van kwaliteitsinformatie te standaardiseren. Hiermee wordt verdere verbetering van de kwaliteit van de gezondheidszorg in Nederland gestimuleerd en wordt ervoor gezorgd dat iedereen toegang heeft tot begrijpelijke en betrouwbare informatie over de kwaliteit van geleverde zorg. Onderzoekers van het Nederlands instituut voor onderzoek van de gezondheidszorg (NIVEL) en adviseurs van Zorginstituut Nederland hebben deze tool samengesteld

### Leeswijzer

Na een korte inleiding over de stappen en mogelijke methoden voor vragenlijstontwikkeling (H1) worden in deze ingegaan op het evalueren en optimaliseren van vragenlijsten (H2).

### Auteurs

Nivel: Mattanja Triemstra en Dolf de Boer

Zorginstituut Nederland: Laura Koopman en Marloes Zuidgeest

# 1. Vragenlijstontwikkeling

Patiëntervaringen kunnen met zogeheten PREM- of PROM-vragenlijsten in beeld worden gebracht. PREMs (Patient Reported Experience Measures) gaan vooral over de ervaringen met zorgverleners/-aanbieders en het zorgverleningsproces. PROMs (Patient Reported Outcome Measures) gaan over de ervaren uitkomsten van de zorg. Voorop staat bij beide soorten vragenlijsten dat ze het patiëntenperspectief weerspiegelen; ze gaan bij uitstek over onderwerpen van de zorg die voor patiënten belangrijk zijn. Om als indicator van de kwaliteit van zorg te kunnen worden gebruikt moeten ze valide<sup>3</sup> en betrouwbaar<sup>4</sup> zijn. Dit betekent dat de vragenlijsten samen met patiënten en op een wetenschappelijk verantwoorde wijze ontwikkeld en geëvalueerd moeten worden.

In Tabel 1 staan de stappen die doorlopen worden bij het ontwikkelen van vragenlijsten (zie voor meer informatie de tool *Kwalitatieve methoden*). Daarvan zijn de laatste stappen '5. Evalueren' en '6. Optimaliseren' zeer belangrijk voor het ontwikkelproces. Deze stappen kunnen – desgewenst – herhaald worden na iedere meting.

**Tabel 1. Stappen bij vragenlijstontwikkeling**

1. Genereren	Inventariseren van relevante onderwerpen voor de vragenlijst.
2. Analyseren	Benoemen, groeperen en ordenen van onderwerpen (door te coderen, clusteren en eventueel te prioriteren).
3. Selecteren	Selecteren van onderwerpen voor de vragenlijst.
4. Formuleren	Vragenlijst opstellen door onderdelen, instructies en vragen te formuleren op basis van de geselecteerde onderwerpen.
5. Evalueren	Vragenlijst uittesten en evalueren.
6. Optimaliseren	Vragenlijst bijstellen en eventueel inkorten.

Het uiteindelijke doel van vragenlijstontwikkeling is dat er een goede, valide en betrouwbare vragenlijst beschikbaar komt voor het meten van patiëntervaringen. Bovendien moet het een 'patiëntvriendelijke' vragenlijst zijn, met bewoordingen en vragen die aansluiten bij de taal van patiënten, een prettige lay-out en een acceptabele lengte. Met kwalitatieve methoden – voor het genereren van relevante onderwerpen (stap 1; zie tool *kwalitatieve methoden*) tot en met het evalueren van vragenlijsten (stap 5) – worden deze eisen al in grote mate gewaarborgd. Vervolgens kan de vragenlijst op basis van een kwantitatieve testfase (een zogeheten pilotmeting, psychometrische of klinimetrische test) verder worden geëvalueerd en geoptimaliseerd (stappen 5 en 6 van vragenlijstontwikkeling).

## 2. Evalueren en optimaliseren van vragenlijsten

Met het *evalueren* van vragenlijsten bedoelen we dat de uitkomsten van de vragen bekeken worden waarbij ook de validiteit en betrouwbaarheid gecheckt wordt. Dit kan tijdens het ontwikkelproces, na iedere meting of na een bepaalde periode plaatsvinden.

Met het *optimaliseren* van vragenlijsten wordt bedoeld dat met de vragenlijst beter (meer valide en betrouwbaar) en efficiënter de kwaliteit van zorg gemeten kan worden. Dit kan middels het stellen van andere vragen, maar ook door het stellen van minder vragen.

### 2.1 Evaluatiecriteria voor goede vragenlijsten

Bij de evaluatie van vragenlijsten kan gebruik gemaakt worden van het bepalen van de meeteigenschappen van vragen (2.1.1), betrouwbaarheidsanalyses (2.1.2) en diverse beoordelingscriteria (2.1.3).

#### 2.1.1. Meeteigenschappen

Bij de evaluatie van vragen is het verstandig de eigenschappen van de vragen (variabelen) te onderzoeken. Het gaat hierbij om de validiteit en verdeling van antwoorden op de vragen. De afwegingen en afkappunten van onderstaande eigenschappen staan beschreven in de tool *Analyses* (Zorginstituut Nederland / NIVEL, 2016).

Denk hierbij aan:

1. Scheefheid van de antwoorden op een vraag. Stel bijvoorbeeld dat meer dan 90% van de patiënten hetzelfde antwoordt op een vraag. Antwoorden die zo scheef verdeeld zijn, tonen geen verschillen tussen patiëntengroepen of zorgaanbieders waardoor dit een eigenschap kan zijn voor het verwijderen van een vraag voor een volgende meting.
2. Missende waarden per vraag. Wanneer een vraag heel vaak niet wordt beantwoord, of wordt beantwoord met "niet van toepassing" of "weet ik niet", dan kan dat een aanwijzing zijn dat de vraag niet goed begrepen wordt. Als vuistregel kunnen vragen met meer dan 5% missende waarden nader worden bekeken.
3. Samenhang tussen vragen. Indien vragen samenhangen met een correlatiecoëfficiënt van  $\geq 0,70$  is er aanzienlijke overlap in de informatie die beide vragen verschaffen. Dit kan aanleiding zijn om één van beide vragen niet verder te analyseren dan wel in een volgende meting met de vragenlijst te verwijderen.
4. Betrouwbaarheid van groepen vragen. Bij een lage (test-hertest) betrouwbaarheid of geen bijdrage aan een betrouwbare factor/ schaal/ thema kan besloten worden een vraag te verwijderen.
5. Onderscheidend vermogen per vraag of thema. Bij weinig of geen onderscheidend vermogen worden geen of nauwelijks verschillen tussen zorgaanbieders gemeten en kan besloten worden vragen niet mee te nemen bij een volgende meting.

Bovengenoemde eigenschappen zijn voorbeelden van redenen om vragen te behouden dan wel te verwijderen. Keuzes kunnen gemaakt worden met belanghebbende partijen, bijvoorbeeld in een project- of werkgroep (zie ook 2.2).

#### 2.1.2. Betrouwbaarheidsanalyses

Indien meerdere vragen een thema (factor, dimensie, composiet of schaal) representeren via een somscore over een aantal vragen of gemiddelde score, is het verstandig om met een betrouwbaarheidsanalyse (Cronbach's alfa coëfficiënt) te onderzoeken of een dergelijk thema betrouwbaar (intern consistent of homogeen) gemeten is bij een nieuwe meting. Dit is ook aan te bevelen als een vraag uit een thema verwijderd is (zie ook tool *Analyses*). Via deze check is zichtbaar of de ingevulde vragen samen een themascore opleveren die betrouwbaar is.

### 2.1.3. Beoordelingscriteria

Voor de beoordeling van patiëntveringsvragenlijsten kunnen algemeen geldende criteria voor vragenlijsten worden toegepast. Bijvoorbeeld die van de International Society for Quality of Life research (ISOQOL; Reeve et al., 2013). Deze criteria hebben achtereenvolgens betrekking op:

- Conceptueel en meetmodel<sup>5</sup>
- Inhoudsvaliditeit<sup>6</sup>
- Constructvaliditeit<sup>7</sup>
- Cross-culturele validiteit<sup>8</sup> (geldt voor vertaalde vragenlijsten)
- Betrouwbaarheid<sup>9</sup>
- Responsiviteit<sup>10</sup> (vooral bij PROMs)
- Hanteerbaarheid<sup>11</sup> en interpreteerbaarheid<sup>12</sup>

Zie voor een uitvoerige beschrijving van deze criteria de handreikingen over PROMs (Van Kessel et al., 2014; p. 18-25; Terwee et al., 2015) en de Indicatorstandaard over methodologische criteria voor de ontwikkeling van betrouwbare kwaliteitsindicatoren (Koolman et al., 2012).

Om er daarnaast zeker van te zijn dat vragenlijsten ook op een juiste manier zijn uitgetest en geëvalueerd, kan gebruik worden gemaakt van de COSMIN checklist (zie [www.cosmin.nl](http://www.cosmin.nl)). Dit is een internationale consensus-based checklist, ontwikkeld in een Delphi studie (Mokkink et al, 2010), voor het beoordelen van de kwaliteit van meeteigenschappen.

## 2.2. Inkorten van vragenlijsten

### 2.2.1 Belang van vragenlijsten inkorten

Patiëntveringsvragenlijsten zijn over het algemeen relatief lang. Dit komt bijvoorbeeld doordat ze vaak worden ontwikkeld voor meerdere gebruiksdoelen en toepassingen, zoals voor:

- publieke informatie (keuze-/consumenteninformatie voor patiënten);
- verbeter-/monitor- of benchmarkinformatie (voor instellingen/zorgaanbieders);
- zorginkoopinformatie (voor zorgverzekeraars); en/of
- als beleids-, verantwoordings- of toezichtinformatie (voor beleidsmakers, overheid en inspectie).

Hierdoor telden de eerder ontwikkelde CQ-index vragenlijsten vaak meer dan 50 vragen. Inmiddels zijn veel van deze vragenlijsten door verdere validatie/doorontwikkeling wel korter geworden.

Sinds 2015 is er een tendens om steeds kortere (PREM of PROM) vragenlijsten te maken, met een kernset bestaande uit circa 10 tot 15 vragen. Dit heeft voordelen voor de hanteerbaarheid en interpretatie, omdat patiënten niet onnodig met lange vragenlijsten worden belast en er maar weinig scores hoeven te worden berekend, gerapporteerd en geïnterpreteerd. Voorbeelden zijn de PREMs voor Ziekenhuiszorg, Oncologie, Geboortezorg, Chronische Zorg en Eerstelijnszorg die op initiatief van Stichting Miletus zijn ontwikkeld. Deze beknopte, generieke basissets kunnen door zorgaanbieders naar eigen inzicht en behoefte uitgebreid worden met gevalideerde optionele vragen of zogeheten 'add-ons' (zorg- of ziektespecifieke modules). Bijvoorbeeld om specifiek inzicht te krijgen in (mogelijkheden voor) kwaliteitsverbetering.

Tegelijk moeten de korte vragenlijsten geschikt zijn om verschillen tussen zorgaanbieders of instellingen te meten. Dat is zeker de wens van zorgverzekeraars, met het oog op de zorginkoop. Daarvoor gaat de aandacht vooral uit naar vragen over dingen die 'het verschil maken', met een groot onderscheidend vermogen. Dit levert ook voor patiënten zinvolle informatie op, omdat ze op basis hiervan zorgaanbieders kunnen vergelijken. Daarentegen zijn zorgaanbieders vooral geïnteresseerd in kwaliteitsverbetering; zij willen inzicht krijgen in het verbeterpotentieel en willen weten of verbetertrajecten resultaat hebben.

### 2.2.2 Informatie voor inkorten van vragenlijsten

De hiervoor genoemde meeteigenschappen en beoordelingslijstjes voor 'goede vragenlijsten' (zie 2.1.1 tot en met 2.1.3) zijn bruikbaar voor het inkorten van vragenlijsten. Naast de resultaten van de kwantitatieve testfase kunnen er ook kwalitatieve argumenten zijn voor het inkorten van vragenlijsten. Zo blijkt bij het uittesten met cognitieve interviews al of een vraag valide is, en kunnen de vragen uitvallen die

slecht begrepen worden. Of vragen kunnen geschrapt worden als ze bij de prioritering (zie concept mapping in de tool *Kwalitatieve methoden*) een relatief lage belangscore krijgen.

Overigens kunnen er naast statistische/methodische criteria ook redenen zijn vanuit de praktijk of het beleid die een rol spelen bij de keuze om items al dan niet te verwijderen. Dit kan in overleg tussen de belanghebbende partijen, bijvoorbeeld in een project- of werkgroep, worden besloten.

In box 1 staan (extra) criteria voor het inkorten van vragenlijsten, per gebruiksdoel ‘verschillen meten’ dan wel ‘kwaliteitsverbetering’, gebaseerd op een eerdere werkinstructie en artikel van Triemstra et al. (2008) over het inkorten van CQI-vragenlijsten.

**Box 1. Criteria voor het inkorten van patiëntervaringsvragenlijsten, afhankelijk van het doel**

Doel:	Criterium voor verwijdering van een item:
<p><b>A. Verschillen meten</b> tussen zorgaanbieders (instellingen of praktijken)</p>	<ul style="list-style-type: none"> <li>• <b>Weinig of geen verklaarde variantie:</b> variantie in de score wordt niet of nauwelijks verklaard door verschillen tussen zorgaanbieders (lage ICC, intraklasse correlatie).</li> <li>• <b>Weinig of geen spreiding tussen zorgaanbieders:</b> score vertoont nauwelijks of geen spreiding, (bijna) alle zorgaanbieders scoren ‘gemiddeld’.</li> <li>• <b>Minimale verschillscore:</b> nauwelijks of geen verschil tussen de beste en slechtste instellings-/praktijkscore.</li> </ul>
<p><b>B. Kwaliteitsverbetering</b> door zorgaanbieders</p>	<ul style="list-style-type: none"> <li>• <b>Weinig of geen verbeterpotentieel:</b> nauwelijks of geen verschil tussen het ‘overall’ item-gemiddelde, en de item-score van ‘best practices’.</li> <li>• <b>Weinig of geen suboptimale ervaringen:</b> nauwelijks of geen patiënten die een ‘suboptimale’ of ‘negatieve’ ervaring hebben (laag % ‘nee’, ‘nooit’/‘soms’, ‘oneens’ etc.).</li> <li>• <b>Relatief minder belangrijk item</b> (lage belangscore).</li> </ul>

Gebaseerd op: Triemstra et al., 2008

Er zijn verschillende manieren om naar het belang en de toepassing van vragen te kijken, en dit is bepalend voor afwegingen over welke vragen wel of niet behouden moeten blijven in een vragenlijst. Voor de besluitvorming is het van belang dat de verschillende belanghebbende partijen in goed overleg de gebruiksdoelen bespreken, en afwegingen maken omtrent de verdere optimalisatie van vragenlijsten.

## Bronnen en verder lezen

- Bos N., Zuidgeest M., Kessel P. van, Boer D. de. Handreiking Ontwikkelen van patiëntervaringsvragenlijsten om kwaliteit van zorg te meten. Utrecht: NIVEL/Zorginstituut Nederland, 2015.
- Kessel P. van, Triemstra M., Boer D. de. Handreiking voor het meten van kwaliteit van zorg met Patient Reported Outcome Measures. Utrecht: NIVEL/Zorginstituut Nederland, 2014.
- Koolman X., Zuidgeest M. Indicatorstandaard 2.0. Diemen: Zorginstituut Nederland, 2012.
- Mokkink L.B., Terwee C.B., Patrick D.L., Alonso J., Stratford P.W., Knol D.L., 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.
- Reeve B.B. et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*, 2013 Oct;22(8):1889-905
- Terwee C.B., Wees P.J. van der, Beurskens S., namens het NFU-expertisenetwerk Patient-reported outcomes. Handreiking voor de selectie van PROs en PROMs. Leiden: NFU- consortium, 2015. Kwaliteit van Zorg.
- Toetsingskader kwaliteitsstandaarden, informatiestandaarden en meetinstrumenten. Zorginstituut Nederland, 2015.
- Tool Analyses voor het vergelijken van patiëntervaringen tussen zorgaanbieders. Zorginstituut Nederland/NIVEL, 2016.
- Tool Dataverzameling voor het meten van patiëntervaringen. Zorginstituut Nederland/NIVEL, 2016.
- Tool Kwalitatieve methoden. Zorginstituut Nederland/NIVEL, 2016.
- Tool Privacy en bescherming van persoonsgegevens. in ontwikkeling.
- Triemstra et al. Doelspecifieke versies van CQ-index meetinstrumenten: korter, krachtiger, en specifiekere meten? *TSG: Tijdschrift voor Gezondheidswetenschappen*: 2008, 86(8), 486-493

## Verklarende woordenlijst

- 1 PREMs: Patient Reported Experience Measures zijn de door patiënten/cliënten gerapporteerde ervaringen met het zorgproces
- 2 PROMs: Patient Reported Outcome Measures zijn de door patiënten/cliënten gerapporteerde uitkomsten van de zorg
- 3 Valide vragenlijst: een geschikte vragenlijst die meet wat hij zou moeten meten
- 4 Betrouwbare vragenlijst: een nauwkeurige vragenlijst die precies meet, zonder meetfouten
- 5 Conceptueel en meetmodel: het concept of begrip dat de vragenlijst beoogt te meten met de onderliggende theorieën en veronderstelde relaties
- 6 Inhoudsvaliditeit: de mate waarin de vragenlijst de betekenis van een begrip (zoals kwaliteit van zorg) in al zijn aspecten omvat
- 7 Constructvaliditeit: de begripsvaliditeit of mate waarin alle vragen samen het beoogde begrip (theoretische construct) dekken, zodat de resultaten van metingen met de vragenlijst een indicator zijn voor het begrip dat je wilt meten (bijvoorbeeld kwaliteit van zorg)
- 8 Cross-culturele validiteit: de mate waarin de vragen en meetresultaten in verschillende landen op eenzelfde manier geïnterpreteerd kunnen worden
- 9 Betrouwbaarheid: de nauwkeurigheid of mate waarin de vragenlijst precies meet, zonder meetfouten (vrij van de invloed van toevallige factoren)
- 10 Responsiviteit: validiteit van een verschilscore van twee metingen (bijvoorbeeld voor en na een behandeling)
- 11 Hanteerbaarheid: bruikbaarheid en toepasbaarheid in de praktijk, zowel voor patiënten (zoals leesbaarheid, belasting, invultijd) als zorgverleners (bijvoorbeeld het gemak waarmee scores kunnen worden berekend)
- 12 Interpreteerbaarheid: de mogelijkheid om scores te interpreteren