



HANDBOEK CQI METINGEN

versie: 2.0
d.d.: 1 november 2008
deel: 2

Handboek CQI Metingen: richtlijnen en voorschriften voor metingen met een CQI meetinstrument

Deel 2: Opschoning databestanden en (vergelijkende) analyses

Eindredactie: Herman Sixma
Michelle Hendriks
Dolf de Boer
Diana Delnoij

NIVEL: Herman Sixma, Diana Delnoij, Janine Stubbe, Mattanja
Triemstra, Olga Damman, Michelle Hendriks, Dolf de Boer
Miletus : Barbara Vriens, Caroline van Weert
Significant: Clare Wilkinson
Mailstreet: Rutger van Zuidam
AMC : Ines Rupp, Guus ten Asbroek

Utrecht, 1 november 2008



HANDBOEK CQI METINGEN

versie: 2.0
d.d.: 1 november 2008
deel: 2

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



Inhoud

Introductie	5
1 Controle databestanden	7
1.1 Het opschonen van gegevensbestanden	7
1.2 Responsanalyses	10
1.3 Verificatie van de taxonomie van het meetinstrument	13
2 Berekening CQI scores	15
2.1 Hercoderen van antwoordcategorieën	15
2.2 Berekening CQI scores per kwaliteitsdimensie	16
3 Vergelijkende analyses en case-mix adjustment	19
3.1 Case-mix adjusters, case-mix adjustment en het bepalen van verschillen tussen zorgaanbieders	19
Literatuur	23



HANDBOEK CQI METINGEN

versie: 2.0
d.d.: 1 november 2008
deel: 2



Introductie

Dit deel van het Handboek CQI metingen geeft richtlijnen en aanbevelingen voor het opschonen en analyseren van een databestand bij een CQI meting.

Deel 2 van dit handboek bevat de volgende hoofdstukken:

1. **controle databestanden**
2. **berekening CQI scores**
3. **vergelijkende analyses en case-mix adjustment**



HANDBOEK CQI METINGEN

versie: 2.0
d.d.: 1 november 2008
deel: 2



1 Controle databestanden

Het is van cruciaal belang dat de databestanden bij een CQI meting zo weinig mogelijk fouten bevatten. In dit hoofdstuk wordt ingegaan op:

- opschonen gegevensbestanden;
- responsanalyses;
- verificatie taxonomie CQI meetinstrument.

1.1 Het opschonen van gegevensbestanden

Deze sectie beschrijft de diverse onderdelen en stappen van het opschonen van CQI gegevensbestanden. Deze opschoningstappen worden uitgevoerd op het responsbestand (zie deel 1, paragraaf 5.2). Het is belangrijk om de opschoningstappen in de hier beschreven volgorde uit te voeren, per stap bij te houden hoeveel lijsten worden verwijderd en in ieder geval na substap 1.2, 3.1 en 3.4 het bestand als een nieuwe versie op te slaan. Dit vergemakkelijkt het berekenen van de parameters ‘netto benaderd’, ‘bruto respons’ en ‘netto respons’, welke in paragraaf 1.2 worden toegelicht. Omdat in de meeste gevallen ook per zorgaanbieder/verzekeraar gecontroleerd wordt hoeveel data er verloren gaat en waaraan dit kan worden toegeschreven, adviseren we om niet alleen het totaal aantal verwijderde lijsten te registreren, maar deze registratie ook uit te splitsen per zorgaanbieder/verzekeraar.

Voor een beknopt overzicht van de opschoningprocedure verwijzen wij naar schema 1.1. Een uitgebreide toelichting op schema 1.1 volgt hieronder:

1 Controle

1.1 Check op de kwaliteit van de data-invoer

Alvorens te beginnen met opschonen dient het databestand te zijn gecontroleerd. Als dat nog niet is gebeurd dienen willekeurig 75 ingevulde vragenlijsten vergeleken te worden met de gegevens in het opgeleverde responsbestand, of in geval van een handmatige invoer dient 10% van de vragenlijsten met een minimum van 10 vragenlijsten twee keer te worden ingevoerd en gecheckt op inconsistenties. Vervolgens dient het databestand aan de hand van frequentieverdelingen te worden gecontroleerd op onmogelijke waarden (bijvoorbeeld de waarde ‘7’ op een vierpuntsschaal). Een uitgebreide instructie voor deze controles is te vinden in het Handboek CQI Metingen deel 1, paragraaf 5.3.

1.2 Controle op dubbel ingevoerde respondenten/ vragenlijsten

Dit komt voor wanneer een vragenlijst twee keer wordt ingevoerd of wanneer een



respondent bij zowel de eerste zending als bij een herinnering een vragenlijst terugstuurt. Bij dubbele lijsten/gegevens wordt de eerst ontvangen versie (eerste ontvangstdatum) bewaard en de latere versie verwijderd, tenzij de eerste versie leeg is en de tweede versie is ingevuld. Dubbele respondenten/vragenlijsten zouden in principe niet voor moeten komen omdat deze in een eerdere stap al verwijderd worden (zie ook paragraaf 5.2 van deel 1 van dit handboek).

2 Mensen verwijderen die ten onrechte waren aangeschreven

2.1 Verwijderen van lijsten retour wegens 'overleden'

Lijsten die retour kwamen omdat de beoogde respondent overleden is worden verwijderd.

2.2 Verwijderen van lijsten 'onbestelbaar retour'

Lijsten die onbestelbaar retour kwamen worden verwijderd.

2.3 Verwijderen van respondenten die niet tot de doelgroep behoren

Respondenten die niet behoren tot de doelgroep: in de eerste vragen in het meetinstrument wordt gevraagd of een bepaald ziektebeeld, behandeling of opname op de respondent van toepassing is. Indien de respondent heeft aangegeven dat dit niet van toepassing is, of indien de respondent niets heeft ingevuld (er blijkt niet dat het wel van toepassing is), wordt de respondent verwijderd. Respondenten waarbij onduidelijk is op welke analyse eenheid hun ervaringen betrekking hebben worden ook verwijderd (bijvoorbeeld als zij een andere zorgaanbieder of verzekeraar opgeven dan waar zij volgens de steekproefopzet geregistreerd staan). Een aandachtspunt hierbij is dat 'lege lijsten' in dit stadium nog behouden moeten blijven zodat we verderop in de procedure inzicht kunnen krijgen in het aantal lege lijsten. Dit betekent dat wanneer een respondent wordt verwijderd onduidelijk is of deze behoort tot de doelgroep, deze respondenten alleen verwijderd dienen te worden als de respondent de vragenlijst wel deels heeft ingevuld.

3 Mensen verwijderen die wel behoren tot de doelgroep, maar die de vragenlijst onvoldoende of niet op de juiste manier hebben ingevuld

3.1 Verwijderen lege cases

Teruggestuurde blanco lijsten dienen te worden verwijderd. We spreken van blanco indien er minder dan 5 vragen zijn beantwoord.

3.2 Verwijderen van respondenten die de lijst niet zelf hebben ingevuld én beantwoord

Enige hulp bij het invullen van de vragenlijst is uiteraard toegestaan. Echter, als respondenten de vragenlijst door een ander hebben laten invullen én beantwoorden, dan dienen zij te worden verwijderd. De optie 'heeft de vragen in mijn plaats beantwoordt' op de vraag 'Hoe heeft die persoon u geholpen' is dan aangevinkt. Bij deze schoningsstap geldt als vuistregel dat we het zekere voor het onzekere nemen. Zo wordt een respondent bijvoorbeeld ook verwijderd indien deze aangeeft hulp te hebben ontvangen, maar niet wat voor hulp.

3.3 Screenervragen

Screenervragen zijn vragen waarin wordt bepaald of een vervolgvraag op de respondent van toepassing is. Achter één van de mogelijke antwoorden op de



screenervraag staat dan de instructie dat de respondent de vervolgvraag of vervolgvragen niet hoeft in te vullen. Voorbeeld: ‘Rookt u?’ (antwoordcategorieën: ‘ja’ en ‘nee’, waarbij de ‘nee’ categorie wordt gevolgd door de verwijzing ‘ga door naar vraag ...’). Voor vervolgvragen geldt dat de screenervraag leidend is. De eventuele antwoorden op vervolgvragen van respondenten die op de screenervraag hebben geantwoord dat de vraag niet op hen van toepassing is, worden verwijderd. Wanneer de screenervraag ‘missing’ is, dienen antwoorden op vervolgvragen ook te worden verwijderd. Deze stap hoeft niet te worden uitgevoerd voor de psychometrische testfase van een CQI Meetinstrument, omdat in dit traject ook inzicht dient te worden verkregen in de werking van de screenervragen.

3.4 *Verwijderen van lijsten waarin onvoldoende vragen zijn beantwoord*

Cases met onvoldoende waarnemingen op sleutelvragen dienen te worden verwijderd. De sleutelvragen zijn vragen die alle respondenten moeten beantwoorden, inclusief de volgende demografische vragen:

- Leeftijd;
- Geslacht;
- Opleiding;
- Taal thuis gesproken;
- Geboorteland vader;
- Geboorteland moeder;
- Geboorteland uzelf;
- Algemene ervaren lichamelijke gezondheid.

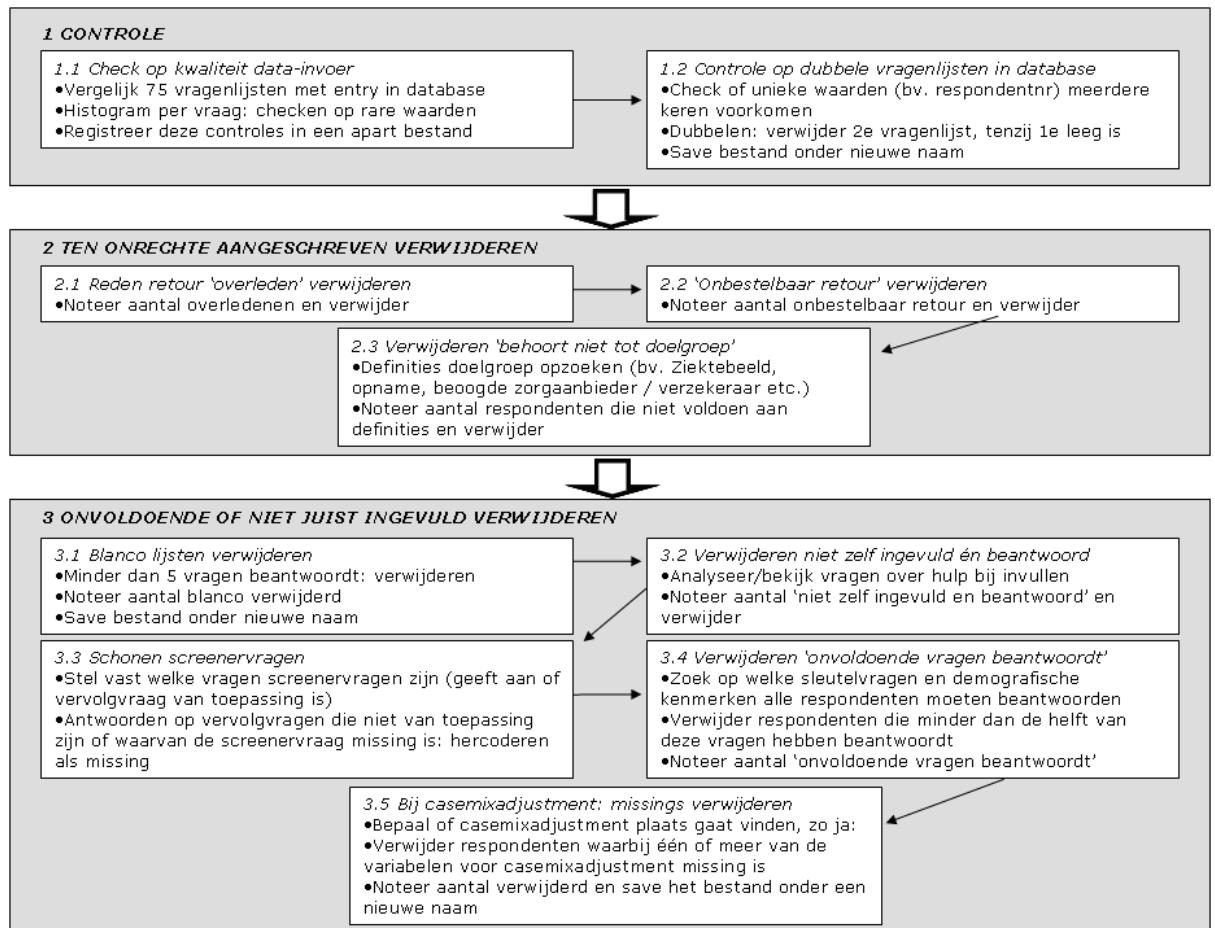
Een vraag is geen sleutelvraag wanneer deze conform een voorafgaande screenervraag mag worden overgeslagen. Van de sleutelvragen, dient 50% te zijn ingevuld. Vragenlijsten waarvan minder dan 50% van de sleutelvragen zijn ingevuld, worden verwijderd.

3.5 *Bij casemixadjustment: verwijderen van respondenten bij wie één of meer van de variabelen voor casemix adjustment missing is*

Casemix adjustment kan alleen worden uitgevoerd wanneer de informatie over variabelen waarvoor gecorrigeerd wordt, compleet is. Wanneer casemix adjustment onderdeel vormt van het project, dienen respondenten waarbij de variabelen voor casemix adjustment niet compleet zijn dan ook te worden geschoond uit de dataset. Dit geldt ook als bij een variabele voor casemix adjustment de categorie ‘anders, namelijk:’ is aangevinkt. Variabelen voor casemix adjustment zijn doorgaans namelijk ordinaal van aard en de antwoordcategorie ‘anders, namelijk:’ is normaal gesproken niet eenduidig in deze orde in te passen. Wanneer er geen casemix adjustment plaats vindt in een project hoeft deze schoningsstap niet te worden uitgevoerd. Dit geldt bijvoorbeeld voor CQ-index ontwikkeltrajecten waarin geen discriminerend vermogen wordt getoetst.



Schema 1.1 Schematisch overzicht van de opschoningprocedure: stappen, substappen en bijbehorende handelingen.



1.2 Responsanalyses

De eerste stap in de analyse is het bepalen van de bruto en netto respons. Daarbij worden de definities gehanteerd zoals weergegeven in schema 1.2. Wanneer data worden verzameld per post of via het internet, wordt vaak gesproken over bijvoorbeeld 'totaal verstuurd' en 'netto verstuurd'. Echter data worden ook wel mondeling verzameld middels bijvoorbeeld interviews en in dergelijke gevallen is geen sprake van het 'versturen' van vragenlijsten. Vandaar dat we in onderstaand schema spreken van 'benaderd' in plaats van 'verstuurd'.

*Schema 1.2 Definities bruto en netto respons*

Term	Definitie
Totaal benaderd	Het totaal aantal benaderde mensen
Netto benaderd	Het totaal aantal benaderde mensen minus: <ul style="list-style-type: none">- gevallen waar de verzekerde overleden is- gevallen waar de verzekerde niet tot de onderzoekspopulatie behoort- vragenlijsten die onbestelbaar retour komen- gevallen die op de rip-lijst stonden
Bruto respons	Het aantal ingevulde vragenlijsten, oftewel 'netto benaderd' minus de lijsten die niet retour of blanco retour zijn (zie ook onder)
Netto respons	Het aantal bruikbare en ingevulde vragenlijsten, oftewel 'bruto respons' minus het aantal vragenlijsten die in de schoningsacties als onbruikbaar zijn bepaald
Netto benaderd percentage	$100 \times (\text{'netto benaderd'} / \text{'totaal benaderd'})$
Bruto responspercentage	$100 \times (\text{'bruto respons'} / \text{'netto benaderd'})$
Netto responspercentage	$100 \times (\text{'netto respons'} / \text{'netto benaderd'})$

Het vaststellen van de parameters 'Netto benaderd', 'Bruto respons' en 'Netto respons' hangt nauw samen met de opschoningprocedure (zie ook paragraaf 1.1):

- Netto benaderd = totaal benaderd – (aantal verwijderd in stap 2.1, 2.2 en 2.3 van de opschoningprocedure, alsmede elders geregistreerde 'overledenen' en 'onbestelbaar retour'). Het is belangrijk om te beseffen dat 'overledenen' en 'onbestelbaar retour' zowel in het bestand kunnen voorkomen als in een aparte registratie. Dit is bijvoorbeeld afhankelijk van in hoeverre lijsten 'onbestelbaar retour' of 'overleden' bij de data-entry terecht komen.
- Bruto respons = aantal cases na stap 3.1 van de opschoningprocedure, gegeven dat de opschoningprocedure is gedaan in de volgorde zoals beschreven in paragraaf 1.1
- Netto respons = aantal cases aan het einde van de opschoningprocedure.

De responsresultaten dienen vergeleken te worden met de bevindingen in het validatierapport over de ontwikkeling van de vragenlijst. Als er sprake is van een beduidend lager responspercentage vergeleken met eerdere (pilot-) onderzoeken dan is het zaak om de mogelijke oorzaken te inventariseren en deze op te nemen in de rapportage en in de meetverantwoording voor het CKZ. Voorbeelden van mogelijke



oorzaken zijn: problemen in de gegevensverzameling, de periode van gegevensverzameling ongelukkig gekozen, verschillen in de doelgroep ten opzichte van de pilotonderzoeken.

Responsanalyse

Vergelijking van netto respons en steekproef

Het centrale doel van een responsanalyse is om te toetsen in hoeverre de netto respons representatief is voor de populatie waarover het onderzoek iets beoogt te zeggen. De representativiteit wordt getoetst door de netto respons te vergelijken met netto benaderd¹, uitgesplitst naar geslacht of uitgesplitst naar leeftijd. Er is sprake van een representatieve netto respons als de netto responspercentages per leeftijdscategorie of sekse ongeveer gelijk is aan het netto benaderd percentage. Dit kan getoetst worden door de responsgegevens in een kruistabel te analyseren met behulp van de chi-kwadraat toets. Bij leeftijd kan ook een t-toets worden gebruikt.

Het is denkbaar dat, in het geval van onder- of oververtegenwoordiging van belangrijke groepen, resultaten die hierdoor beïnvloed worden niet worden gepresenteerd. We spreken van onder- of oververtegenwoordiging indien:

- Aanzienlijke verschillen optreden in de netto respons percentage en netto benaderd percentage (gebruik hierbij de chi-kwadraat toets).
- Deze verschillen inconsistent zijn met het validatierapport. In het validatierapport wordt een diepgaande analyse uitgevoerd van de representativiteit. In enkele gevallen volgt hieruit dat onder- of oververtegenwoordiging van bepaalde groepen geaccepteerd is voor het betreffende CQI Meetinstrument.

Vaak is onder- of oververtegenwoordiging niet echt een probleem: als onder- of oververtegenwoordiging op bepaalde variabelen de data beïnvloedt, dan wordt dat doorgaans ondervangen door een case-mix adjustment op die variabelen (zie ook hoofdstuk 3). Niettemin dient men wel aan deze onder- of oververtegenwoordiging te refereren bij de interpretatie van de resultaten en aan te geven wat de invloed daarvan is. Uiteraard geldt dit wanneer de chi-kwadraat toets een significant verschil oplevert.

Vergelijking netto respons en nonrespondenten

Naast de responsanalyse voor representativiteit dient men ook een responsanalyse te doen om inzicht te krijgen in selectiebias. In dit geval wordt de netto respons vergeleken met de mensen die wel waren aangeschreven, maar niet tot de netto respons behoren. Ook hierbij wordt de chi-kwadraat toets gebruikt en worden de resultaten vergeleken met het validatierapport. Wanneer er een significant verschil optreedt tussen de netto-respons en geadresseerden die niet tot de netto-respons behoorden, dan volgt een vergelijk tussen de netto-respons en de onderzoekspopulatie. Dit is nodig om vast te stellen in hoeverre een selectiebias bij het terugsturen van vragenlijsten ten koste gaat van de representativiteit

Het is hier wellicht verleidelijk om de netto respons te vergelijken met totaal verstuurd (oftewel: totale steekproef). Dit doen we echter niet omdat totaal verstuurd nog personen bevat die ten onrechte waren aangeschreven en waarnaar we dus niet willen generaliseren.



ten opzichte van de onderzoekspopulatie.

Controle van gegevensverzameling

Uiteraard zijn er verschillen in respons tussen zorgaanbieders/verzekeraars.

Responsgegevens kunnen ook worden gebruikt ter controle van de gegevensverzameling.

Deze controle vindt plaats per zorgaanbieder / verzekeraar en omvat het volgende:

- De gegevensverzameling wordt gecontroleerd door per zorgaanbieder de netto respons te bekijken. Dit doet men middels een overzicht per zorgaanbieder/verzekeraar van de parameters ‘totaal benaderd’, ‘netto benaderd’, ‘bruto respons’ en ‘netto respons’ (zie schema 1.3 voor een voorbeeld).
- Dit overzicht maakt het mogelijk om grote verschillen in de netto respons tussen zorgaanbieders/verzekeraars toe te schrijven aan elementen uit de gegevensverzameling. Dit is van belang om een lage netto respons voor bepaalde zorgaanbieders te verantwoorden. Een lage respons voor een zorgaanbieder kan bijvoorbeeld voorkomen als een zorgaanbieder een steekproef aanlevert waarin relatief veel mensen zitten die niet voldoen aan de steekproefcriteria.

Een belangrijke vraag in dit verband is wanneer een verschil in netto respons tussen zorgaanbieders groot is (oftewel: welke zorgaanbieders/verzekeraars dienen verantwoording af te leggen wegens een lage respons). Op dit moment zijn er geen harde cut-off criteria te noemen; het CKZ vormt hier een beeld van op basis van de beschikbare empirische gegevens. In afwachting hiervan adviseren we contact op te nemen met het CKZ bij twijfel over een lage respons.

Schema 1.3 Voorbeeld van een responstabel per zorgaanbieder

Zorgaanbieder	Totaal benaderd	Netto benaderd	Bruto respons	Netto respons
#	#	# (% van totaal benaderd)	# (% van netto benaderd)	# (% van netto benaderd)
1	390	360 (92%)	276 (77%)	220 (61%)
2	405	385 (95%)	283 (74%)	210 (55%)

1.3 Verificatie van de taxonomie van het meetinstrument

Tijdens het ontwikkeltraject van de vragenlijst zijn schalen en dimensies benoemd (zie ‘Handboek CQI Ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument’, paragraaf 5.3.3 en 5.3.4). Bij analyse van een CQI meting dient te worden gecontroleerd of de oorspronkelijke schalen en dimensies nog betrouwbaar zijn. Bij metingen in meerdere instellingen dient dit niet per instelling plaats te vinden, maar op een groot bestand met daarin alle, of in ieder geval een groot deel van deze instellingen. De betrouwbaarheid van de schalen uit het ontwikkeltraject wordt onderzocht door berekening van de betrouwbaarheidscoëfficiënt (Cronbach’s alfa). Bij



HANDBOEK CQI METINGEN

versie: 2.0
d.d.: 1 november 2008
deel: 2

grote verschillen in Cronbach's alfa ten opzichte van eerdere metingen en rapporten dienen verklaringen te worden onderzocht, bijvoorbeeld door te kijken naar mogelijke verschillen in de (steekproef)populatie of case-mix verschillen.



2 Berekening CQI scores

Bij het berekenen van CQI-scores wordt onderscheid gemaakt tussen een gemiddelde score per kwaliteitsdimensie en een gemiddelde verdeling van scores over antwoordcategorieën van een kwaliteitsdimensie. In een aantal gevallen vinden ook hercoderingen plaats, waarbij diverse antwoordcategorieën samen worden genomen.

In dit hoofdstuk wordt achtereenvolgens aandacht geschonken aan:

- Hercoderen van antwoordcategorieën
- Berekening CQI scores per kwaliteitsdimensie

2.1 Hercoderen van antwoordcategorieën

Voor enkele ervaringsvragen en respondentkenmerken worden in rapportages niet de oorspronkelijke antwoorden gebruikt maar een verdere categorisering daarvan. Voorbeelden van vragen die vaak worden gehercodeerd zijn algemene waarderingsvragen (rapportcijfers) waarbij de antwoorden 0 t/m 5 worden samengenomen, en vragen naar geboorteland ouders². Daarnaast worden de vragen met de antwoordcategorieën ‘nooit’, ‘soms’, ‘meestal’ en ‘altijd’ gehercodeerd naar drie antwoordcategorieën door ‘nooit’ en ‘soms’ samen te nemen. Het kan ook voorkomen dat bijvoorbeeld de vraag naar opleiding of de vraag naar leeftijd verder wordt gecategoriseerd omdat het aantal categorieën tot onoverzichtelijke resultaten leidt. De wijze van verdere categorisatie van antwoordcategorieën kan verschillen tussen meetinstrumenten en settings. In geval van een op zichzelf staand onderzoeksproject is het aan de onderzoeker om, in overeenstemming met de andere stakeholders, te bepalen welke hercategorisatie het beste aansluit bij de situatie. Echter, wanneer resultaten van verschillende partijen zoveel mogelijk op elkaar aan moeten sluiten is het van belang dat dergelijke hercategorisaties overal gelijk verlopen. Dit geldt bijvoorbeeld voor landelijke metingen waarbij meetbureaus resultaten terugkoppelen aan hun klanten, maar waarbij er ook nog centrale

² Indien een variabele ‘etniciteit’ benodigd is, kan deze conform de CBS-definitie als volgt worden geconstrueerd:

- Allochtoon: ten minste één van beide ouders van de respondent is in het buitenland geboren;
- Autochtoon: beide ouders zijn in Nederland geboren.



bewerkingen plaatsvinden.

Nagenoeg alle CQI-vragen zijn gesloten vragen met een beperkt aantal antwoordcategorieën.

Echter, in sommige vragen wordt een (open) toelichting gevraagd op de antwoordcategorie 'anders'. Schema 2-1 geeft een voorbeeld. De categorie 'anders, namelijk...' wordt als één categorie beschouwd.

Schema 2-1 Voorbeeld van een gesloten vraag met een open toelichting

<p>Verplicht: In welke taal praat u thuis het meeste?</p> <p><input type="checkbox"/> Nederlands <input type="checkbox"/> Fries <input type="checkbox"/> Nederlands dialect <input type="checkbox"/> Indonesisch <input type="checkbox"/> Sraran (Surinaams) <input type="checkbox"/> Marokkaans-Arabisch <input type="checkbox"/> Turks <input type="checkbox"/> Duits <input type="checkbox"/> Papiaments (Nederlandse Antillen) <input type="checkbox"/> Anders, namelijk: (a..u.b. in blokletters)</p>

2.2 Berekening CQI scores per kwaliteitsdimensie

Bij de berekening van gemiddelde scores voor de kwaliteitsdimensies wordt onderscheid gemaakt tussen de gemiddelde score en de gemiddelde verdeling over de antwoordcategorieën. De gemiddelde score wordt gebruikt voor het toetsen op significante verschillen, terwijl de gemiddelde verdeling over de categorieën van belang is voor de grafische weergave in staafdiagrammen.

Berekening van gemiddelde score per kwaliteitsdimensie

Alleen respondenten die de helft of meer van de vragen per dimensie hebben beantwoord, worden meegenomen in de bepaling van het dimensiegemiddelde. Vragen die beantwoord zijn met 'weet ik niet' of 'niet van toepassing' gelden hierbij als missing omdat zij geen logische plaats hebben in de volgorde van antwoordcategorieën. Vragen die conform een screenvraag niet ingevuld hoefden te worden gelden ook als missing. Het kwantitatieve gemiddelde wordt als volgt bepaald:

- Bepaal per dimensie het gemiddelde per respondent. Dit is gelijk aan het (ongewogen) gemiddelde van de scores op de afzonderlijke vragen;
- Het totale gemiddelde per dimensie is het gemiddelde van de respondentgemiddelden.



De berekeningen van het dimensiegemiddelde worden geïllustreerd aan de hand van een voorbeeld zoals weergegeven in schema 2-2. Dit schema bevat de scores van vijf respondenten op drie vragen die tot één kwaliteitsdimensie behoren. In de laatste kolom worden de berekeningen weergegeven voor het kwantitatieve dimensiegemiddelde.

Schema 2-2 Voorbeeld berekening gemiddelde per dimensie;

	Vraag 1	Vraag 2	Vraag 3	Dimensiegemiddelde
Respondent 1	2	4	4	$10/3 = 3,333$
Respondent 2	3	4	4	$11/3 = 3,666$
Respondent 3	2	3	Missing	$5/2 = 2,5$
Respondent 4	4	4	3	$11/3 = 3,666$
Respondent 5	Missing	1	Missing	Mag niet meedoen
Totaal				3,292

Berekening van gemiddelde verdeling over antwoordcategorieën van een kwaliteitsdimensie

De gemiddelde verdeling over de categorieën wordt als volgt bepaald (zie schema 2-4 voor een voorbeeld):

- Bepaal per vraag de percentuele verdeling over antwoordcategorieën;
- De percentuele verdeling voor de dimensie is het gemiddelde van de percentuele verdelingen per vraag.

Alleen respondenten die meer dan de helft van de vragen per dimensie hebben beantwoord, worden meegenomen in de analyse

In tabel 2-3 wordt de verdeling van de antwoorden van de vijf respondenten op de drie vragen weergegeven. In de onderste rij wordt de gemiddelde verdeling voor het thema bepaald.

Schema 2-3 Voorbeeld berekening gemiddelde verdeling over categorieën

	Categorie 1	Categorie 2	Categorie 3	Categorie 4
Vraag 1	0%	25%	25%	50%
Vraag 2	0%	0%	25%	75%
Vraag 3	0%	33%	33%	33%
Gemiddelde verdeling	0%	$58\%/3 = 19\%$	$83\%/3 = 28\%$	$158\%/3 = 53\%$

Bijzondere situaties

Soms komt het voor dat de onderliggende vragen van een dimensie niet op een consistente manier zijn gesteld. Dat vraagt de nodige aandacht bij de bepaling van het



themagemiddelde. Over het algemeen verdient het de voorkeur om dergelijke bijzondere situaties al bij de ontwikkeling van de vragenlijst te vermijden maar dat is niet altijd mogelijk. Het betreft de volgende situaties:

- 1) ***Sommige vragen zijn andersom gesteld.*** De meeste ervaringsvragen zijn zodanig gesteld dat het antwoord ‘altijd’ of ‘ja’ op een positieve ervaring duidt. Sommige ervaringsvragen worden echter andersom gesteld. Een voorbeeld is de vraag ‘hoe vaak had u overlast van andere patiënten op de kamer’ (bron: de CQI heup/knie operatie’) waarbij de antwoordcategorie ‘altijd’ op een negatieve ervaring duidt. Als dergelijke ‘andersom’-vragen onderdeel uitmaken van een dimensie, moeten zij vóór het bepalen van het dimensiegemiddelde zodanig gecodeerd worden dat de meest positieve antwoordcategorie hetzelfde cijfer krijgt als de meest positieve antwoordcategorie voor de overige vragen (bijvoorbeeld bij een vraag met vier antwoordcategorieën: 1 wordt 4; 2 wordt 3; 3 wordt 2; 4 wordt 1).
- 2) ***De vragen binnen een kwaliteitsdimensie maken gebruik van verschillende aantallen antwoordcategorieën.*** Het kan zijn dat een dimensie bestaat uit een combinatie van vragen met twee antwoordcategorieën (‘ja’ en ‘nee’) en vragen met vier antwoordcategorieën (‘nooit’, ‘soms’, ‘meestal’, ‘altijd’). In dat geval moeten de vragen met twee antwoordcategorieën vóór het bepalen van het dimensiegemiddelde als volgt gecodeerd worden:
 - het antwoord ‘ja’ krijgt dezelfde cijfer als het antwoord ‘altijd’ (de waarde ‘4’)
 - het antwoord ‘nee’ krijgt dezelfde cijfer als het antwoord ‘nooit’ (de waarde ‘1’).
- 3) ***Sommige vragen binnen een kwaliteitsdimensie worden voorafgegaan door een screenervraag.*** Het kan zijn dat sommige vragen binnen een kwaliteitsdimensie alleen van toepassing zijn op een deel van de respondenten, te bepalen door een screenervraag vooraf. Een voorbeeld van een screenervraag is de vraag ‘Hebt u tijdens deze ziekenhuisopname naar pijnstillers moeten vragen?’ (bron: de algemene ziekenhuis vragenlijst). Alleen respondenten die bevestigend beantwoorden krijgen vervolgens een vraag over de ervaring met pijnstillers. De manier van berekening van het dimensiegemiddelde blijft ongewijzigd in het geval van screenervragen. Echter de kans dat respondenten uitvallen door de randvoorwaarde dat respondenten meer dan de helft van de vragen per dimensie moeten hebben beantwoord, wordt met screenervragen groter. Toch moet deze randvoorwaarde blijven – immers het dimensiegemiddelde moet zo veel mogelijk een weergave zijn van de ervaring op alle onderliggende vragen.



3 Vergelijkende analyses en case-mix adjustment

Bij vergelijkende analyses dienen de data te worden gecorrigeerd voor case-mix om invloeden van elementen waar zorgaanbieders niets aan kunnen/moeten doen eruit te halen (bijvoorbeeld leeftijd en geslachtssamenstelling van respondenten). Case-mix gecorrigeerde data worden gebruikt om gemiddelde scores per kwaliteitsdimensie te berekenen en te vergelijken, en om de verdeling over antwoordcategorieën te berekenen.

3.1 Case-mix adjusters, case-mix adjustment en het bepalen van verschillen tussen zorgaanbieders³

Via case-mix adjustment wordt gecorrigeerd voor verschillen tussen zorgaanbieders (of verzekeraars) in de samenstelling van de respondenten die van invloed zijn op de verschillen, maar die niet te maken hebben met kwaliteit van zorg of service (O'Malley, A. J., Zaslavsky, A. M., Elliott, M. N., Zaborski, L., & Cleary, P. D., 2005; Zaslavsky, A. M., 2001; Zaslavsky, A. M. et al., 2008). Het doel is een eerlijke vergelijking van zorgaanbieders of zorgverzekeraars. In de validatie dient per vraag te worden bepaald of deze voor casemix correctie in aanmerking komt en zo ja, welke respondentkenmerken moeten worden meegenomen als case-mix adjuster. Kenmerken die veel gebruikt worden als case-mix adjuster zijn leeftijd, opleiding en algemene ervaren gezondheid (O'Malley, A. J., Zaslavsky, A. M., Elliott, M. N., Zaborski, L., & Cleary, P. D., 2005; Zaslavsky, A. M. et al., 2008). Voor meer informatie over de motivatie en de overwegingen bij case-mix adjustment wordt de lezer verwezen naar het Handboek CQI Ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument. Case-mix adjustment wordt gebruikt bij de analyse van ervaringsvragen, thema's en waarderingvragen. Waar totaalcijfers over alle respondenten en zorgaanbieders worden berekend is alleen case-mix adjustment nodig wanneer deze totaalcijfers worden vergeleken met andere (totaal)cijfers in andere steekproeven.

³ Het CKZ schrijft voor dat case-mix adjustment 'multi-level' moet worden uitgevoerd (tenzij er een goede reden is om hiervan af te wijken). Diverse meet- en analyseorganisaties zullen hiervoor niet zelf de expertise in huis hebben. Organisaties die niet over deze expertise beschikken kunnen de case-mix adjustment het beste outsourcen. Overleg hiervoor ook met het CKZ.



Uitvoering case-mix adjustment en formaat case-mix adjusters

De uitvoering van de case-mix adjustment vindt op twee manieren plaats afhankelijk van het doel van de analyse:

- Als het doel van de analyse betreft het onderzoeken van significante verschillen of het berekenen van gecorrigeerde waarderingscijfers, dan worden de case-mix adjusters meegenomen in een multilevel regressie.
- Als het doel van de analyse betreft het berekenen van de verdeling over de categorieën van ervaringsvragen of thema's na case-mix adjustment, dan wordt de data *per respondent* gecorrigeerd in een multilevel regressieanalyse. Vervolgens wordt de verdeling berekend met de gecorrigeerde data.

De case-mix adjusters kunnen als kwantitatieve variabele worden meegenomen, maar ook als categorie variabele. Hier wordt een beslissing over genomen bij het bepalen van het discriminerend vermogen, welke gevolgd moet worden (zie ook 'Handboek CQI Ontwikkeling: richtlijnen en voorschriften voor de ontwikkeling van een CQI meetinstrument', hoofdstuk 6). Deze beslissing wordt ook vermeld in de vragenlijst specifieke richtlijnen.

Toepassing van modellen voor casemix adjustment

Vanuit pragmatisch oogpunt wordt de case-mix adjustment bij voorkeur uitgevoerd met behulp van een lineaire regressie, omdat de meeste statistische software, zoals SPSS, SAS en MLwiN, lineaire regressies multi-level uit kunnen voeren. Echter, lineaire regressies zijn niet in alle gevallen geschikt:

- Lineaire regressies mogen worden toegepast voor case-mix adjustment indien de te corrigeren waarde zich bevindt op een schaal die ten minste ordinaal van aard is en bestaat uit ten minste drie categorieën. Voorbeelden hiervan zijn vragen met de antwoordcategorieën 'nooit', 'soms', 'meestal' en 'altijd' en vragen met de antwoordcategorieën 'geen probleem', 'klein probleem' en 'groot probleem';
- Indien er maar twee antwoordcategorieën zijn of de antwoordcategorieën zijn niet ordinaal van aard, dan dient men een logistische regressie uit te voeren. Voorbeelden hiervan zijn vragen met antwoordcategorieën 'ja/nee' of bijvoorbeeld de vraag 'Wat voor soort verzekering heeft u?' met antwoordcategorieën 'zorg in natura', 'restitutie' en 'combinatie van zorg en restitutie'. SPSS is niet in staat een logistische regressie multi-level uit te voeren maar in SAS en MLwiN kan dat wel.

Case-mix adjustment wordt toegepast om vergelijkingen tussen zorgaanbieders eerlijker te maken. Case-mix adjustment hoeft dus alleen te worden toegepast bij vragen waarop een vergelijking tussen zorgaanbieders plaats vindt. Dergelijke vragen kennen doorgaans drie of vier antwoordcategorieën die ordinaal van aard zijn, waardoor lineaire regressie kan worden toegepast.

Het berekenen van gecorrigeerde waarderingscijfers en het bepalen van significante verschillen

De case-mix adjustment wordt uitgevoerd door de case-mix adjusters als voorspellende variabelen mee te nemen in een lineaire regressieanalyse. Een voorbeeld van een dergelijk regressiemodel is:



$Y_{ij} = a + \text{zorgaanbieder}_j + b_2 \text{leeftijd}_{ij} + b_3 \text{opleiding}_{ij} + b_4 \text{gezondheid}_{ij} + \text{error}_{ij}$
waarbij Y_{ij} de score is van respondent i en zorgaanbieder j , 'zorgaanbieder' de variabele waar men in geïnteresseerd is en 'leeftijd' en 'opleiding' de case-mix adjusters. De parameters a , b_1 enzovoort worden geschat door het regressiemodel te fitten op de CQI vragenlijst dataset. Vervolgens kan de gecorrigeerde score per zorgaanbieder berekend worden door in het regressiemodel referentiewaarden in te vullen voor de case-mix adjusters; voor deze referentiewaarden wordt in veel gevallen de gemiddelde waarde over alle zorgaanbieders gekozen.

Als onderdeel van de analyse wordt een toets uitgevoerd op de significantie van verschillen tussen zorgaanbieders. Individuele aanbieders worden daarin direct vergeleken met elkaar en met het gemiddelde over alle aanbieders. Deze informatie is nodig om het belang van verschillen tussen zorgaanbieders te kunnen beoordelen. Alleen waar verschillen significant zijn, kan gesproken worden van feitelijke verschillen die niet het gevolg zijn van toevalligheden in de meting.

De toets wordt uitgevoerd door het 95% vergelijkingsinterval te berekenen voor de gemiddelde *case-mix gecorrigeerde* score van de individuele zorgaanbieder. Dit vergelijkingsinterval wordt berekend als het casemixgecorrigeerde gemiddelde plus of min 1,39 maal de standaarderror (Goldstein, H. & Healy, M. J. R., 1995). Als twee vergelijkingsintervallen geen overlap vertonen is er sprake van een significant verschil. Belangrijk is dat de case-mix adjusters meegenomen worden in de berekening van de gemiddelden en de vergelijkingsintervallen.

In bovenstaande toets worden individuele aanbieders direct vergeleken met elkaar en met het gemiddelde over alle aanbieders. Hiernaast is er nog een andere toets waarmee wordt vastgesteld of de totale variantie tussen zorgaanbieders significant groter dan nul is. Dit wordt gedaan door de zogenoemde intra-class correlatie op te vragen in de Multi-level analyse. De intra-class correlatie staat voor het percentage van de totale variantie dat kan worden toegeschreven aan verschillen tussen zorgaanbieders (Snijders, T. A. B & Bosker, R. J., 1999).

Toepassing van een lineaire regressie veronderstelt dat de responsvariabele normaal verdeeld is. Hoewel de meeste variabelen uit de CQ-index strikt genomen niet normaal verdeeld zijn, is lineaire regressie toch een redelijke strategie om een aantal redenen. Ten eerste geldt voor waarderingsvragen dat zij 11 antwoordcategorieën kennen, waardoor zij bij benadering als normaal verdeeld worden beschouwd. Ten tweede geldt voor de ervaringsvragen en thema's, die twee tot vier antwoordcategorieën kennen, dat lineaire regressie een valide methodiek is voor de toetsing van significantie als er sprake is van een groot aantal waarnemingen. Ten derde is het toetsen van significantie met behulp van lineaire regressie eenduidig en makkelijk uit te voeren met behulp van de gangbare statistische software.



Case-mix adjustment voor het berekenen van de verdeling over de categorieën van ervaringsvragen of thema's

Globaal omvat de strategie voor het berekenen van een case-mix gecorrigeerde verdeling per zorgaanbieder over de categorieën van ervaringsvragen of thema's het volgende: er wordt een casemix gecorrigeerde score berekent *per respondent* waarna de gecorrigeerde scores worden verdeeld over categorieën.

De case-mix gecorrigeerde score per respondent wordt geschat met behulp van het volgende model:

$Y_{ij} = a + b_1 \text{leeftijd}_{ij} + b_2 \text{opleiding}_{ij} + b_3 \text{gezondheid}_{ij} + \text{zorgaanbieder}_j + \text{error}_{ij}$, waarbij Y_{ij} , de score is van respondent i behorende tot zorgaanbieder j , en 'leeftijd' en 'opleiding' de case-mix adjusters. Het instellingseffect (instelling) wordt hier ook geschat opdat dit effect behouden kan blijven na case-mix adjustment. De case-mix gecorrigeerde score wordt berekend door voor de case-mix adjusters (leeftijd en gezondheid e.d.) de referentiewaarde in te vullen; voor deze referentiewaarden wordt in veel gevallen de gemiddelde waarde over alle zorgaanbieders gekozen.

De uitkomsten van deze procedure dienen te worden afgerond. Bij een vraag met drie antwoordcategorieën (1, 2 en 3) gaat dit bijvoorbeeld als volgt: 0 tot 1,5 wordt 1; 1,5 tot 2,5 wordt 2 en 2,5 en hoger wordt 3. Bij een vraag met vier antwoordcategorieën ziet dat er als volgt uit: 0 tot 1,5 wordt 1; 1,5 tot 2,5 wordt 2; 2,5 tot 3,5 wordt 3; 3,5 en hoger wordt 4.



Literatuur

Goldstein, H. & Healy, M. J. R. (1995). The Graphical Presentation of a Collection of Means. *Journal of the Royal Statistcial Society. Seris A (Statistics in Society)*, 158, 175-177.

O'Malley, A. J., Zaslavsky, A. M., Elliott, M. N., Zaborski, L., & Cleary, P. D. (2005). Case-mix adjustment of the CAHPS Hospital Survey. *Health Serv.Res.*, 40, 2162-2181.

Snijders, T. A. B. & Bosker, R. J. (1999). *Multilevel Analysis: An introduction to basic and advanced multilevel modeling*. Sage Publishers.

Zaslavsky, A. M. (2001). Statistical issues in reporting quality data: small samples and casemix variation. *Int.J.Qual.Health Care*, 13, 481-488.

Zaslavsky, A. M., Zaborski, L. B., Ding, L., Shaul, J. A., Cioffi, M. J., & Cleary, P. D. (2008). Adjusting Performance Measures to Ensure Equitable Plan Comparisons. *Health care financing review*, 22, 109-126.