



## Centraal Bureau voor de Statistiek

Divisie Methodologie en Kwaliteit  
Sectoren Methodologie Den Haag en Heerlen

---

### METHODOLOGISCH ONDERZOEK BIJ HET HERONTWERP VAN DE GEZONDHEIDSENQUÊTE 2010

**Fannie Cobben, Joeri Roels, Fong Yee Wong, Bart Buelens, Reinder  
Banning, Jan van den Brakel en Barry Schouten**

*Samenvatting: In het kader van het Herontwerp Persoonsenquêtes is de waarneming van de Gezondheidsenquête veranderd. Het face-to-face onderzoek is vervangen door een mixed-mode design waarbij naast face-to-face ook Internet waarneming en telefonische waarneming worden ingezet. Daarnaast zijn er ook een aantal wijzigingen in de vragenlijst doorgevoerd. In deze nota worden methodologische onderzoeken samengevat die in verband met deze veranderingen uitgevoerd zijn. In het bijzonder wordt ingegaan op methodebreuken die door de wijzigingen ontstaan. Daarnaast worden andere aanpassingen in het productieproces beschreven.*

*Trefwoorden: Kwaliteit, methodenbreuken, weging*

#### 1. Inleiding

Het programma Herontwerp Persoonsenquêtes (HPE) is in 2007 geïnitieerd om de persoonsenquêtes van het CBS efficiënter en flexibeler te maken tegen acceptabele kwaliteit (Cuppen, 2007). De belangrijkste aspecten van het herontwerp zijn de invoering van een basisvragenlijst en de overgang naar mixed-mode waarneming. Het design dat daarvoor is gekozen, is internetwaarneming gevolgd door een mix van face-to-face waarneming en telefonische waarneming.

In 2010 is voor de Gezondheidsenquête (GE) de mixed-mode waarneming ingevoerd. Van februari tot en met april en van juli tot en met oktober 2010 is het oude design nog uitgevoerd naast het nieuwe design; het paralleltraject. Het doel van

---

Projectnummer:

206413

BPA-nummer:

DMK-2011-12-12-FCBN-JRLS-FWNG-BBUS-RBNG-JBRL-BSTN

Datum:

12 december 2011

het paralleltraject is het analyseren van methodebreuken en het vergelijken van verschillende dimensies van kwaliteit voor het oude en het nieuwe design.

In deze nota beschrijven we de verschillende methodologische onderzoeken die zijn uitgevoerd in het deelproject Kwaliteit en Breuken voor de invoering van het mixed-mode design bij de GE in 2010. De onderzoeken hebben geleid tot aanpassingen in het productieproces die noodzakelijk zijn om rekening te houden met de veranderingen als gevolg van het HPE. Met deze nota krijgen interne gebruikers van de GE een overzicht over de aanpassingen. Meer informatie over de afzonderlijke onderzoeken is te vinden in de nota's uit de referentielijst.

In paragraaf 2 worden het oude en het nieuwe design van de GE beschreven. In paragraaf 3 gaan we in op de verschillen in kwaliteit tussen het oude en het nieuwe design. Vervolgens wordt in paragraaf 4 uitgelegd hoe de weging van de GE is aangepast aan het nieuwe design. In paragraaf 5 wordt ingegaan op de analyse van de breuken als gevolg van de vragenlijst verandering en de mode verandering. We eindigen de nota met conclusies en aanbevelingen.

## **1. Vergelijking oude en nieuwe design**

In het kader van het Herontwerp Persoonsenquêtes (HPE; Cuppen, 2007) is de waarnemingstrategie aangepast. Onder het oude design werden de personen door een interviewer thuis benaderd en werd de elektronische vragenlijst via een persoonlijk vraaggesprek ingevuld (Computer Assisted Personal Interviewing, CAPI). Onder het nieuwe design wordt een mixed-mode strategie toegepast: allereerst worden alle personen benaderd om via Internet mee te doen met het onderzoek (Computer Assisted Web Interviewing, CAWI). De non-respondenten van de Internet benadering worden vervolgens telefonisch benaderd (Computer Assisted Telephone Interviewing, CATI) als het CBS het telefoonnummer kan achterhalen, en de overige personen worden face-to-face benaderd (CAPI).

Oorspronkelijk was het de bedoeling om zowel vaste telefoonnummers als ook mobiele nummers te gebruiken bij de CATI-benadering. Omdat de respons via mobiele nummers echter bijzonder laag was, is vanaf februari 2011 ervoor gekozen om mobiele nummers via CAPI te benaderen. Vanaf augustus 2011 zijn de mobiele nummers echter wel weer via CATI benaderd.

Onder het nieuwe design is ook de vragenlijst aangepast. Dit heeft te maken met twee ontwikkelingen. Ten eerste is in 2009 de basisvragenlijst ingevoerd (Cuppen et al., 2009). Het is de bedoeling dat deze vragenlijst in de toekomst in alle CBS-enquêtes afgenomen wordt. Sommige vragen uit deze vragenlijst kwamen van oudsher in de GE voor, andere vragen in de basisvragenlijst zijn nieuw voor de GE. Ten tweede is in de vragenlijst de vraagstelling over medische consumptie

veranderd<sup>1</sup>. Daarnaast is de vragenlijst op een paar punten aangepast om de duur van het interview via CATI te verkorten.

Voor de opbouw van het mixed-mode design is in december 2009 gestart met Internet waarneming. Nonrespondenten via Internet met een geregistreerd telefoonnummer zijn in januari met CATI herbenaderd. De overige nonrespondenten zijn in februari met CAPI herbenaderd. Vanaf februari 2010 was het volledige mixed-mode design in het veld, d.w.z. dat de GE elke maand zowel via Internet als via CATI en CAPI waargenomen is.

Tegelijkertijd is er nog waargenomen volgens het oude (CAPI) design om de effecten van het herontwerp goed te kunnen vaststellen (Cuppen et al., 2009; Roels, Van den Brakel en Van Hertem, 2009). In mei en juni heeft echter de waarneming van het oude design stil gelegen vanwege een andere prioritering van enquêtes. In juli is de waarneming van het oude design weer hervat. De periode waarin het oude en het nieuwe design parallel uitgevoerd zijn is van februari tot en met april en juli tot en met oktober, zie Tabel 2.1.

*Tabel 2.1: Overzicht van het nieuwe en het oude design in 2010.*

	2009		2010										
	dec	jan	Feb	mrt	apr	mei	jun	jul	aug	sep	okt	nov	dec
<b>Nieuwe design</b>	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI	CAWI
		CATI	CATI	CATI	CATI	CATI	CATI	CATI	CATI	CATI	CATI	CATI	CATI
			CAPI	CAPI	CAPI	CAPI	CAPI	CAPI	CAPI	CAPI	CAPI	CAPI	CAPI
<b>Oude design</b>			CAPI	CAPI	CAPI			CAPI	CAPI	CAPI	CAPI	CAPI	CAPI

## 2. Kwaliteitsmeting

In 2008 is binnen het deelproject Kwaliteit en Breuken een kwaliteitskader voor het HPE opgesteld (Projectgroep HPE Kwaliteit en Breuken, 2008a en 2008b). Hiervoor zijn een aantal kwaliteitsdimensies en bijbehorende indicatoren uitgewerkt, die voor de enquêtes in het HPE vóór en na het herontwerp berekend worden. De veranderingen in kwaliteit als gevolg van het herontwerp kunnen hiermee in kaart gebracht worden.

In de kwaliteitsmeting van de kernvariabelen in de Gezondheidsenquête in 2010 (Banning et al, 2011) worden verschillende kwaliteitsdimensies berekend voor de Gezondheidsenquête 2010 in het oude design en in het nieuwe design. Doordat de designs parallel aan elkaar uitgevoerd zijn kan de kwaliteit van de verzamelde data in beide designs goed vergeleken worden. In de kwaliteitsmeting beperken we ons

<sup>1</sup> De revisie van medische consumptie was al gepland en is daarmee geen onderdeel van het HPE.

tot de maanden waarin zowel het oude als nieuwe design zijn gebruikt<sup>2</sup>, zie Tabel 2.1.

Binnen de kwaliteitsmeting is gekeken naar de volgende kernvariabelen (van Hertem, 2009): Ervaren gezondheid, Aanwezigheid van langdurige ziekten, aandoeningen of handicaps, Contact met de huisarts afgelopen 14 dagen<sup>3</sup>, Functionele beperkingen, Mentale gezondheidsfactor, Gelukkig zijn, Rookgedrag, Overgewicht, en Marihuana-gebruik.

In de volgende paragrafen bespreken we alleen de kwaliteitsdimensies waarop het herontwerp invloed heeft gehad.

## 2.1 Nauwkeurigheid

*Overdekking:* Het percentage kaderfouten stijgt van 1,6% in het oude naar 2,0% in het nieuwe design. Dit is een marginale stijging.

*Unit non-respons:* Voor unit non-respons wordt gekeken naar twee indicatoren: het responspercentage en de R-indicator. De R-indicator meet de afwijking in samenstelling van respons ten opzichte van een vaste set registerkenmerken. Binnen HPE wordt telkens de set type huishouden, leeftijd, herkomst, stedelijkheid, WOZ en werkzaamheid in loondienst gebruikt. Het responspercentage is in het nieuwe design ongeveer 2% hoger, 60,9% om 63,2%. De R-indicator is lager in het nieuwe design, namelijk 80,0% om 78,9%. Deze daling is echter niet statistisch significant. Tabel 3.2.1 geeft de respons en de representativiteit van het nieuwe design per survey mode weer.

Tabel 3.2.1 Unit respons van de GEZO mixed-mode enquête, uitgesplitst naar mode.

	CAWI	CATI	CAPI
Respons	5.323 (35,6%)	2.732 (45,6%)	1.398 (41,0%)
Non-respons	9.641 (64,4%)	3.256 (54,4%)	2.010 (59,0%)
Totaal	14.964 (100%)	5.988 (100%)	3.408 (100%)
R-indicator	77,6%	82,7%	79,4%

*Item non-respons:* Ook voor item-nonrespons is gekeken naar responspercentage en R-indicator. Echter, item-nonrespons is specifiek voor een survey variabele. De kernvariabelen zijn hier dus apart beschouwd. Tabellen 3.2.2 en 3.2.3 bevatten de percentages en R-indicatoren voor de kernvariabelen. In tabel 3.2.3 is de analyse beperkt tot variabelen die meer dan 1% item non-respons hebben. Behalve Overgewicht zijn dat allemaal vragen die uit het tweede deelonderzoek komen. Uit de cijfers volgt dat in het nieuwe design meer item non-respons voorkomt, met name

---

<sup>2</sup> Hiervoor verschillen alleen de modes, en niet de vragen. De vergelijking maakt dus het verschil in kwaliteit door een andere mode strategie inzichtelijk.

<sup>3</sup> Deze variabele is gereviseerd als onderdeel van het blok vragen over medische consumptie en daarmee geen onderdeel van het HPE.

op vragen die in het tweede deelonderzoek zijn gesteld<sup>4</sup>. Dit heeft echter geen grote invloed gehad op de samenstelling van de item-respons; deze is in de meeste gevallen slechts beperkt afgenomen.

Tabel 3.2.2: Gewogen item respons voor de kernvariabelen.

Kernvariabele	Mixed-mode	Traditioneel
	2010	2010
	Respons	Respons
Ervaren gezondheid	99,8	100
Aanwezigheid langdurige ziekten	99,8	100
Contact met huisarts afgelopen 14 dagen	99,8	100
Functionele beperkingen	54,4	67,6
Mentale gezondheidsindicator	52,5	64,4
Gelukkig zijn	53,7	66,6
Roken	99,8	100
Overgewicht	95,4	97,0
Marihuana-gebruik	52,8	66,6

Tabel 3.2.3: R-indicator voor item non-respons voor de kernvariabelen.

Kernvariabele	Mixed-mode	Traditioneel
	2010	2010
	$I_2$ (%)	$I_2$ (%)
Functionele beperkingen	78,3	80,7
Mentale gezondheidsindicator	78,6	81,9
Gelukkig zijn	78,6	80,1
Overgewicht	96,6	97,0
Marihuana-gebruik	80,1	80,0

*Precisie:* Door het hogere responscijfer is de precisie in het nieuwe design groter voor variabelen in het eerste deelonderzoek. De precisie voor het tweede deelonderzoek is lager door tussentijdse uitval.

*Meting:* Hiervoor is gekeken naar de samenhang tussen variabelen door een regressiemodel op te stellen voor ervaren gezondheid gegeven de andere kernvariabelen. De verklarende kracht van dit model en, belangrijker, de richting van de verbanden is niet aangetast door het nieuwe design. Door de afnemende precisie van sommige kernvariabelen zijn verbanden echter wel minder eenvoudig te vinden.

*Netto impact unit- en item non-respons:* Wanneer zowel unit- als item non-respons beschouwd worden neemt de totale respons af. De R-indicator voor de samenstelling van de respons kan echter zowel groter als kleiner worden. Tabel 3.2.4 geeft een

<sup>4</sup> Hierbij is rekening gehouden met het feit dat sommige vragen niet alle leeftijdsgroepen worden gesteld. De item non-respons wordt vastgesteld ten opzichte van de respondenten op het eerste deelonderzoek. (Unit) non-respons op het tweede deelonderzoek wordt hierdoor dus ook als item non-respons beschouwd.

overzicht van de respons en de representativiteit. Het nieuwe design heeft een lagere respons op sommige kernvariabelen maar levert vergelijkbare representativiteit.

*Tabel 3.2.4: Netto responspercentage en R-indicator voor de kernvariabelen.*

Kernvariabele	Mixed-mode 2010		Traditioneel 2010	
	Respons	R-indicator	Respons	R-indicator
Ervaren gezondheid (alle leeftijden)	63,1	79,0	60,9	80,2
Aanwezigheid langdurige ziekten (alle leeftijden)	63,1	79,0	60,9	80,2
Contact met huisarts afgelopen 14 dagen (alle leeftijden)	63,1	79,0	60,9	80,2
Functionele beperkingen (vanaf 12 jaar)	33,2	80,9	39,8	79,3
Mentale gezondheidsindicator (vanaf 12 jaar)	32,0	80,8	37,9	80,5
Gelukkig zijn (vanaf 12 jaar)	32,8	80,9	39,3	79,5
Roken (vanaf 12 jaar)	60,9	80,1	58,9	82,4
Overgewicht (vanaf 20 jaar)	57,2	80,0	56,6	82,2
Marihuana-gebruik (15 – 64 jaar)	32,2	81,0	38,3	79,5

## 2.2 Tijdigheid

Over 2010 werd pas rond 1 juli 2011 gepubliceerd, om zo voldoende tijd te hebben om de gevolgen van het Herontwerp van de GEZO in kaart te brengen. Het streven is om dit latere rapportagemoment eenmalig te laten zijn. Het nieuwe design heeft dus eenmalig tot een vertraging geleid van een kwartaal.

## 2.3 Vergelijkbaarheid

De indicator voor de vergelijkbaarheid geeft aan hoe lang de huidige waarnemingstrategie van de enquête ongewijzigd in gebruik is. 2010 is het eerste jaar waarbij is overgegaan op een mixed-mode waarnemingstrategie. Er geldt daarom formeel dat de vergelijkbaarheid in jaren gelijk is aan nul jaren. Het is vanzelfsprekend van groter belang te bezien of, en in welke mate, breuken voorkomen. Paragraaf 5 gaat hier uitvoerig op in. Binnen de kwaliteitsmeting is slechts gekeken naar de tijdreeksen voor elk van de kernvariabelen en is beoordeeld of de nieuwe waarneming past in de reeks. Uit de reeksen is geconcludeerd dat alleen hasjgebruik een waarde laat zien die buiten de trend valt. Alle overige reeksen worden plausibel bevonden.

In paragraaf 5 wordt beschreven hoe deze reeksen zijn gebruikt om een schatting te maken van de breuk.

## 2.4 Samenvatting kwaliteitsmeting

De kwaliteitsmeting biedt over de volle breedte genomen een positief beeld op van de overgang van het oude naar het nieuwe design. Het responspercentage is in het nieuwe design iets hoger maar de representativiteit is lager, zij het niet significant.

Daarnaast is er in het nieuwe design meer item non-respons, met name op de vragen die in het tweede deelonderzoek zijn gesteld.

### **3. Voorstudie weging**

Vanwege de overgang van het oude (CAPI) design op het mixed-mode design is onderzocht of de weging zoals deze werd toegepast in het CAPI design aangepast moet worden, zie Buelens (2010).

Een aantal verschillende mogelijkheden voor het aanpassen van de weging zijn onderzocht:

1. Varianten op de weging m.b.t. insluitgewichten en indeling van variabelen;
2. Uitbreiding van het weegmodel met nieuwe weegtermen die samenhangen met telefoonbezit en/of Internet;
3. Opname van een stabiliserende term (vaste verdeling van modes) voor het voorkomen van menging van mode- en selectieeffecten;
4. Afzonderlijk wegen van het tweede deelonderzoek vanwege mogelijk selectieve uitval.

De verschillende mogelijkheden worden hieronder apart besproken. De algemene conclusie is dat weegmodel zoals gebruikt in het verleden, ook voor de nieuwe GE kan worden ingezet.

#### **3.1 Varianten op de weging**

De insluitgewichten worden in het CAPI weegmodel gelijk genomen, hoewel strikt gezien de steekproef wordt uitgedund naar veldwerkregio en per maand. Hiermee wordt rekening gehouden door in het weegmodel deze termen op te nemen. Wanneer de insluitgewichten worden aangepast voor de uitdunning, en dus met ongelijke insluitkansen wordt gewogen, blijkt dit nauwelijks invloed te hebben op de cijfers. In het nieuwe design wordt niet langer uitgedund per regio, dus is correctie hiervoor niet meer van toepassing.

Een andere variant betreft de indeling van de variabele 'etniciteit'. Voor de Gezondheidsenquête wordt hierbij ook onderscheid gemaakt in generaties. Dit is voor andere onderzoeken niet gebruikelijk en is bovendien veranderlijk. Het verschil is echter klein; bovendien als er gepubliceerd wordt naar generatie allochtoniteit wordt aangeraden dit kenmerk onveranderd te laten.

Indien gewenst kan de weging vereenvoudigd worden door bovengenoemde varianten.

#### **3.2 Uitbreiding weegmodel**

Vanwege de sequentiële benadering is het niet zo dat een bepaalde deelpopulatie via een mode wordt benaderd en een andere deelpopulatie via een andere. Daarom kan

niet worden gewogen naar de responsmode. Wel kan het kenmerk telefoonbezit opgenomen worden in de weging. Het is bekend dat niet-telefoonbezitters een ander responsgedrag vertonen dan telefoonbezitters. Het opnemen van telefoonbezit gekruist met leeftijd blijkt echter nauwelijks effect te hebben op de uitkomsten. De overige termen in het weegmodel verklaren het verschil in responsgedrag tussen telefoonbezitters en niet-telefoonbezitters.

Daarnaast is gekeken wat het effect is van het uitbreiden van het weegmodel met de gemiddelde WOZ-waarde en het inkomen. Deze termen hangen mogelijk samen met telefoonbezit en/of beschikking over Internet. Zowel WOZ-waarde als inkomen blijken echter geen significant effect te hebben op de uitkomsten.

### **3.3 Opname stabiliserende term**

Wanneer de samenstelling van de respons over de tijd anders zou zijn met betrekking tot de responsmode, en het gaat daarbij om dezelfde groepen die responderen, dan krijgt men te maken met een verschuiving in meetfouten. Dit kan het geval zijn bij afnemende telefoondekking, toenemende Internetdekking of veranderingen in deelnamebereidheid voor zover die niet selectief zijn. Op basis van één onderzoek kan hierover niets gezegd worden. Wel kan op deze eventuele verschuivingen geanticipeerd worden door te bekijken hoe gevoelig de uitkomsten van het onderzoek zijn voor de samenstelling van modes. Dit is gedaan door een term op te nemen in het weegmodel met betrekking tot de responsmode. Hierbij zijn verschillende verhoudingen tussen de modes bekeken, van de huidige verhouding tot verhoudingen die in de toekomst denkbaar zijn.

Hieruit blijkt dat kleine toevallige verschuivingen in de verdeling van modes vermoedelijk geen of weinig effect hebben op de uitkomsten. Grotere verschuivingen kunnen echter destabiliserend werken voor sommige variabelen.

Er is nog geen nette methodologische oplossing om hiermee om te gaan. Voorlopig wordt geadviseerd om niet te wegen naar responsmode. Wanneer in volgende edities van de Gezondheidsenquête de samenstelling van de respons naar mode erg verschilt, kan alsnog de responsmode als term in het weegmodel worden opgenomen.

In de tussentijd is het aan te bevelen te streven naar een zo stabiel mogelijke verdeling van de respons over de modes. Belangrijk hierbij is het nauwgezet uitvoeren van het veldwerk met steeds evenveel rappels en belpogingen en het volledig uitbaten van de steekproef.

### **3.4 Afzonderlijk wegen van de vervolgmodule**

Voor de analyse van de weging waren de doelvariabelen van het tweede deelonderzoek nog niet beschikbaar. Echter uit de voorstudie is wel gebleken dat wegen naar uitkomsten van de GE belangrijk kan zijn voor het vervolgonderzoek. Indien niet wordt teruggewogen, dan blijft de respons op het vervolgonderzoek selectief met betrekking tot de uitkomsten van de GE. Er is inmiddels een



afzonderlijke weging voor de vervolgmodule ontwikkeld, zie Bruggink en Van Herten (2011).

#### 4. Breukanalyse

Om de effecten van de veranderingen in de Gezondheidsenquête met betrekking tot de mode en de vragenlijst in kaart te kunnen brengen, is een transitietraject opgesteld, zie Roels et al. (2009). Hiermee kunnen de vragenlijst- en mode-effecten afzonderlijk geanalyseerd worden, evenals de totaaleffecten, zie Wong et al. (2011b). In de volgende paragrafen wordt de opzet van het transitietraject beschreven, de vergelijking van de resultaten en mogelijke oorzaken van de gevonden verschillen. De breukanalyse is een gedetailleerde uitwerking van de kwaliteitsdimensie vergelijkbaarheid, zie paragraaf 3.4.

##### 4.1 Opzet van het transitietraject

Naast het totaaleffect zijn ook het effect van de verandering van de vragenlijst en waarnemingsmode afzonderlijk in kaart gebracht. Deze effecten kunnen zichtbaar gemaakt worden met behulp van de volgende behandelingen:

1. Een schatting voor de doelvariabelen op basis van de oude vragenlijst en de oude waarnemingsmethodiek. De beschikbare reeksen, die zijn waargenomen vanaf 1981 tot en met 2009 (jaarcijfers), zijn gemodelleerd met een structureel tijdreeksmodel. Vervolgens is een voorspelling gemaakt voor de doelvariabelen voor 2010, zie Wong et al. (2011).
2. Een schatting voor de doelvariabelen via de nieuwe vragenlijst en de nieuwe waarnemingsmethodiek. Omdat het HPE in januari 2010 is ingevoerd, is voor 2010 een schatting beschikbaar gekomen op basis van de reguliere steekproef. De waargenomen cijfers zijn conform het weegmodel van Bruggink en van Herten (2011) opgehoogd tot een jaarcijfer<sup>5</sup>.
3. Een schatting voor de doelvariabelen waargenomen via de nieuwe vragenlijst onder de oude waarnemingsmethodiek. Via het paralleltraject is voor deze variabelen een schatting beschikbaar gekomen op basis van een steekproef van beperkte omvang<sup>6</sup>. De steekproefaantallen liggen ongeveer 70% lager dan de nieuwe opzet (behandeling 2) die gedurende een periode van 7 maanden in 2010 verzameld zijn. Deze zijn opgehoogd tot een jaarcijfer, zie Roels en van den Brakel (2011).

---

<sup>5</sup> Van een aantal variabelen is geen informatie beschikbaar gekomen: *Aantal contacten met fysiotherapeut*, *Perc langdurige aandoening* en *Perc norm gezond bewegen*.

<sup>6</sup> Bij deze behandeling zijn de variabelen *Aantal contacten met fysiotherapeut* en *Perc norm gezond bewegen* niet gepubliceerd.

Door de schattingen van bovengenoemde behandeling 1 en 2 met elkaar te vergelijken, kan het totaaleffect van alle wijzigingen bekeken worden. Bij vergelijking van behandeling 3 en 2 is de invloed van de mode te bepalen. Het effect van de vragenlijstwijzigingen is te bepalen door het verschil tussen behandeling 1 en 3 te analyseren. Dit wordt beschreven in de volgende paragraaf.

## 4.2 Vergelijking behandelingen

Uit de resultaten blijkt dat bij sommige variabelen significante verschillen gevonden zijn tussen de schattingen bij de verscheidene behandelingen. Dit effect kan worden veroorzaakt door verschillen in de vragenlijst, door verandering van mode, of beide. In Tabel 5.3.1 worden alleen de variabelen met significante verschillen weergegeven. Hierbij is voor een variabele, indien het een meerkeuzevraag is, voor elke behandeling de schatting voor de 1<sup>e</sup> categorie in de tabel weergegeven.

Tabel 5.3.1: Variabelen die effect ondervinden (1<sup>e</sup> categorie)

Vragenblok	Variabelen	Schatting 1 - 2 Totaaleffect	Schatting 3 - 2 Mode-effect	Schatting 1 - 3 Vragenlijsteffect
Contact met de specialist en polikliniek	Aantal contacten met specialist	-0,44 *	-0,3	-0,14
Contact met de specialist en polikliniek	Perc contact met specialist	6,18% *	1,50%	4,68% *
Medicijngebruik	Perc voorgeschreven medicijn	4,58% *	1,00%	3,58% *
Medicijngebruik	Perc niet-voorgeschreven medicijn	4,97% *	6,00% *	-1,03%
Roken	Perc zware rokers	1,09% *	2,10% *	-1,01%
Roken	Perc rokers	0,83%	2,80% *	-1,97%
Overige gezondheidsvoorzieningen	Perc contact met alternatief genezer	0,90%	1,60% *	-0,70%

*De significante breuken worden met een \* aangegeven.*

De significante afwijking bij *Perc voorgeschreven medicijn* en *Perc contact met specialist* is waarschijnlijk te verklaren door de verandering van de vragenlijst, aangezien er geen mode-effect is waargenomen. Het vragenlijsteffect heeft ook doorgewerkt bij het totaaleffect.

De verandering van waarnemingsmode is een mogelijke verklaring voor de significante afwijking bij de variabelen *Perc niet-voorgeschreven medicijn* en *Perc zware rokers*. Dit effect is ook zichtbaar in het totaaleffect.

Bij de variabelen *Perc contact met alternatief genezer* en *Perc rokers* is geen sprake van een significant totaaleffect, maar wel van een mode-effect en een klein (niet significant) vragenlijsteffect. Het vragenlijsteffect is tegengesteld van richting vergeleken met het mode-effect, waardoor het totaaleffect niet meer significant is.

Een significant totaaleffect blijkt enkel te gelden voor de variabele *Aantal contacten met specialist*. De verandering van vragenlijst en waarnemingsmode hebben een klein verschil tot gevolg, die afzonderlijk niet-significant zijn, maar elkaar wel versterken (dezelfde richting).

Bij twee van de vragenblokken zijn geen significante effecten gevonden, maar lijken alle doelvariabelen binnen dat blok wel last te hebben van een mode-effect en/of vragenlijsteffect. Bij het blok vervolgonderzoek (PAPI/websurvey) lijkt er een negatief mode-effect en een positief vragenlijsteffect te zijn. Het blok Contact met

de Tandarts, Mondhygiënist en Orthodontist lijkt een negatief mode-effect te hebben. Dit blok bestaat echter maar uit twee variabelen.

### 4.3 Mogelijke oorzaken

Voor de variabelen, waarbij significante verschillen gevonden zijn, wordt gekeken wat de mogelijke oorzaken zijn. Allereerst worden de significante verschillen besproken, die met een vragenlijsteffect te verklaren zijn (paragraaf 4.3.1). Vervolgens komen de variabelen met een mode-effect aan bod (paragraaf 4.3.2).

#### 4.3.1 Vragenlijst-effecten

Voor de variabelen *Perc voorgeschreven medicijn* en *Perc contact met specialist* is onderzocht wat mogelijke redenen zijn van een significante breuk. *Aantal contacten met specialist* en *Perc contact met alternatief genezer* en *Perc rokers* hebben een klein (niet significant) vragenlijsteffect en een tegengesteld mode-effect. Aangezien de variabele *Perc contact met specialist* nauw samenhangt met variabele *Aantal contacten met specialist*, wordt verondersteld dat de vragenlijsteffecten van *Perc contact met specialist* doorwerkt op *Aantal contacten met specialist*. Voor *Perc rokers* is geen verklaring gevonden. Voor een uitgebreide beschrijving van de verschillen en mogelijke gevolgen zie Wong et al. (2011b). De belangrijkste verschillen zijn:

- De volgorde van de themablokken is aangepast. Dit heeft mogelijk invloed op de interpretatie van een vragenblok en het kan geheugeneffecten beïnvloeden.
- Er is besloten de vragenblokken in de verschillende modes hetzelfde te houden. Daardoor zijn in de nieuwe vragenlijst alle themablokken ingekort, om te voorkomen dat de respondent vroegtijdig afhaakt bij websurvey of CATI.
- Het gebruik van antwoordkaarten/lijsten kan leiden tot geheugeneffecten.
- Een extra toelichting/introductie bij een vraag kan leiden tot andere interpretatie en beantwoording van de vraag. Dit geldt ook voor het aanscherpen van een hoofdvraag of voor het toevoegen van één of meerdere vragen.
- Een andere referentieperiode voor de vraag.
- Een andere routing.

#### 4.3.2 Mode-effect

De verandering van CAPI naar mixed-mode Web, CATI en CAPI heeft geleid tot een mode-effect bij de variabelen *Perc niet-voorgeschreven medicijn*, *Perc zware rokers*, *Perc rokers* en *Perc contact met alternatief genezer*. Bij *Perc niet-voorgeschreven medicijn* en *Perc zware rokers* is het mode-effect terug te zien in het totaal effect. De variabelen *Perc rokers* en *Perc contact met alternatief genezer* laten

een mode-effect zien, maar dat is niet terug te zien in het totaaleffect, aangezien beide variabelen een vragenlijsteffect hebben die tegengesteld van richting is als het mode-effect.

Het mode-effect kan veroorzaakt worden door de aanzienlijk grotere bijdrage van respons via Web dan CATI of CAPI, zie Tabel 5.4.1.

*Tabel 5.4.1: Respons-overzicht behandeling 2 basisvragenlijst (nieuwe vragenlijst, mixed-mode waarnemingsstrategie)*

<i>Mode van behandeling 2</i>	<i>Respons</i>	<i>Percentage</i>
Web	9302	56,9%
CATI	4838	29,6%
CAPI	2201	13,5%
Totaal	16341	100%

Voor een aantal variabelen kan een specifiekere oorzaak aangewezen worden voor het mode-effect:

- Een andere snelheid van het vraaggesprek beïnvloedt het resultaat van *Aantal contacten met specialist*.
- Het tonen van “Weet niet” en “Weiger” als zichtbare toegestane antwoordcategorieën bij websurveys, kan ervoor zorgen dat bij behandeling 2 vaker gekozen wordt voor “Weet niet” en “Weiger” dan bij behandeling 3. Dit was het geval bij *Perc contact met alternatief genezer*.
- Bij *Perc rokers* speelt het selectie-effect mogelijk een belangrijke rol. Verondersteld wordt dat over het algemeen hoog opgeleide mensen minder roken dan laag opgeleide mensen. Respondenten van de websurvey zijn iets vaker hoog opgeleide mensen zijn. Opleiding wordt in het weegmodel van het eerste deelonderzoek niet meegenomen. Dit kan betekenen dat hoog-opgeleiden mogelijk iets oververtegenwoordigd zijn in de schatting.

## **5. Conclusies**

In de evaluatie van het oude en nieuwe design voor de Gezondheidsenquête is gekeken naar een breed aantal aspecten: een kwaliteitsmeting op een groot aantal kwaliteitsdimensies voor oud en nieuw design, de invloed op de wegging van nonrespons, en een gedetailleerde analyse van breuken. Deze brede blik en de beschikbaarheid van een dubbeldraaitraject zijn op zichzelf zeer sterk punten van de evaluatie.

De evaluatie biedt in het algemeen een positief beeld. De scores op de verschillende kwaliteitsdimensies wijken weinig af tussen de designs, de wegging hoeft niet te worden aangepast en de breukanalyse laat een relatief klein aantal breuken zien. De reeksen zijn volgens inhoudelijke experts nog steeds plausibel, met uitzondering van

hasjgebruik, en een regressiemodel voor de samenhang tussen variabelen laat geen grote invloed zien van het herontwerp.

Er zijn echter ook enige aandachtspunten:

- De item non-respons op sommige kernvariabelen is toegenomen door de lage respons op het tweede deelonderzoek. Hoewel niet blijkt dat dit een grote, directe invloed heeft gehad op de samenstelling van respons, zorgt de uitval tussen het eerste en het tweede deelonderzoek voor een afname in precisie.
- Daarnaast heeft een aantal tijdreeksen van kernvariabelen te maken met breuken. Dit vermindert de vergelijkbaarheid voor die variabelen voor enkele jaren. Aangezien met het dubbeldraaitraject geen indicatie gekregen kan worden van de afzonderlijke mode effecten van de modes in het nieuwe ontwerp, is onzeker wat voor invloed schommelingen in respons en dekking zullen hebben op vergelijkbaarheid in de tijd. Voor die variabelen waar een breuk is geconstateerd, kan vooraf gesteld worden dat deze het meest gevoelig zullen zijn voor dergelijke schommelingen. Immers deze variabelen laten nu al een mode effect zien. Echter, ook voor de andere reeksen zijn schommelingen mogelijk als mode effecten voor die variabelen bestaan die nu nog onzichtbaar zijn in de mix van 2010.

## Referenties

- Banning, R., Bruggink, J.-W. en Cobben, F. (2011), Kwaliteitsmeting kernvariabelen Gezondheidsenquête 2010, BPA-nummer DMV-2011-10-25-RBNG-JBRK-FCBN, CBS Den Haag/Heerlen
- Brakel, J.A van den en J. Roels (2010). *Intervention analysis with state-space models to estimate discontinuities due to a survey redesign*. The Annals of Applied Statistics , 2010, Vol.0, No.00, 1-37.
- Bruggink, J.-W. en van Herten, M. (2011). Weging Gezondheidsenquête, vanaf 2010, <http://www.cbs.nl/NR/rdonlyres/F6C15B98-5281-48ED-8968-069BA70C789D/0/2011weginggevanaf2010.pdf>
- Buelens, B. (2010). Voorstudie weging Gezondheidsenquête 2010. BPA-nummer DMH-2010-11-16-bbus. Centraal Bureau voor de Statistiek, Heerlen.
- Cuppen, M. (2007). Programma Herontwerp Persoonsenquêtes. Project Initiatie Document. Versie 1.0 (23 november 2007). BPA-nummer SOO-2007-H211. Centraal Bureau voor de Statistiek, Heerlen.
- Cuppen, M., B. Janssen, M. Vosmer, I. Beckers, J.-W. Bruggink, S. te Riele, M. van Herten, M. Houben en A. van Rozendaal (2009). Integraal Onderzoeksdesign 2010, BPA-nummer SOO-2009-H2. Centraal Bureau voor de Statistiek, Heerlen.

- Projectgroep HPE Kwaliteit en Breuken (2008a), Kwaliteitsmetingen binnen het Herontwerp Persoonsenquêtes, Interne nota, BPA DMV-2008-03-31-BSTN, CBS, Voorburg.
- Projectgroep HPE Kwaliteit en Breuken (2008b), Uitwerking kwaliteitskader HPE, Interne nota, BPA DMV-2008-06-10-BSTN, CBS, Voorburg.
- Roels, J. & J. van den Brakel (2011). *Test op correctie op Gezondheidsenquête cijfers voor het 7 maandse paralleltraject 2010*. Centraal Bureau voor de Statistiek. Versie 1 april 2011.
- Roels, J., J. Van den Brakel en M. Van Herten (2009). Transitietraject Gezondheidsenquête bij het Herontwerp Persoonsenquête, BPA-nummer DMH-2009-12-04-jrls. Centraal Bureau voor de Statistiek, Heerlen.
- Wong, F.Y., J. Roels & J. van den Brakel (2011). *Schattingen voor de Gezondheidsenquête 2010 op basis van voorspellingen met behulp van structurele tijdreeksmodellen*. Centraal Bureau voor de Statistiek. Versie 3.4.
- Wong, F.Y., J. Roels & J. van den Brakel (2011b). *Uitkomsten analyse transitietraject Gezondheidsenquête bij het Herontwerp Persoonsenquête*. Centraal Bureau voor de Statistiek. Versie 1.5.p6.