

Analyse methodebreuken
Gezondheidsenquête
bij het herontwerp
Persoonsenquête

Fong Yee Wong, Joeri Roels en Jan van den Brakel

Statistische Methoden (20113)



Verklaring van tekens

.	=	gegevens ontbreken
*	=	voorlopig cijfer
**	=	nader voorlopig cijfer
x	=	geheim
–	=	nihil
–	=	(indien voorkomend tussen twee getallen) tot en met
o (o,o)	=	het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	=	een cijfer kan op logische gronden niet voorkomen
2010–2011	=	2010 tot en met 2011
2010/2011	=	het gemiddelde over de jaren 2010 tot en met 2011
2010/'11	=	oogstjaar, boekjaar, schooljaar enz., beginnend in 2010 en eindigend in 2011
2008/'09–	=	oogstjaar, boekjaar enz., 2008/'09 tot en met 2010/'11

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

Colofon

Uitgever

Centraal Bureau voor de Statistiek
Henri Faasdreef 312
2492 JP Den Haag

Prepress

Centraal Bureau voor de Statistiek
Grafimedia

Omslag

Teldesign, Rotterdam

Inlichtingen

Tel. (088) 570 70 70
Fax (070) 337 59 94
Via contactformulier:
www.cbs.nl/infoservice

Bestellingen

E-mail: verkoop@cbs.nl
Fax (045) 570 62 68

Internet

www.cbs.nl

ISSN: 1876-0333

© Centraal Bureau voor de Statistiek,
Den Haag/Heerlen, 2011.
Verveelvoudiging is toegestaan,
mits het CBS als bron wordt vermeld.

ANALYSE METHODEBREUKEN GEZONDHEIDSENQUÊTE BIJ HET HERONTWERP PERSOONSENQUÊTE

Samenvatting: Bij de Gezondheidsenquête wordt met ingang van 2010 overgegaan op een nieuwe waarnemingsmethodiek en een nieuwe vragenlijst. Het transitieplan dat ontworpen is om effecten van het herontwerp te kwantificeren, is beschreven in Roels, van den Brakel en Herten (2009). In deze nota worden de resultaten van de analyse gepresenteerd, met mogelijke verklaringen voor de gevonden verschillen tussen schattingen van variabelen van verscheidene behandelingen.

Trefwoorden: HPE, gezondheidsenquête, transitie vergelijking.

1. Inleiding

Het CBS verzamelt al sinds 1974 informatie over gezondheid. Vanaf 1974 vond het LeefSituatie-onderzoek (LSO) plaats. In 1989 is deze doorgegaan als Doorlopend Leefsituatie Onderzoek (DLO). Daarnaast is in 1981 de GezondheidsEnquête van start gegaan, die uiteindelijk in 1997 is opgegaan in het Permanent Onderzoek Leef-Situatie (POLS). DLO hield in 1996 op. Door deze overgangen in het verleden kunnen er methodebreuken zijn ontstaan. Daarmee wordt bedoeld dat de veranderingen in het survey proces een systematisch effect op de uitkomsten hebben, waardoor de continuïteit van opgebouwde reeksen verstoord wordt.

In het kader van het Herontwerp Persoonsenquêtes (HPE) wordt de Gezondheidsenquête in 2010 wederom herzien. De belangrijkste veranderingen in het HPE is de toevoeging van een basisvragenlijst en een verandering van de waarnemingsmethodiek. Verder is in de vragenlijst de vraagstelling over medische consumptie veranderd. Al deze veranderingen in het proces kunnen leiden tot verschillen in de uitkomsten. Op verzoek van SAH wordt onderzocht wat de effecten zijn van de verandering van de vragenlijst, de waarnemingsmethodiek en het gezamenlijk effect.

Voor het bepalen van de effecten worden drie verschillende schattingen voor de onbekende doelvariabelen waargenomen of geconstrueerd: een schatting op basis van een voorspelling waarbij data zijn gebruikt, waargenomen via de oude vragenlijst en de oude waarnemingsmethodiek, een schatting via de nieuwe vragenlijst en de oude waarnemingsmethodiek, en (door middel van het

paralleltraject in 2010) via de nieuwe vragenlijst en de oude waarnemingsmethodiek. Door de bovengenoemde drie schattingen te vergelijken, kunnen vragenlijst-, mode- en totaaleffecten geïdentificeerd worden.

Allereerst wordt een uitgebreide beschrijving gegeven van de Gezondheidsenquête. Er wordt algemene informatie gegeven over de Gezondheidsenquête, maar ook de verschillen tussen POLS (tot en met 2009) en de herziene Gezondheidsenquête (vanaf 2010) worden expliciet aangegeven. Doordat de Gezondheidsenquête is herzien, zijn verschillende effecten opgetreden: vragenlijst-, mode- en totaaleffecten. Er is behoefte aan om inzicht te krijgen hoe groot de afzonderlijke effecten zijn. De effecten kunnen zichtbaar gemaakt worden met het transitieplan dat is beschreven in paragraaf 3. In paragraaf 4 worden de resultaten van het veldwerk en de trajectvergelijkingen gepresenteerd. De mogelijke oorzaken van de verschillen worden beschreven in paragraaf 5, gevolgd door een conclusie en aanbevelingen in paragraaf 6.

2. Beschrijving van de Gezondheidsenquête

Voor de Gezondheidsenquête wordt een steekproef getrokken uit de personen van de Nederlandse bevolking. De steekproef bestaat over het algemeen uit personen van alle leeftijden (minimaal 0 jaar). Bij sommige vragenblokken binnen de enquête wordt nog een aparte leeftijdsgrens gehanteerd, zoals bij het vragenblok Betaald werk OP, waarbij een leeftijdsgrens van minimaal 15 jaar geldt. Tot en met 2009 werden jaarlijks ongeveer 10.000 personen aangeschreven. In 2010 waren dat ongeveer 16.000 personen. De steekproef is gelijkmatig verspreid over de maanden van het jaar. Op basis van de antwoorden van de respondent worden jaarcijfers gemaakt, die inzicht geven in de ontwikkelingen in de gezondheid, medische consumptie, leefstijl en preventief gedrag van de Nederlandse bevolking. Doordat er verschillen zijn tussen de samenstelling van de respons en de totale bevolking, wordt de verzamelde data gecorrigeerd met een weegfactor dat is gebaseerd op de kenmerken geslacht, leeftijd, burgerlijke staat, stedelijkheidsgraad, huishoudgrootte, landsdeel, enquetemaand (in 2010 seizoen) en provincieplus (de provincies plus de vier grote steden). In 2010 is ook herkomstgroepering in de weging opgenomen.

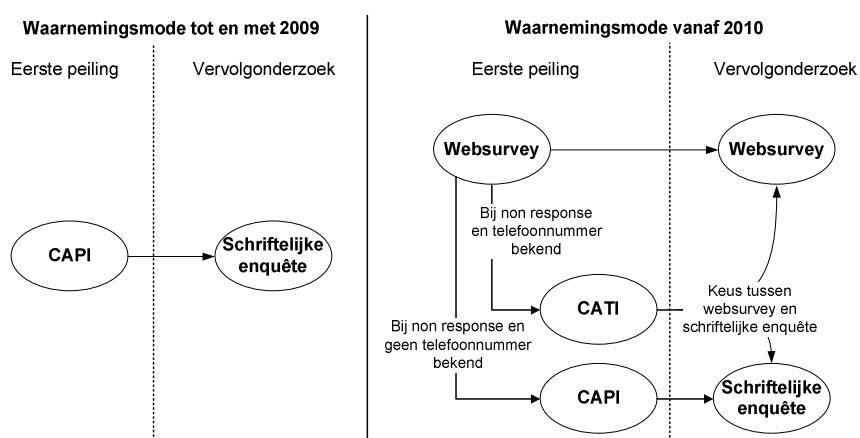
De Gezondheidsenquête bestaat uit twee delen: een eerste peilingsonderzoek en een vervolgonderzoek. Het eerste peilingsonderzoek bevat een basisvragenlijst en een aantal vragenblokken over de gezondheid, zoals Gezondheid en beweging, Gezondheid algemeen en Huisarts. In het vervolgonderzoek worden vragen gesteld, die getypeerd worden als gevoelige onderwerpen. Wegens het gevoelige karakter van de vragen, worden deze vragen schriftelijk door de respondent, zonder tussenkomst van een interviewer, beantwoord. Deze vragenlijst is alleen bestemd voor personen van 12 jaar of ouder.

Voor de breukanalyse van de Gezondheidsenquête zijn 23 doelvariabelen geselecteerd. In sommige gevallen is de uiteindelijk te publiceren variabele gebaseerd op meerdere enquêtevragen. Per vraag zijn er vaak meerdere

antwoordmogelijkheden. In bepaalde gevallen wordt dat vervolgens gedichotomiseerd tot een ja/nee variabele. Vier doelvariabelen zijn continue variabelen. Bij de meeste vragen, wordt vaak alleen de eerste (eventueel gedichotomiseerde) categorie gepubliceerd.

De belangrijkste verschillen tussen het oude ontwerp (tot en met 2009) en het nieuwe ontwerp (vanaf 2010) zijn:

De waarnemingsmode van het eerste peilingsonderzoek bij het oude ontwerp is CAPI en bij het vervolgonderzoek een schriftelijke enquête. In 2010 werd er voor het eerst mixed-mode waargenomen bij de Gezondheidsenquête. De respondenten konden in eerste instantie via een websurvey de vragenlijst invullen. Bij non-respons werd gekeken of van de respondent een telefoonnummer bekend was. Indien dit het geval was, werd de respondent CATI benaderd, en anders CAPI. De mode van het vervolgonderzoek is afhankelijk van de responsmode van het eerste peilingsonderzoek. Indien bij het eerste peilingsonderzoek voor de websurvey was gekozen, kreeg de respondent het vervolgonderzoek ook via web aangeboden. Indien de mode bij het eerste onderzoek CAPI was, werd ook meteen de papieren vragenlijst voor het vervolgonderzoek bij de respondent thuis achtergelaten. CATI-respondenten kregen de keus tussen websurvey of schriftelijke enquête (zie Figuur 1).



Figuur 1: Waarnemingsmode tot en met 2009 en vanaf 2010

- Doordat vanaf 2010 mixed-mode wordt waargenomen en ervoor gekozen is om de vragenlijst bij elke mode op dezelfde manier af te nemen, is de vragenlijst ingekort. De doorlooptijd van het antwoordproces dient bij een websurvey en CATI korter te zijn dan bij CAPI.
- Verder is de tekst van een aantal vragen in 2010 aangepast. Dit heeft onder andere te maken met dat bij CATI geen antwoordkaart aangeboden kan worden en de vragen voor alle modes hetzelfde moesten blijven. Daarnaast zijn ook vragen aangepast ter verduidelijking van de vraag. Bij een websurvey wordt namelijk geen interviewer betrokken om de vragen te verduidelijken. De vragen over zorggebruik (medische consumptie) zijn ook per 2010 aangepast. Dat is gedaan om de vragen beter aan te laten sluiten op de ontwikkelingen die zich in de laatste jaren in de gezondheidszorg hebben voorgedaan.
- Het vragenblok Alcohol en de vragenblokken over drugsgebruik zijn in 2010 verplaatst van eerste peilingsonderzoek naar vervolgonderzoek. Hierdoor is ook de mode voor de vragenblokken over Alcohol en drugsgebruik veranderd.
- De nieuwe vragenlijst bevat ook vragen over vrijwilligerswerk, dat gebruikt wordt voor het onderzoek Vrijwilligerswerk en Sociale Samenhang. In POLS werden ook vragen over Sociale Samenhang gesteld, alleen was dat minder uitgebreid dan bij de vragenlijst van 2010.

3. Beschrijving transitieplan

Een belangrijk aspect van een enquête is continuïteit. Om deze reden worden enquêtes zo lang mogelijk onveranderd gehouden. Om de efficiëntie en kwaliteit van een onderzoek te verhogen worden soms veranderingen in het proces doorgevoerd. Hierdoor kunnen allerlei factoren beïnvloed worden die een systematisch effect hebben op de te meten parameters. Dit systematische effect wordt ook wel methodebreuk genoemd.

Veranderingen in de Gezondheidsenquête zijn o.a. aanpassingen in de vragenlijst en verandering van de waarnemingsmethodiek (zie ook paragraaf 2). Vanuit SAH is aangegeven dat het wenselijk is beide effecten afzonderlijk in kaart te brengen. Hiervoor is een transitieplan opgesteld, waarin wordt beschreven dat er drie schattingen gemaakt worden (paragraaf 3.1) die onderling met elkaar vergeleken kunnen worden (paragraaf 3.2). Verondersteld wordt dat er geen interactie effecten optreden tussen de verandering in de waarnemingsmethodiek en de vragenlijst. In de volgende paragrafen wordt hier verder op in gegaan.

3.1 Beschrijving van de schattingen

Er zijn drie behandelingen te onderscheiden, waarvoor data beschikbaar zijn gekomen. Dit resulteert in drie verschillende schattingen voor de onbekende doelvariabelen voor 2010:

1. Een schatting voor de doelvariabelen op basis van de oude vragenlijst en de oude waarnemingsmethodiek. De beschikbare reeksen die zijn waargenomen tot en met 2009, 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.
3. Een schatting voor de doelvariabelen waargenomen via de nieuwe vragenlijst onder de oude waarnemingsmethodiek. Via het paralleltraject is voor deze variabele een schatting beschikbaar gekomen op basis van een steekproef van beperkte omvang. 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 zijn tot een jaarcijfer, zie Roels en van den Brakel (2011).

Deze drie schattingen zijn nodig om de verschillende effecten te kunnen analyseren. Hieronder wordt in het kort uitgelegd hoe de drie schattingen tot stand zijn gekomen. De schattingen op basis van een structureel tijdreeksmodel wordt uitgebreider beschreven in Wong, et al (2011) en de schattingen van het paralleltraject in Roels en van den Brakel (2011).

3.1.1 Schattingen op basis van een structureel tijdreeksmodel

De tijdreeksen, die zijn waargenomen tot en met 2009, kunnen worden gebruikt om via een tijdreeksmodel een voorspelling te maken voor de doelvariabelen in 2010. Het model bestaat o.a. uit een stochastisch trendmodel (smooth trend) voor elke antwoordcategorie. Vanwege veranderingen in het survey proces (zie Stam en Knoops, 2009), zijn interventievariabelen mee gemodelleerd, waarbij de bijbehorende regressiecoëfficiënten kunnen worden geïnterpreteerd als methodebreuken (zie ook van den Brakel en Roels, 2010). Met het pakket OxMetrics (gebruikmakend van subroutines uit Ssfpack 3.0) kunnen optimale schattingen voor de modelparameters berekend worden. Zie Koopman et al. (2008), Harvey (1989) en Durbin en Koopman (2001) voor technische details.

Om de validiteit van het gefitte model te analyseren, is o.a. onderzocht of de innovaties standaard normaal verdeeld zijn en of er geen heteroskedasticiteit of seriële correlatie in de innovaties zit. Zie Koopman et al. (2008) voor definitie en berekening. Het blijkt dat het structurele tijdreeksmodel bruikbaar is voor de waarnemingen tot en met 2009. Verder hebben inhoudelijke deskundigen

aangegeven dat er geen aanwijzingen zijn dat er (grote) veranderingen zijn geweest in 2010, waardoor de voorspellingen van het structureel tijdreeksmodel voor 2010 niet veel zullen afwijken van de reële ontwikkelingen in 2010. De voorspellingen zullen gebruikt worden bij de transitie vergelijkingen.

3.1.2 Schattingen nieuwe opzet

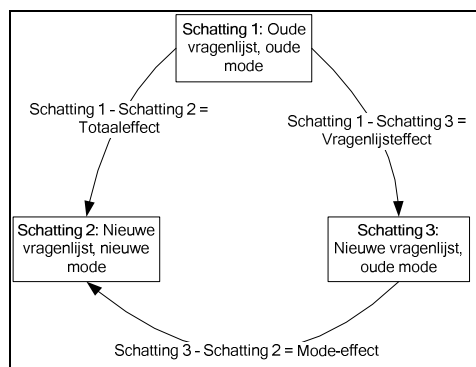
De waargenomen cijfers zijn conform het weegmodel van Bruggink en van Herten (2011) opgehoogd tot een jaarcijfer. Van een aantal variabelen is bij behandeling 2 geen informatie beschikbaar gekomen: *Aantal contacten met fysiotherapeut*, *Perc langdurige aandoening* en *Perc norm gezond bewegen*.

3.1.3 Schattingen paralleltraject

Bij de Gezondheidsenquête worden elk jaar cijfers gepubliceerd die op 12 maanden zijn gebaseerd. Door budgettaire redenen en veldwerkcapaciteitsproblemen is voor de derde behandeling in 2010 slechts 7 maanden data verzameld. Omdat er mogelijk seizoenseffecten in de cijfers optreden, kunnen door het ontbreken van 5 maanden, de cijfers van de drie behandelingen wellicht niet goed vergeleken worden. Om te weten te komen bij welke variabelen een systematisch effect optreedt, worden de cijfers van 2009 gebaseerd op 12 maanden vergeleken met de cijfers van 2009 gebaseerd op 7 maanden. De cijfers worden vergeleken met behulp van een design-based variant van een T-Toets met een Holm-Bonferroni correctie, zie Holm (1979). Hieruit is gebleken dat geen significant seizoenseffect aanwezig is. De cijfers voor behandeling 3 zullen dan ook niet gecorrigeerd worden (zie ook Roels en van den Brakel, 2011). Er zijn geen schattingen beschikbaar gekomen voor de variabelen *Aantal contacten met fysiotherapeut* en *Perc norm gezond bewegen*.

3.2 Vergelijkingen van verschillende behandelingen

Zoals in paragraaf 3.1 beschreven is, dienen drie schattingen berekend te worden om de vragenlijst-, mode- en totaaleffect van de veranderingen in 2010 te bepalen. Doordat verondersteld is dat er geen interactie-effecten optreden tussen de verandering in de waarnemingsmethodiek en de vragenlijst, kan met de trajectvergelijkingen, weergegeven in Figuur 2, de afzonderlijke effecten gemeten worden.



Figuur 2: trajectvergelijkingen voor het meten van vragenlijst-, mode- en totaaleffecten

Door de schattingen van behandeling 1 en 2 met elkaar te vergelijken, kan het totaaleffect van alle wijzigingen bekeken worden. Bij vergelijking van behandeling 2 en 3 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. Variabelen van behandelingen worden vergeleken door middel van een T-Toets waarbij het significantie niveau waartegen getoetst wordt, gecorrigeerd wordt via de methode van Holm-Bonferroni, zie Holm (1979).

Hierbij moet nog opgemerkt worden dat het niet helemaal zuiver is om het verschil tussen schatting 3 en 1 alleen maar vragenlijsteffect te noemen. Bij het ontwerp van de nieuwe vragenlijst zijn namelijk de vragenblokken over Alcohol en drugsgebruik van het eerste peilingsonderzoek naar het vervolgonderzoek verplaatst. Het eerste peilingsonderzoek is bij behandeling 3 met CAPI waargenomen en het vervolgonderzoek met een schriftelijke enquête. Verondersteld wordt dat de modeverandering geen gevolgen heeft voor de meeste variabelen, met uitzondering voor de doelvariabelen van het vragenblok Alcohol zelf. Voor de analyse zijn geen variabelen over drugsgebruik geselecteerd. Dit betekent dat voor de variabelen over Alcohol het afzonderlijke vragenlijst- en mode-effect niet te meten is met bovengenoemd transitieplan.

4. Resultaten

De enquêtes zijn conform de beschrijving van paragraaf 2 en 3 uitgezet bij de respondenten. Op basis van de respons is voor elke behandeling een schatting per variabele bepaald (zie Bijlage A). Het bepalen van de schattingen komt uitgebreid aan bod in Wong, et al. (2011) voor behandeling 1, en in Roels en van den Brakel (2011) voor behandeling 3. Voor het bepalen van de schattingen voor behandeling 2 is het weegmodel van Bruggink en van Hertem (2011) gebruikt. Dit document richt zich voornamelijk op het vergelijken van de verschillende schattingen, waarmee de verschillende effecten bepaald kunnen worden.

Allereerst wordt in paragraaf 4.1 inzicht gegeven in het veldwerkproces en dan met name de respons van de mixed-mode waarneming. Vervolgens worden in paragraaf 4.2 de schattingen van de verschillende behandelingen met elkaar vergeleken, waarbij ook enige aandacht is of er sprake is van een structureel effect (positief of negatief) bij één of meerdere vragenblokken.

4.1 Veldwerkproces

De schattingen van behandeling 1 zijn gebaseerd op de enquêtedata, waargenomen tot en met 2009. De respondenten zijn voor het eerste peilingsonderzoek face-to-face benaderd. Vervolgens heeft de interviewer een schriftelijke vragenlijst achtergelaten voor het vervolgonderzoek. Voor benadering 3 zijn de respondenten op dezelfde manier benaderd, alleen dan wel in 2010.

De waarneemstrategie van behandeling 2 is afwijkend. Om kosten te besparen is in eerste instantie aan alle respondenten gevraagd om via websurvey het eerste peilingsonderzoek in te vullen. Indien de respondent niet reageerde, werd afhankelijk van de beschikbaarheid van een telefoonnummer, de respondent telefonisch of face-to-face benaderd. 56,9% van de totale respons op het eerste peilingsonderzoek heeft de vragenlijst via web ingevuld. Voor CATI is het percentage 29,6% en CAPI 13,5%.

Voor het vervolgonderzoek waren twee modes beschikbaar: websurvey of schriftelijke enquête. De respondenten, die in eerste instantie de vragenlijst op internet hebben gebruikt, kregen voor het vervolgonderzoek ook een websurvey. De CAPI-respondenten kregen net zoals bij de oude waarneemstrategie een papieren vragenlijst aangeboden. De CATI-respondenten konden kiezen tussen websurvey of een schriftelijke enquête. 60,3% van de respondenten van het vervolgonderzoek heeft het eerste peilingsonderzoek via internet ingevuld. Voor 25,8% van de respondenten van het vervolgonderzoek geldt dat ze via CATI benaderd zijn om het eerste peilingsonderzoek te beantwoorden. Bij de rest, 13,9% van de totale respons op het vervolgonderzoek, is een papieren vragenlijst achtergelaten bij het afnemen van het eerste peilingsonderzoek via CAPI. Zie ook Tabel 1.

Tabel 1: Respons-overzicht behandeling 2 (nieuwe vragenlijst, mixed-mode waarnemingsstrategie)

Mode van behandeling 2	Responspercentages	
	Eerste peilingsonderzoek	Vervolgonderzoek
Web	56,9%	60,3%
CATI	29,6%	25,8%
CAPI	13,5%	13,9%
Totaal	100%	100%

Het veldwerk van de Gezondheidsenquête, uitgevoerd door de telefonische interviewers en CAPI-interviewers, is geëvalueerd door de divisie Dataverzameling (Schulte en Mohren, 2010). Uit de evaluatie komt onder andere naar voren dat doordat de telefonische interviewers andere personen zijn dan CAPI-interviewers en de CAPI-interviewers al enige jaren ervaring hebben kunnen opdoen, dat beide type interviewers het afnemen van de Gezondheidsenquête anders hebben ervaren. De CATI-interviewers moesten nog aan de vragenlijst wennen.

Zowel de CATI- als CAPI-interviewers hebben met respondenten te maken, die in eerste instantie hebben geweigerd om via web mee te werken aan de Gezondheidsenquête. Uit de evaluatie blijkt dat iets meer dan 12% van de CAPI-interviewers heeft ervaren dat het moeilijk tot heel moeilijk was om medewerking te verkrijgen. Bij de CATI-interviewers was dat iets minder dan 10%. Omdat verwacht werd dat voor de CATI en CAPI een moeilijkere groep respondenten over zou blijven, doordat ze in eerste instantie al geweigerd hadden voor de websurvey, lijken deze percentages mee te vallen.

Uit de evaluatie van de interviewers blijkt dat respondenten hebben aangegeven dat ze problemen hadden ondervonden bij het invullen van de websurvey. Ze liepen vaak vast in de vragenlijst en de inlogcodes klopten niet altijd.

De overgrote meerderheid van de interviewers heeft volgens de evaluatie aangegeven dat het interview tamelijk tot heel soepel is verlopen. De interviewduur wordt echter wel als (erg) lang ervaren. De werkelijke gemiddelde gespreksduur van het CAPI-interview (35,2 minuten) blijkt wel korter te zijn dan de vooraf ingeschatte tijdsduur van het CAPI-interview (40 minuten), terwijl het CATI-interview gemiddeld iets langer heeft geduurd (20,1 minuten) dan ingeschat (20 minuten).

De opbouw van de vragenlijst werd als onlogisch en rommelig ervaren. Dit kwam vooral doordat de vragen van de basisvragenlijst en de gezondheidsvragen door elkaar gevraagd werden. Ook was de overgang naar een nieuw blok ook niet altijd logisch voor de respondent. Bij de CAPI vragenlijst volgde op het eind nog een vragenblok Sociale Samenhang wat door de respondent ook als onlogisch ervaren werd, aangezien hier geen vragen over gezondheid in zitten. Daarnaast zijn er opmerkingen gemaakt over dat bij een aantal vragen de vraagteksten te lang was. Met name het blok Contact met de Huisartsenpraktijk werd genoemd. Er was onder andere onduidelijkheid over de werkzaamheden en aanwezigheid van de praktijkondersteuner.

4.2 Analyse van de resultaten

Van de drie behandelingen zijn schattingen beschikbaar gekomen. Deze schattingen zijn met elkaar vergeleken. Er 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, zie Tabel 2, aangezien er is aangenomen dat er in 2010 geen (grote) veranderingen zijn geweest, waardoor de voorspelling van het structureel tijdreeksmodel voor 2010 zullen afwijken van de reële ontwikkelingen in 2010. Hierbij is voor een variabele, indien het een meerkeuzevraag is, voor elke behandeling de schatting voor de eerste categorie in de tabel weergegeven, met uitzondering van *Perc matig en ernstig overgewicht*. Daarvan worden de eerste twee categorieën weergegeven in de tabel, omdat die doelvariabele uit drie antwoordcategorieën bestaat.

Uit Tabel 2 blijkt dat bij vijf vragenblokken alle doelvariabelen van dat blok een mode- en/of vragenlijsteffect hebben, die allemaal dezelfde richting op gaan: het vervolgonderzoek (schriftelijke enquête/websurvey) en de vragenblokken Contact met de Tandarts, Mondhygiënist en Orthodontist, Overige gezondheidsvoorzieningen en Roken. De doelvariabelen van het vervolgonderzoek vertonen allemaal een negatief mode-effect en een positief vragenlijsteffect. De effecten zijn over het algemeen vrij klein (niet significant). Het vragenblok Lengte en gewicht bevat slechts één doelvariabele met meerdere antwoordcategorieën, waardoor het niet mogelijk is om over effect binnen blok te spreken.

Er is bij alle doelvariabelen van vragenblok Contact met de Tandarts, Mondhygiënist en Orthodontist een negatief mode-effect geconstateerd. De verschillen zijn niet groot (niet significant).

Tabel 2: Resultaten vergelijkingen tussen behandelingen

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%
Contact met de Tandarts, Mondhygiënist en Orthodontist	Aantal contacten met tandarts	-0,23	-0,2	-0,03
Contact met de Tandarts, Mondhygiënist en Orthodontist	Perc contact met tandarts	0,43%	-1,30%	1,73%
Contacten met de huisartsenpraktijk	Aantal contacten met huisarts	-0,25	-0,2	-0,05
Contacten met de huisartsenpraktijk	Perc contact met huisarts	2,16%	0,60%	1,56%
Stemgedrag en religie (gezondheid en beweging)	Perc goede ervaren gezondheid	1,08%	1,40%	-0,32%
Schriftelijke enquête/websurvey	Perc beperking bewegen	0,38%	-0,10%	0,48%
Schriftelijke enquête/websurvey	Perc beperking horen	0,02%	-0,50%	0,52%
Schriftelijke enquête/websurvey	Perc beperking zien	-0,11%	-0,90%	0,79%
Schriftelijke enquête/websurvey (in 2009 Alcohol)	Perc zware drinkers	0,67%	-0,90%	1,57%
Schriftelijke enquête/websurvey	Perc langdurige aandoening	Vergelijking niet mogelijk		1,27%
Schriftelijke enquête/websurvey	Perc norm gezond bewegen	Vergelijking niet mogelijk		
Lengte en gewicht	Perc matig en ernstig overgewicht (categorie 1)	-1,50%	0,50%	-2,00%
Lengte en gewicht	Perc matig en ernstig overgewicht (categorie 2)	1,11%	1,70%	-0,59%
Medicijngebruik	Perc anticonceptiepil	-0,81%	-1,50%	0,69%
Medicijngebruik	Perc niet-voorgeschreven medicijn	4,97%	6,00%	-1,03%
Medicijngebruik	Perc voorgeschreven medicijn	4,58%	1,00%	3,58%
Overige gezondheidsvoorzieningen	Aantal contacten met fysiotherapeut	Vergelijking niet mogelijk		
Overige gezondheidsvoorzieningen	Perc contact met alternatief genezer	0,90%	1,60%	-0,70%
Overige gezondheidsvoorzieningen	Perc contact met fysiotherapeut	-0,55%	1,80%	-2,35%
Roken	Perc rokers	0,83%	2,80%	-1,97%
Roken	Perc zware rokers	1,09%	2,10%	-1,01%
Ziekenhuisopname en dagopname	Perc ziekenhuisopname	0,42%	-0,70%	1,12%

De significante breuken worden dik gedrukt weergegeven.

Het vragenblok Overige gezondheidsvoorzieningen heeft een negatief vragenlijsteffect en een positief mode-effect. Het mode-effect bij variabele *Perc contact met alternatief genezer* is bovendien nog significant. Mogelijke verklaringen worden gegeven in paragraaf 5. Het gebruik van antwoordkaarten (oude vragenlijst wel, nieuwe vragenlijst niet) is één van de verschillen, die bij de vergelijking tussen de vragenlijsten is gevonden.

Bij de doelvariabelen van het vragenblok Roken is juist het tegenovergestelde zichtbaar: een positief mode-effect en een negatief vragenlijsteffect. Het gaat hier slechts om twee variabelen. Aangezien bij de doelvariabelen van Roken ook een significante afwijking is waargenomen, zal later in dit document de variabele nog nader bekeken worden.

Bij zeven variabelen is bij de transitievergelijkingen een significant verschil gevonden. De significante afwijking bij *Perc voorgeschreven medicijn* en *Perc*

contact met specialist is waarschijnlijk te verklaren door de verandering van de vragenlijst. Bij de vergelijking van behandeling 1 met behandeling 3 is bij deze twee variabelen een significante afwijking gevonden en niet bij de vergelijking tussen behandeling 2 en behandeling 3. De significante afwijking werkt door bij het totaaleffect (behandeling 1 en 2).

De effecten bij de variabelen *Perc niet-voorgeschreven medicijn* en *Perc zware rokers* kunnen waarschijnlijk verklaard worden door het effect van de verandering van waarnemingsmode (vergelijking behandeling 2 en 3). Deze variabelen hebben ook een significante afwijking bij de vergelijking van het totaaleffect (behandeling 1 en 2). De significante afwijking van het totaaleffect zal waarschijnlijk door de waarnemingsmodeverandering verklaard worden.

Bij de variabelen *Perc contact met alternatief genezer* en *Perc rokers* is er geen sprake van een significant totaaleffect. Deze variabelen hebben naast de waarnemingsmodeverandering ook te maken hebben met een klein niet significant vragenlijsteffect. Het vragenlijsteffect is tegengesteld van richting vergeleken met het effect van de waarnemingsmode, waardoor het totaaleffect niet meer significant is.

Een significant verschil tussen alleen behandeling 1 en behandeling 2 (totaaleffect) blijkt enkel te gelden voor de variabele *Aantal contacten met specialist*. De verandering van vragenlijst (behandeling 1 en 3) en de verandering van waarnemingsmode (behandeling 2 en 3) hebben een klein verschil tot gevolg, die afzonderlijk niet-significant zijn, maar elkaar wel versterken (dezelfde richting).

5. 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 5.1). Vervolgens komen de variabelen met een mode-effect aan bod (paragraaf 5.2) en variabelen waar alleen een totaaleffect speelt (paragraaf 5.3).

5.1 Vragenlijst

Door de Gezondheidsenquête van 2010 met die van 2009 te vergelijken, kunnen mogelijke verklaringen voor het vragenlijsteffect gevonden worden. De belangrijkste verschillen, die zijn geconstateerd bij de vergelijking van de vragenlijsten, en de mogelijke gevolgen zijn:

- De volgorde van de themablokken zijn aangepast. De interpretatie van een vragenblok kan mede bepaald worden door de voorgaande vragenblokken. Tevens kan dit geheugeneffecten beïnvloeden.
- In de nieuwe vragenlijst zijn de themablokken ingekort. Veel vragen uit de oude vragenlijst zijn weggevallen. De meeste vragenblokken bevatten

weinig vragen. Doordat de overgang van vragenblok naar vragenblok bij de nieuwe vragenlijst vrij snel gaat, kan dat als rommelig en onlogisch ervaren worden. Zie ook Schulte en Mohren (2010). Het kan daardoor moeilijker zijn om de aandacht erbij te houden.

- Er zijn ook veranderingen binnen blokken waargenomen, zoals de formulering van de vraag. Dit zou onder andere kunnen leiden tot andere interpretatie van de vraag door de respondent. Twee andere voorbeelden van veranderingen binnen blokken zijn verschillen gevonden in de routing en het gebruik van antwoordkaarten:
 - Verschillen in de routing: bij een weiger-antwoord op een vraag wordt bij de nieuwe vragenlijst vaak nog een vervolgvraag gesteld, zoals bij de variabele *Perc contact met specialist* het geval is. Bij de oude vragenlijst is dat niet het geval. Dit betekent dat bij de nieuwe vragenlijst geprobeerd wordt om een antwoord te krijgen van een weigeraar. (Jansen en Joostens, 1998)
 - Er worden geen antwoordkaarten gebruikt bij de nieuwe vragenlijst en wel bij de oude vragenlijst. Sommige van deze vragen zijn geschrapt in de nieuwe vragenlijst. Andere vragen zijn opgesplitst in meerdere vragen. Dit kan tot de volgende gevolgen leiden:
 - Bij twijfel over een antwoord, kan een antwoordkaart gebruikt worden om het geheugen van de respondent te prikkelen. De respondent kan zo erachter komen dat een antwoord toch van toepassing is op hem of haar.
 - Bij het opsplitsen van een vraag met een antwoordkaart in meerdere vragen wordt nadrukkelijk bij elke antwoordmogelijkheid stilgestaan. Bij een antwoordkaart bepaalt de respondent zelf hoeveel aandacht er aan elke antwoordcategorie geschonken wordt.

Omdat bij de variabelen *Perc voorgeschreven medicijn* en *Perc contact met specialist* significante verschillen zijn geconstateerd, zullen ze in de volgende subparagrafen nog nader bekeken worden. Er zal dan alleen ingegaan worden op de veranderingen binnen blokken.

5.1.1 *Perc voorgeschreven medicijn*

De variabele *Perc voorgeschreven medicijn* valt binnen het blok Medicijngebruik. Volgens de voorspelling van behandeling 1 zou 40,4% van de populatie gebruik maken van voorgeschreven medicijnen en 59,6% niet (met een standaardfout van 0,4). Uit het paralleltraject blijkt dat 36,8% wel voorgeschreven medicijnen gebruikt en 63,2% niet (met een standaardfout van 0,8). Zie Tabel 3.

Tabel 3: Resultaten *Perc voorgeschreven medicijn*

Variabele	Wel		Niet	
Perc voorgeschreven medicijn (behandeling 1)	40,4%	(0,4)	59,6%	(0,4)
Perc voorgeschreven medicijn (behandeling 3)	36,8%	(0,8)	63,2%	(0,8)

De standaardfouten zijn tussen haakjes achter de schattingen vermeld.

De overgang naar de nieuwe vragenlijst heeft geleid tot een lager percentage bij Perc voorgeschreven medicijn. De volgende veranderingen in de nieuwe vragenlijst kunnen effect hebben op deze variabele:

- De nieuwe vraag over het gebruik van voorgeschreven medicijn bevat een toelichting, die ontbreekt bij de oude vraag: “Sommige medicijnen of voedingssupplementen zijn alleen verkrijgbaar wanneer de huisarts of specialist hiervoor een recept uitschrijft.” Doordat deze toelichting niet bij de oude vragenlijst is gegeven, is het de vraag of respondenten beide vragen op dezelfde manier zullen interpreteren en beantwoorden. Er zijn ook medicijnen, die voorgeschreven kunnen worden door de huisarts of specialist en die ook gewoon zonder recept te krijgen zijn. Deze vorm van medicijngebruik wordt mogelijk niet meegenomen bij de nieuwe vragenlijst.
- De interviewer beschikte bij de oude vragenlijst over een antwoordkaart bij de vraag over welk voorgeschreven medicijn gebruikt is door de onderzoekspersoon. De antwoordkaart kan beschouwd worden als een soort checklist. De respondenten kunnen door de antwoordkaart mogelijk erachter komen dat er toch voorgeschreven medicijnen gebruikt worden, terwijl ze er anders niet aan gedacht zouden hebben (Jansen en Joostens, 1998). Volgens het vraagschema van Geers (2009) zou dit niet mogen voorkomen, omdat de vraag met de antwoordkaart alleen gesteld mag worden als de vorige vraag (wel of geen gebruik voorgeschreven medicijn) met “ja” beantwoord wordt. Als echter bij twijfel bij de respondent de interviewer toch de antwoordkaart ter inzage aan de respondent geeft, kan dit leiden tot een hoger medicijngebruik in de oude vragenlijst. De vraag over welke voorgeschreven medicijnen gebruikt worden ontbreekt volledig in de nieuwe vragenlijst.
- Voedingssupplementen worden specifiek genoemd in de nieuwe vragenlijst. Bij de oude vragenlijst ontbreekt het woord voedingssupplementen. Hierdoor kan het mogelijk zijn dat die respondenten voedingssupplementen niet hebben meegenomen in hun beoordeling in de oude vragenlijst. Bij deze aanpassing zou men echter verwachten dat in de nieuwe vragenlijst het medicijngebruik toeneemt. Het kan echter zijn dat bovengenoemde twee effecten sterker zijn. Dit kan dus betekenen dat als voedingssupplementen in de nieuwe vragenlijst niet specifiek genoemd zou zijn, het vragenlijsteffect mogelijk groter zou kunnen zijn.

5.1.2 Perc contact met specialist

De waarden van de variabelen *Perc contact met specialist* is gebaseerd op de antwoorden van de respondenten in het themablok Specialist (oude vragenlijst) en Contact met Specialist en Polikliniek (nieuwe vragenlijst). Volgens de resultaten van behandeling 1 zou 44,0% van de populatie contact hebben gehad met een specialist

en 56,0% niet, met een standaardfout van 0,7. Uit het paralleltraject komen waarden van 39,3% voor wel contact en 60,7% voor geen contact met een specialist, met een standaardfout van 0,8. Zie Tabel 4. Het percentage wel contact met de nieuwe vragenlijst ligt dus hoger dan met de oude vragenlijst.

Tabel 4: Resultaten Perc contact met specialist

Variabele	Wel contact		Geen contact	
Perc contact met specialist (behandeling 1)	44,0%	(0,7)	56,0%	(0,7)
Perc contact met specialist (behandeling 3)	39,3%	(0,8)	60,7%	(0,8)

De standaardfouten zijn tussen haakjes achter de schattingen vermeld.

De volgende veranderingen in de nieuwe vragenlijst kunnen effect hebben op deze variabele:

- De introducties van de genoemde themablokken vertonen enkele verschillen. In de introductie van de nieuwe vragenlijst wordt nadrukkelijk vermeld dat het over de contacten met specialisten in Nederland gaat. De afbakening naar Nederland wordt niet in de introductie van de oude vragenlijst vermeld, maar wel op het einde. De vraagvolgorde kan volgens Dilman en Christian (2005) invloed hebben op het antwoord.
- Bij de oude vragenlijst wordt eerst gevraagd hoeveel contacten er met een specialist waren (maximaal 97) en op het einde hoeveel contacten met een specialist in het buitenland erbij zaten (maximaal 6). Als één van de onderzoekspersonen meer dan 6 contacten in het buitenland zou hebben, dan zouden de twee vragenlijsten andere informatie opleveren. Verwacht wordt dat de kans wel klein is dat dit voorkomt.
- In het vraagschema van de nieuwe vragenlijst (Cremers en Schulte, 2010) staat nadrukkelijk in de introductie vermeld dat contacten met specialisten in het ziekenhuis (niet tijdens een ziekenhuis- of dagopname) wel gezien wordt als contact met specialist en contacten met een psychiater niet. De introductie van de oude vragenlijst is veel beperkter in de toelichting. Het is hierdoor mogelijk dat een respondent anders zou antwoorden in de ene vragenlijst dan in de andere vragenlijst (Giesen, et al., 2010).
- De referentieperiode van de nieuwe vragenlijst is korter dan die van de oude vragenlijst, namelijk 4 weken in plaats van 2 maanden. Het is mogelijk dat hier geheugeneffecten een rol kunnen spelen, maar dat hoeft niet noodzakelijk zo te zijn (Giesen, et al., 2010).
- De routing is een aspect dat mogelijk mee kan spelen in het verschil in vragenlijst. Op de vraag hoeveel contact er in een bepaalde referentieperiode met een specialist is geweest zijn verschillende antwoorden mogelijk: een reëel getal tussen de 0 en 97, “Weet niet” en “Weiger”. De antwoordcategorieën zijn in beide vragenlijsten gelijk aan elkaar. Volgens het schema van de oude vragenlijst (Geers, 2009) worden er verder geen vragen meer gesteld in het themablok als de respondent weigert om een antwoord te geven. Volgens het vraagschema van Cremers en Schulte (2010) krijgen de weigeraars bij de nieuwe vragenlijst nog een aanvullende vraag,

namelijk “Heeft u in de afgelopen 12 maanden voor uzelf contact gehad met de specialist?”.

- Bij de oude vragenlijst wordt meer doorgevraagd over de contacten met een specialist dan bij de nieuwe vragenlijst. Dit hoeft niet noodzakelijkerwijs te leiden tot verschil in antwoorden.

5.2 Mode-effect

De volgende variabelen hebben te maken gehad met het zogenoemde mode-effect (zie ook Tabel 5):

- Perc niet-voorgescreven medicijn.
- Perc zware rokers.
- Perc rokers.
- Perc contact met alternatief genezer.

De variabelen *Perc niet-voorgescreven medicijn* en *Perc zware rokers* hebben een mode-effect wat terug te zien is 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.

Tabel 5: resultaten variabelen met mode-effect

	Behandeling 1				Behandeling 2				Behandeling 3			
	Cat. 1		Cat. 2		Cat. 1		Cat. 2		Cat. 1		Cat. 2	
Perc niet-voorgescreven medicijn	39,2%	(0,7)	60,8%	(0,7)	34,2%	(0,4)	65,8%	(0,4)	40,2%	(0,8)	59,8%	(0,8)
Perc zware rokers	6,3%	(0,3)	93,7%	(0,3)	5,2%	(0,2)	94,8%	(0,2)	7,3%	(0,5)	92,7%	(0,5)
Perc rokers	27,6%	(0,6)	72,4%	(0,6)	26,8%	(0,4)	73,2%	(0,4)	29,6%	(0,8)	70,4%	(0,8)
Perc contact met alternatief genezer	6,9%	(0,3)	93,1%	(0,3)	6,0%	(0,2)	94,0%	(0,2)	7,6%	(0,4)	92,4%	(0,4)

De standaardfouten zijn tussen haakjes achter de schattingen vermeld.

Gekeken naar de verschillende responspercentages per mode in mixed-mode waarneming (Tabel 1), kan opgemerkt worden dat er aanzienlijk veel meer via webrespons is dan CATI of CAPI. Op basis van deze constatering kan een aantal mogelijke verklaringen gegeven worden voor het mode-effect:

- Selectie-effect: verschillende type respondenten kunnen een voorkeur hebben voor verschillende modes (Vannieuwenhuyze, 2010). Maar ook niet elke respondent kan even goed bereikt worden met elke mode. Websurveys zijn alleen mogelijk indien de respondent Internet heeft. Bepaalde deelpopulaties hebben vaker geen Internet dan andere deelpopulaties, zoals ouderen, laag-opgeleiden en mensen met een laag inkomen. Daarnaast spelen ook andere factoren een rol mee of met een bepaalde mode gerespondeerd wordt, zoals de vertrouwdheid om met een bepaalde media om te gaan, zoals computerervaring bij websurveys (De Leeuw, 2005; Feskens, et al., 2010; Bowling, 2005; Beukenhorst & Kerssemakers, 2010).

Sommige mode-effecten zullen meer zichtbaar zijn bij bepaalde type respondenten dan bij andere respondenten.

- Interviewer-effect
 - Beïnvloeden vraaggesprek: de aanwezigheid van een interviewer heeft invloed op het responsproces. Op het moment dat een bepaalde vraag niet duidelijk is voor een respondent of dat de respondent de vraag niet goed interpreteert, kan de interviewer (zowel CAPI, als CATI) de respondent helpen door een extra toelichting te geven. Aan de andere kant kan een interviewer sturend vragen door de respondent een antwoord te suggereren (Beukenhorst & Kerssemakers, 2010; Giesen, et al., 2010).
 - Communicatie: een interviewer heeft op het moment dat er een face-to-face interview gehouden wordt, meer mogelijkheden om te communiceren dan telefonisch of bij een websurvey het geval is. Naast mondelinge communicatie zijn ook andere communicatiemogelijkheden mogelijk, zoals het aanreiken van een antwoordkaart aan de respondent, het gebruik van een vertaalde schriftelijke vragenlijst voor respondenten die de Nederlandse taal niet machtig zijn en non verbale communicatie. Bij CATI is alleen mondelinge communicatie mogelijk en in beperkte mate non-verbale communicatie, zoals kuchen. Bij websurveys gaat het voornamelijk om visuele communicatie.(De Leeuw, 2005; Feskens, et al., 2010; Beukenhorst & Kerssemakers, 2005).
 - Sociaal wenselijke antwoorden: de aanwezigheid van een interviewer kan tot sociaal gewenste antwoorden leiden (Giesen, et al., 2010; Bowling, 2005; Beukenhorst & Kerssemakers, 2005). Gevoelige vragen kunnen het beste visueel, zonder aanwezigheid van een interviewer beantwoord worden (web of schriftelijke enquête), zodat de antwoorden zo min mogelijk worden beïnvloed. Het daadwerkelijk wel (CAPI) of niet (CATI) zien van een interviewer kan ook een effect hebben op het antwoord van de respondent.
 - Stimulans: de interviewer kan de stimulerende factor voor de respondent zijn om mee te doen aan het onderzoek en ook om alle vragen te beantwoorden. Daarnaast kan de interviewer ervoor zorgen dat de vragen “serieuzer” ingevuld wordt (Beukenhorst & Kerssemakers, 2005).
 - Uit de evaluatie van de interviewers (Schulte en Mohren, 2010) blijkt dat de telefonische interviewers het veldwerk anders hebben ervaren dan de CAPI-interviewers. Dit wordt voornamelijk veroorzaakt doordat de vragenlijst voor telefonische interviewers nieuw is, terwijl CAPI-interviewers vaak al jaren de Gezondheidsenquête hebben afgenomen bij de respondent.

- Bij satisficing wordt het vraag-antwoordproces suboptimaal doorlopen (Beukenhorst & Kerssemakers, 2005). De respondent heeft geen tijd of zin om het antwoord goed te begrijpen en te beantwoorden. Dit treed eerder op als er geen interviewer bij aanwezig is. Er wordt dan bijvoorbeeld eerder voor Weet niet of weiger-antwoorden gekozen. “Weet niet” en “Weiger” worden bij CATI en CAPI verborgen gehouden voor de respondenten. Bij een websurvey is dat niet gebeurd en zijn die antwoordcategorieën vaker gebruikt bij de variabele *Perc contact met alternatief genezer*.
- Tijd: drie tijdsaspecten spelen een belangrijke rol bij de verschillende modes.
 - Het moment dat de vragenlijst wordt afgenomen. Bij CATI en CAPI wordt dat bepaald door de aanwezigheid van de interviewer en respondent. De respondenten kunnen bij een websurvey zelf bepalen wanneer ze daar tijd voor vrij maken. Volgens De Leeuw (2005) zijn de resultaten bij open vragen, waarbij de respondent de meeste controle heeft en zelf kan bepalen wanneer de enquête ingevuld wordt, het meest betrouwbaar en het meest consistent, vergeleken met interviewgesprekken. De verschillen zijn echter maar klein.
 - Doorlooptijd interview: de maximale doorlooptijd van een websurvey moet veel korter zijn dan die bij CATI en CAPI. CATI moet ook korter zijn dan CAPI (Bowling, 2005). De fysieke aanwezigheid van een interviewer kan ervoor zorgen dat de respondent langer gemotiveerd blijft om een vragenlijst serieus in te vullen.
 - De snelheid waarmee door de vragen gelopen wordt, is bij elke mode anders. Bij een websurvey heeft de respondent het zelf in de hand (Beukenhorst & Schouten, 2009; Ariel, et al., 2005). Als bij telefonische interviews lange stiltes vallen, kan dat storend werken. Bij face-to-face interviews speelt de interviewer daar een belangrijke rol in. De snelheid waarmee door de vragen gelopen wordt, bepaalt in bepaalde mate ook hoe lang over een vraag nagedacht wordt.
- Lay-out: bij visuele vragenlijsten speelt de lay-out een belangrijke rol bij het antwoordproces van de respondent. Verschillende lay-outs kunnen leiden tot verschillende antwoorden (Beukenhorst & Kerssemakers, 2005; Toepoel, et al., 2009)

Primacy and recency effecten kunnen spelen op het moment dat er een lijst met antwoordcategorieën gegeven wordt. Er wordt onderscheid gemaakt tussen visueel de antwoordcategorieën presenteren of de respondent de antwoordcategorieën laten horen. Indien de antwoordcategorieën getoond worden, wordt vaak voor één van de eerste antwoordcategorieën gekozen, dat enigszins klopt (primacy effect). Hoort de respondent de antwoordmogelijkheden, dan heeft hij of zij eerder de neiging om één van de laatst vernomen opties te kiezen (recency effect), zie Beukenhorst en

Schouten (2009). Aangezien de meeste vragen van de Gezondheidsenquête uit twee antwoordcategorieën (ja/nee) bestaat, wordt aangenomen dat primacy en recency effecten niet of minimaal zullen optreden.

De meeste genoemde effecten kunnen in een bepaalde mate invloed hebben op alle variabelen. Daarnaast kan voor een aantal variabelen nog een specifieke oorzaak aangewezen worden:

- Wie de snelheid van het vraaggesprek bepaalt, kan het resultaat van *Aantal contacten met specialist* beïnvloeden. Het is mogelijk dat de respondent het aantal contactmomenten moet uitzoeken om de vraag te kunnen beantwoorden en kan mogelijk meer geneigd zijn om de zoekactie uit te voeren als hij zelf mag bepalen wanneer de zoekactie zal plaatsvinden, dan wanneer een ander voor hem het moment bepaalt (wanneer het interview afgenomen wordt). Zo kan de respondent mogelijk op meer contactmomenten uitkomen, dan als hij/zij het bijvoorbeeld niet had opgezocht in de agenda.
- 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*. Dit leidt tot verschillen, indien deze antwoorden niet willekeurig uit de categorieën komen. Voor de variabelen *Perc niet-voorgeschreven medicijn* en *Perc rokers* geldt dat niet. *Perc zware rokers* wordt onder andere afgeleid van de vervolgvraag op *Perc rokers*. Er is dus een afhankelijkheid tussen *Perc rokers* en *Perc zware rokers*. Bij de vervolgvraag *Perc zware rokers* zijn de antwoordcategorieën “Weet niet” en “Weiger” niet toegestaan. Bij *Perc contact met alternatief genezer* heeft ongeveer 1% van de alle respondenten voor “Weet niet” of “Weiger” gekozen. Die 1% is allemaal afkomstig van de websurvey. Bij de andere variabelen is het percentage dat gekozen heeft voor “Weet niet” of “Weiger” (nagenoeg) gelijk aan nul.
- Bij *Perc rokers* zal selectie-effect mogelijk een belangrijke rol spelen. Verondersteld wordt dat over het algemeen hoog opgeleide mensen minder roken dan laag opgeleide mensen. Verder hebben inhoudelijke deskundigen bij SAH aangegeven dat de websurvey-respondenten iets vaker hoog opgeleide mensen zijn. Opleiding wordt in het weegmodel van het eerste peilingsonderzoek niet meegenomen. Dit kan betekenen dat hoog-opgeleiden mogelijk iets oververtegenwoordigd wordt in de schatting.

De variabelen *Perc rokers* en *Perc contact met alternatief genezer* laten afgezien van een significant mode-effect ook een klein niet significant tegengesteld vragenlijsteffect zien. Behalve dat bovengenoemde mogelijke oorzaken voor het mode-effect ook toepasbaar zijn op de twee genoemde variabelen, wordt in onderstaande twee subparagrafen ingegaan op deze twee variabelen.

5.2.1 Perc rokers

Perc rokers komt uit het themablok Roken OP. Er is geen significant verschil gevonden tussen de resultaten van behandeling 1 en behandeling 3 en bij de vergelijking van het totaaleffect, maar bij de vergelijking tussen behandeling 2 en 3 is wel een significant verschil gevonden (mode-effect). De percentages, die bij deze variabele horen, staan weergegeven in Tabel 6. De verandering van mode zorgt voor een lager cijfer. De verandering van vragenlijst zorgt voor een hoger cijfer. De effecten samen zorgen ervoor dat het significante verschil niet zichtbaar is in het totaal effect, terwijl dit wel het geval is tussen de verschillende modes. Het vragenlijst effect is minimaal, maar niet significant.

Tabel 6: Resultaten Perc rokers

Variabele	Rokers		Niet rokers	
Perc rokers (behandeling 1)	27,6%	(0,6)	72,4%	(0,6)
Perc rokers (behandeling 2)	26,8%	(0,4)	73,2%	(0,4)
Perc rokers (behandeling 3)	29,6%	(0,8)	70,4%	(0,8)

De standaardfouten zijn tussen haakjes achter de schattingen vermeld.

De vragen in de oude vragenlijst en nieuwe vragenlijst zijn bijna identiek aan elkaar. Er zijn kleine verschillen geconstateerd. Die verschillen zijn echter geen verklaring voor het kleine vragenlijst effect. Wat mogelijk wel een rol kan meespelen, is de volgorde van de verschillende themablokken.

5.2.2 Perc contact met alternatief genezer

In de themablok Overige gezondheidsvoorzieningen is een aantal vragen opgenomen over *Perc contact met alternatief genezer*. Volgens Tabel 7 veroorzaakt de nieuwe vragenlijst een hogere waarde voor Perc contact met alternatief genezer, terwijl de mixed-mode waarneemstrategie een lagere waarde tot gevolg heeft.

Tabel 7: Resultaten Perc contact met alternatief genezer

Variabele	Wel contact		Geen contact	
Perc contact met alternatief genezer (behandeling 1)	6,9%	(0,3)	93,1%	(0,3)
Perc contact met alternatief genezer (behandeling 2)	6,0%	(0,2)	94,0%	(0,2)
Perc contact met alternatief genezer (behandeling 3)	7,6%	(0,4)	92,4%	(0,4)

De standaardfouten zijn tussen haakjes achter de schattingen vermeld.

De belangrijkste verandering in de vragenlijst is dat bij de nieuwe vragenlijst ook een vraag gesteld wordt of de huisarts ook alternatieve geneeswijzen toepast. Bij de oude vragenlijst wordt de huisarts in de themablok “Overige gezondheidsvoorzieningen” niet meegeteld als bijzondere genezer. De lijst van type alternatieve geneeswijze/genezers in de nieuwe vragenlijst is langer dan die van de oude vragenlijst: antroposofie, chiropractie, osteopathie, fythotherapie of kruidengeneeswijze ontbraken in de oude lijst. Deze typen van geneeswijzen vallen onder categorie “Andere genezers” bij de oude vragenlijst. Een dergelijke checklist heeft invloed op de respondenten. De respondenten zouden sommige antwoorden eerder selecteren, omdat ze op een antwoordmogelijkheid geattendeerd worden, terwijl ze er anders niet aan gedacht zouden hebben (Jansen en Joostens, 1998).

5.3 Totaaleffect

Bij de variabele *Aantal contacten met specialist* is een totaaleffect gevonden, maar geen afzonderlijk vragenlijsteffect of mode-effect. Het gemiddeld *Aantal contacten met specialist* is volgens de voorspelling 1,9 met een standaardfout van 0,1. Bij het paralleltraject is dat 2,0, met eveneens een standaardfout van 0,1. Zie Tabel 8. Het *Aantal contacten met een specialist* is hoger bij de nieuwe vragenlijst, dan bij de oude vragenlijst. Bij de nieuwe mode is het *Aantal contacten met een specialist* ook hoger. Beide effecten zijn niet significant, maar samen zorgen ze wel voor een significant hoger totaaleffect. Aangezien de variabele *Perc contact met specialist* nauw samenhangt met variabele *Aantal contacten met specialist*, is het aannemelijk dat hetzelfde vragenlijsteffect hier ook speelt. Voor mogelijke verklaringen voor het kleine mode-effect, zie paragraaf 5.2.

Tabel 8: Resultaten *Aantal contacten met specialist*

Variabele	Aantal contacten	
Aantal contacten met specialist (behandeling 1)	1,9	(0,1)
Aantal contacten met specialist (behandeling 2)	2,3	(0,1)
Aantal contacten met specialist (behandeling 3)	2,0	(0,1)

De standaardfouten zijn tussen haakjes achter de schattingen vermeld.

6. Conclusie en aanbevelingen

De verklaringen bij mode-effecten zijn slechts mogelijke verklaringen. Het is mogelijk om nog vervolganalyses uit te voeren om verklaringen uit te sluiten of om te onderzoeken of bepaalde verklaringen ondersteund kunnen worden. Voorbeelden hiervan zijn:

- onderzoeken welke type respondenten (deelpopulaties) responderen via websurveys. Dezelfde type respondenten kunnen dan uit het bestand van de paralleltraject gefilterd worden. Er kan gekeken worden of verschillen zijn tussen de antwoorden via websurvey en CAPI (paralleltraject).
- Onderzoeken in hoeverre “Weet niet” en “Weiger” invloed hebben op de variabele *Perc contact met alternatieve genezer*. Er kan onderzocht worden welke achtergrondkenmerken de respondenten hebben, die contact hebben met een alternatieve genezer en welke type respondenten niet. Vervolgens kan gekeken worden welk percentage van die type respondenten “Weet niet” en “Weiger” hebben gekozen bij behandeling 2.

Referenties

- Ariel, A., D. Giesen, F. Kerssemakers & R. Vis-Visschers (2008). *Literature Review on Mixed-Mode Studies*. Centraal Bureau voor de Statistiek. Versie 16 april 2008.
- Beukenhorst, D.J. & B. Schouten (2009). *Methodenreeks: Mixed Mode benadering*. Centraal Bureau voor de Statistiek. Versie 23 oktober 2009.
- Beukenhorst, D.J. & F. Kerssemakers (2010). *Methodenreeks: Hoofdhema: Benaderstrategieën; Deelthema: algemene beschrijving en uni-mode designs*. Versie 7 januari 2010.
- Bowling, A. (2005). *Mode of questionnaire administration can have serious effects on data quality*. Advance Access Publication. Volume 27, No.3, pp.281-291
- 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. & M. van Herten (2011). *Weging Gezondheidsenquête 2010*. Centraal Bureau voor de Statistiek. Versie mei 2011.
- Cremers, M. & M. Schulte (2010) *Gezondheidsenquête 2010: Vraagteksten en schema's CATI/CAPI*. Centraal Bureau voor de Statistiek, versie 1.0.p14
- Dilman, D.A. & L. M. Christian (2005). *Survey Mode as a Source of Instability in Responses across Surveys*. Field Methods. Volume 17, No. 1, pp. 30-52
- Feskens, R., J. Kappelhof, J. Dagevos & I. Stoop (2010). *Minderheden in de Mixed-Mode? Een inventarisatie van voor- en nadelen van mixed-mode dataverzameling onder niet-westerse migranten*. Sociaal en Cultureel Planbureau, versie augustus 2010
- Geers, X. (2009) *Gezondheidsenquête 2009: Vraagteksten en schema's (basis)vragenlijst*. Centraal Bureau voor de Statistiek, versie 2009C 1.0.p2
- Giesen, D., V. Meertens, R. Vis-Visschers & D. Beukenhorst (2010) *Methodenreeks: Thema: Vragenlijstontwikkeling*. Centraal Bureau voor de Statistiek, versie 24 februari 2010
- Holm, S. (1979). *A simple sequential rejective multiple test procedure*. Chalmers University of Technology, Göteborg. Scandinavian Journal of Statistics 6:65-70, 1979.
- Jansen, E.P.W.A. & Th. H. Joostens (1998). *Enquêteren: Het opstellen en gebruiken van vragenlijsten*. Wolters Noordhoff Groningen, 2^e druk.
- Leeuw, E.D. de (2005). *To Mix or Not to Mix Data Collection Modes in Surveys*. Journal of Official Statistics. Volume 21, No. 2, pp. 233-255

- 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 & M. van Herten (2009). *Transitietraject gezondheidsenquête bij het herontwerp persoonsenquête*. Centraal Bureau voor de Statistiek. Versie 4 december 2009.
- Schulte, M & D. Mohren (2010). *Samenvatting Evaluatie GEZO en V&S (CAPI en CATI)*. Centraal Bureau voor de Statistiek. Versie 18 november 2010
- Toepoel, V., M. Das & A. van Soest (2009). *Design of Web Questionnaires: The Effect of Layout in Rating Scales*. Journal of Official Statistics. Volume 25, No. 4, pp. 509-528
- Vannieuwenhuyze, J., G. Loosveldt & G. Molenberghs (2010). *A Method for Evaluating Mode Effects in Mixed-Mode Surveys*. Public Opinion Quarterly. Volume 75, No. 5, pp. 1027-1045
- 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.

Bijlage A

Tabel 9: Cijfers voor 2010 op basis van voorspellingen (Behandeling 1).

Variabele	Cat 1	Cat 2	Cat 3	Sf. cat 1	Sf. Cat 2	Sf. Cat 3
Aantal contacten met fysiotherapeut	3,36			0,18		
Aantal contacten met huisarts	3,95			0,15		
Aantal contacten met specialist	1,86			0,04		
Aantal contacten met tandarts	2,07			0,10		
Perc anticonceptiepil	38,19	61,81		1,49	1,49	
Perc beperking bewegen	7,78	92,22		0,53	0,53	
Perc beperking horen	2,92	97,08		0,17	0,17	
Perc beperking zien	5,79	94,21		0,55	0,55	
Perc contact met fysiotherapeut	21,45	78,55		0,65	0,65	
Perc contact met huisarts	74,46	25,54		0,75	0,75	
Perc contact met specialist	43,98	56,02		0,72	0,72	
Perc contact met tandarts	78,83	21,17		0,69	0,69	
Perc langdurige aandoening	43,07	56,93		0,35	0,35	
Perc niet-voorgescreven medicijn	39,17	60,83		0,68	0,68	
Perc norm gezond bewegen	57,13	42,87		0,48	0,48	
Perc rokers	27,63	72,37		0,57	0,57	
Perc voorgeschreven medicijn	40,38	59,62		0,39	0,39	
Perc ziekenhuisopname	6,92	93,08		0,32	0,32	
Perc zware drinkers	10,37	89,63		0,55	0,55	
Perc zware rokers	6,29	93,71		0,26	0,26	
Perc goede ervaren gezondheid	81,38	18,62		0,52	0,52	
Perc contact met alternatief genezer	6,90	93,10		0,28	0,28	
Perc matig en ernstig overgewicht	35,50	12,01	52,59	0,59	0,39	0,64

Tabel 10: Cijfers voor 2010 nieuwe opzet (Behandeling 2).

Variabele	Cat 1	Cat 2	Cat 3	Sf. cat 1	Sf. Cat 2	Sf. Cat 3
Aantal contacten met fysiotherapeut						
Aantal contacten met huisarts	4,2			0,1		
Aantal contacten met specialist	2,3			0,1		
Aantal contacten met tandarts	2,3			0,1		
Perc anticonceptiepil	39,0	61,0		0,8	0,8	
Perc beperking bewegen	7,4	92,6		0,3	0,3	
Perc beperking horen	2,9	97,1		0,2	0,2	
Perc beperking zien	5,9	94,1		0,3	0,3	
Perc contact met fysiotherapeut	22,0	78,0		0,3	0,3	
Perc contact met huisarts	72,3	27,7		0,4	0,4	
Perc contact met specialist	37,8	62,2		0,4	0,4	
Perc contact met tandarts	78,4	21,6		0,3	0,3	
Perc langdurige aandoening						
Perc niet-voorgescreven medicijn	34,2	65,8		0,4	0,4	
Perc norm gezond bewegen						
Perc rokers	26,8	73,2		0,4	0,4	
Perc voorgeschreven medicijn	35,8	64,2		0,4	0,4	
Perc ziekenhuisopname	6,5	93,5		0,2	0,2	
Perc zware drinkers	9,7	90,3		0,4	0,4	
Perc zware rokers	5,2	94,8		0,2	0,2	
Perc goede ervaren gezondheid	80,3	19,7		0,3	0,3	
Perc contact met alternatief genezer	6,0	94,0		0,2	0,2	
Perc matig en ernstig overgewicht	37,0	10,9	52,1	0,4	0,3	0,5

Tabel 11: Cijfers voor 2010 op basis van 7 maanden (Behandeling 3).

Variabele	Cat 1	Cat 2	Cat 3	Sf. cat 1	Sf. Cat 2	Sf. Cat 3
Aantal contacten met fysiotherapeut						
Aantal contacten met huisarts	4			0,2		
Aantal contacten met specialist	2			0,1		
Aantal contacten met tandarts	2,1			0,1		
Perc anticonceptiepil	37,5	62,5		1,6	1,6	
Perc beperking bewegen	7,3	92,7		0,6	0,6	
Perc beperking horen	2,4	97,6		0,3	0,3	
Perc beperking zien	5	95		0,5	0,5	
Perc contact met fysiotherapeut	23,8	76,2		0,7	0,7	
Perc contact met huisarts	72,9	27,1		0,7	0,7	
Perc contact met specialist	39,3	60,7		0,8	0,8	
Perc contact met tandarts	77,1	22,9		0,7	0,7	
Perc langdurige aandoening	41,8	58,2		1	1	
Perc niet-voorgeschreven medicijn	40,2	59,8		0,8	0,8	
Perc norm gezond bewegen						
Perc rokers	29,6	70,4		0,8	0,8	
Perc voorgeschreven medicijn	36,8	63,2		0,8	0,8	
Perc ziekenhuisopname	5,8	94,2		0,4	0,4	
Perc zware drinkers	8,8	91,2		0,6	0,6	
Perc zware rokers	7,3	92,7		0,5	0,5	
Perc goede ervaren gezondheid	81,7	18,3		0,6	0,6	
Perc contact met alternatief genezer	7,6	92,4		0,4	0,4	
Perc matig en ernstig overgewicht	37,5	12,6	49,9	0,9	0,6	0,9