Centraal Bureau voor de Statistiek

Aanvullend Voorzieningengebruik Onderzoek (AVO) 2007

Onderzoeksdocumentatie

Onderzoekspartners

In deze rapportage zijn de belangrijkste achtergrondgegevens opgenomen van het proces van het Aanvullend Voorzieningengebruik Onderzoek (AVO) die (mede) in opdracht van het Sociaal Cultureel Planbureau (SCP) en door het Centraal Bureau voor de Statistiek (CBS) in de laatste maanden van 2007 is verricht.

De onderzoeksdocumentatie is samengesteld door A.B.P. Buis
telefoon: 045 5706919 email: apb.buis@cbs.nl

Samenvatting

In deze onderzoeksdocumentatie worden de voorbereiding, verzameling en verwerking van de vragenlijstgegevens in al hun facetten gedocumenteerd. Daarnaast fungeert dit document als een contactmedium voor de diverse geledingen die aan het onderzoek bijdragen en als een naslagwerk voor onderzoekers die meer inzicht wensen in de achtergronden van het onderzoek en in het tot stand komen van de analysebestanden.

Inhoudsopgave

1. Doel van het onderzoek

Het Aanvullend Voorzieningengebruik Onderzoek (AVO) is een 4-jaarlijks onderzoek, met als doel:

Het inzicht verkrijgen in de mate waarin door de Nederlandse huishoudens, alsmede door alle personen van 6 jaar of ouder binnen deze huishoudens, gebruik wordt gemaakt van sociale en culturele voorzieningen van zeer uiteenlopende aard.

Het onderzoek wordt ‘aanvullend‘ genoemd omdat de informatie - naar onderwerp en/of naar detaillering- complementair is ten opzichte van informatie uit andere bronnen. Het onderzoek is in 2007 door het CBS uitgevoerd op verzoek van en in samenwerking met het Sociaal en Cultureel Planbureau (SCP). Het AVO 2007 is het achtste onderzoek in de vierjaarlijkse reeks die is gestart in 1979. Hieronder staan de respons resultaten, de steekproefmethoden, de veldwerkperiodes en de uitvoerders van de verschillende Vooronderzoeken door de jaren heen.

AVO'79

Uitvoerder veldwerk NSS / Marktonderzoek
Veldwerkperiode september 1979 - november 1979
Steekproefmethode enkelvoudige aselecte adressensteekproef
Steekproefomvang 9915 huishoudens
Respons 6431 huishoudens; 17.232 personen (65%)

AVO'83

Uitvoerder veldwerk NSS / Marktonderzoek
Veldwerkperiode september 1983 - november 1983
Steekproefmethode enkelvoudige aselecte adressensteekproef
Steekproefomvang 9908 huishoudens
Respons 5774 huishoudens; 14.869 personen (58%)

AVO'87

Uitvoerder veldwerk NSS / Marktonderzoek
Veldwerkperiode oktober 1987 - december 1987
Steekproefmethode enkelvoudige aselecte adressensteekproef, met extra adressen in vier grote steden + Haarlem
Steekproefomvang 10.302 huishoudens
Respons 6496 huishoudens; 16.151 personen (63%)

AVO'91

Uitvoerder veldwerk NSS / Marktonderzoek
Veldwerkperiode september 1991 - december 1991
Steekproefmethode tweetrapssteekproef: gemeenten/adressen; stratificatie naar gemeentegrootte
Steekproefomvang 12.797 huishoudens
Respons 5458 huishoudens; 13.105 personen (43%)

AVO'95

Uitvoerder veldwerk GFK Interact
Veldwerkperiode september 1995 - januari 1996
Steekproefmethode enkelvoudige aselecte adressensteekproef
Steekproefomvang 9305 huishoudens
Respons 6421 huishoudens; 14.489 personen (69%)

AVO'99

Uitvoerder veldwerk GfK Nederland
Veldwerkperiode september 1999 - februari 2000
Steekproefmethode enkelvoudige aselecte adressensteekproef
Steekproefomvang 9216 huishoudens
Respons 6125 huishoudens; 13.490 personen (66%)

AVO2003

Uitvoerder veldwerk GfK Panel Services Benelux
Veldwerkperiode september 2003 - maart 2004
Steekproefmethode enkelvoudige aselecte adressensteekproef
Steekproefomvang 10.680 huishoudens (inclusief niet bereikt)
Respons 6404 huishoudens; 13.776 personen (60%)

2. Design AVO 2007

2.1 Vragenlijst

Het onderzoek kent twee fases. De eerste fase betreft het afnemen van een huishoudvragenlijst die wordt beantwoord door iemand uit de huishoudkern (degene die alles regelt in het huishouden of zijn/haar partner). Deze vragenlijst (de zogenaamde huishouden-vragenlijst) wordt face-to-face afgenomen (CAPI). In de huishoudvragenlijst wordt geworven voor de tweede fase van het onderzoek, waarin ieder lid van het huishouden van 6 jaar of ouder een schriftelijke vragenlijst invult (PAPI).

2.2 Schriftelijke vragenlijsten

Er worden twee schriftelijke vragenlijsten gebruikt:

I. Een vragenlijst voor personen in de leeftijd van 6 tot en met 15 jaar; de Jeugdvragenlijst.

II. Een vragenlijst voor personen van 16 jaar of ouder; de Volwassenenvragenlijst.

Voor het invullen van de schriftelijke vragenlijsten door kinderen wordt door het CBS de volgende standaard werkwijze gehanteerd:

Voor kinderen jonger dan 12 jaar wordt de Jeugdvragenlijst door de ouders/verzorgers ingevuld (proxi).
Voor kinderen in de leeftijd van 12 jaar tot en met 15 jaar wordt aan de ouders/verzorgers gevraagd of het kind de vragenlijst zelf mag invullen. Indien dit niet het geval is wordt de Jeugdvragenlijst door de ouders/verzorgers ingevuld (proxi).
Kinderen in de leeftijd van 16 jaar of ouder vullen zelf de Volwassenenvragenlijst in.

2.3 Invultijd van de enquêtes

De duur van de huishoudvragenlijst bedraagt gemiddeld 25 tot 30 minuten. De duur van de Volwassenenvragenlijst bedraagt gemiddeld 30 tot 45 minuten en die van de Jeugdvragenlijst gemiddeld 20 tot 30 minuten.

2.4 Split half

De structuur van de schriftelijke vragenlijsten blijft grotendeels ongewijzigd ten opzichte van de schriftelijke vragenlijsten van het AVO-onderzoek van 2003.
Het verschil met ‘AVO 2003’ bestaat eruit dat een split-half design wordt toegepast, waarbij vragenblokken in twee versies worden aangeboden. Het gaat hierbij om de sport- en cultuurvragen. In de praktijk betekent dit dat er twee versies van beide vragenlijsten zijn.

Om te bepalen wie welke vragenlijst moet krijgen is gekeken naar het identificatienummer; het zogenaamde WE-ID. Als het - op één na laatste- cijfer 'even' is, krijgen de leden van het huishouden blauwe vragenlijsten; donkerblauwe vragenlijsten voor volwassenen en lichtblauwe vragenlijsten voor kinderen. Is het -op één na- laatste cijfer oneven, dan krijgen de leden van het huishouden groene vragenlijsten; donkergroene vragenlijsten voor volwassenen en lichtgroene vragenlijsten voor kinderen. Zie Staat 2.4 Split half vragenlijsten.

WE_ID (voorlaatste cijfer)	**Volwassen (16 jaar en ouder)**	**Jeugd (jonger dan 16 jaar)**
even	kleur: donkerblauw	kleur: lichtblauw
	volgnummer start met (A)	volgnummer start met (C)
oneven	kleur: donkergroen	kleur: lichtgroen
	volgnummer start met (B)	volgnummer start met (D)

Staat 2.4: Split-half vragenlijsten

2.5 Navraagformulier

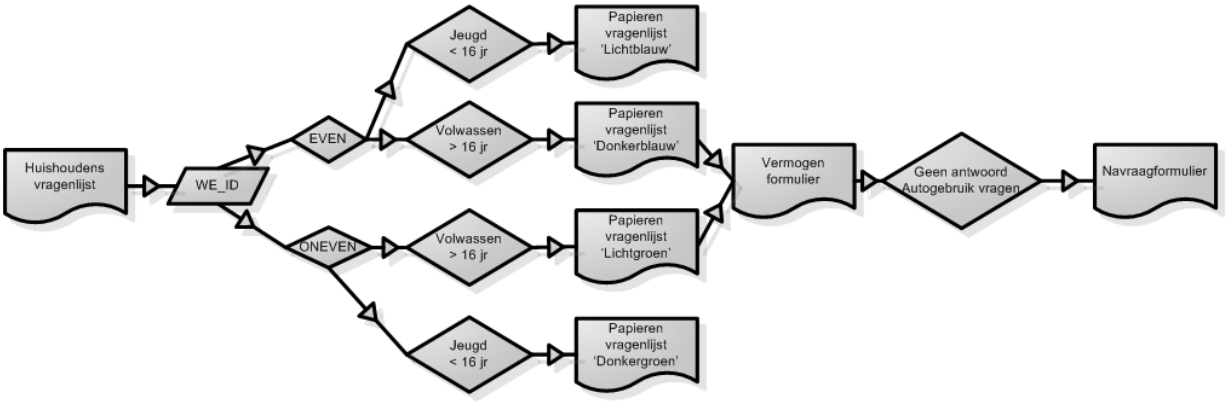
Het kan voorkomen dat een respondent tijdens het face-to-face interview niet direct alle antwoorden op de vragen, gesteld in de huishoudvragenlijst over het autogebruik, voorhanden heeft. De interviewer laat in dit geval het zogenaamde Navraagformulier achter, waarop de respondent de ontbrekende informatie kan invullen. Het Navraagformulier wordt door de interviewer tezamen met de schriftelijke vragenlijsten achtergelaten bij de respondent na het afnemen van de huishoudvragenlijst.

2.6 Vermogensformulier

De informatie over bezittingen en schulden die via andere bronnen beschikbaar is, is niet toereikend. Daarom wordt er naast de andere papieren vragenlijsten, aan de persoon die de financiën van het huishouden regelt, gevraagd om het zogenaamde Vermogensformulier in te vullen.

2.7 Schema vragenlijsten

In figuur 2.7 is de beschrijving van de bovenstaande paragrafen verduidelijkt in een illustratie.



Schema 2.7: AVO vragenlijsten

2.8 Planning

Uitgaande van de start van het AVO 2007 in september 2007 en een dataverzamelingsperiode van drie maanden ziet de planning er op hoofdlijnen uit zoals vermeld in Staat 2.8.

	start	eind
Onderzoeksdesign	11-apr	1-mei
Onderzoeksdefinitie	25-apr	1-jun
Steekproef	25-apr	19-jul
Weging	10-dec	28-mrt-08
Vragenlijst	10-apr	6-aug
Integrale logistieke test	28-mei	6-aug
Instructie interviewers	28-mei	23-aug
Verwerking	1-sept	31-mrt 08

Staat 2.8: Planning op hoofdlijnen

2.9 Benaderingsstrategie

Alle steekproefadressen worden in de eerste fase van het onderzoek aan-huis benaderd. Toegepast op het AVO ziet deze afgesproken benaderingsstrategie er in grote lijnen als volgt uit:

2.9.1 Instructie

Alle interviewers zijn specifiek getraind in “medewerking verkrijgen” van respondenten. Ten behoeve van het AVO zullen alle interviewers een schriftelijke instructie ontvangen. Daarnaast zullen interviewers die na 1-1-2006 bij het CBS in dienst zijn gekomen een mondelinge instructie ontvangen. “Medewerking verkrijgen” vormt, voor het AVO, onderdeel van deze instructies.

2.9.2 Aanschrijfbrief

De adressen in een maandportie ontvangen voorafgaand aan het eerste bezoek een zogenoemde aanschrijfbrief waarin het bezoek van de interviewer wordt aangekondigd en het doel van het onderzoek wordt toegelicht. Bij de aanschrijfbrief is een incentive ter waarde van 10 postzegels toegevoegd.

2.9.3 Bezoeken

In een maandportie worden alle steekproefadressen in de eerste helft van de betreffende maand ten minste één keer bezocht. Wordt de beoogde respondent thuis getroffen dan probeert de interviewer een afspraak te maken voor een interview zodanig dat een spreiding van de responsen in de tijd wordt bereikt.

Bij de eerste drie vergeefse bezoeken wordt een kaartje achtergelaten. Vanaf het derde bezoek probeert de interviewer telefonisch contact te leggen met de beoogde respondent om een afspraak te maken voor een interview. De interviewer blijft ondertussen het adres bezoeken.

Een adres kan uitsluitend worden afgeboekt met als eindresultaat ‘geen contact’ na zes vergeefse bezoeken, die gelijkmatig zijn gespreid over de gehele veldwerkperiode van een maand.

2.9.4 Onderzoek

Indien de respondent bereid is om deel te nemen aan het onderzoek wordt de Huishoudvragenlijst afgenomen; de eerste fase van het onderzoek. Vervolgens proberen de interviewers de respondent te werven voor de tweede fase van het onderzoek waarin ieder lid van het huishouden in de leeftijd van 6 jaar of ouder een schriftelijke vragenlijst moet invullen. De interviewer noteert in de huishoudvragenlijst bij wie welke schriftelijke vragenlijst is achtergelaten. Ook noteert de interviewer op de schriftelijke vragenlijsten welk lid van het huishouden welke schriftelijke vragenlijst moet invullen.

De interviewer kent de respondent een incentive toe van 10 euro zodra deze aangeeft mee te willen werken aan het vervolgonderzoek. Respondenten die bereid zijn mee te werken krijgen de keuze tussen een Irischeque ter waarde van 10 euro of een cadeau van dezelfde waarde. De interviewer krijgt vanuit de Huishoudvragenlijst teruggekoppeld welke versie van de schriftelijke vragenlijsten moet worden achtergelaten in verband met het split-half design (zie schema 2.7). Dit betekent dat de interviewers van beide versies van de schriftelijke vragenlijsten voldoende voorraad moeten hebben.

2.9.5 Navraagformulier

Het kan voorkomen dat een respondent tijdens het face-to-face interview niet direct alle in de huishoudvragenlijst gevraagde informatie voorhanden heeft. De interviewer laat in dit geval het zogenaamde Navraagformulier achter, waarop de respondent de ontbrekende informatie kan invullen. De interviewer noteert in de Huishoudvragenlijst of bij het huishouden een Navraagformulier is achtergelaten. Ook noteert de interviewer op het Navraagformulier een met het huishouden overeenkomstig identificatienummer, zodat de informatie op het Navraagformulier aan de Huishoudvragenlijst gekoppeld kan worden.

2.9.6 Vermogensformulier

De informatie over bezittingen en schulden die via andere bronnen beschikbaar is, is niet toereikend. Daarom wordt er naast de andere papieren vragenlijsten, aan de persoon die de financiën van het huishouden regelt, gevraagd om het zogenaamde Vermogensformulier in te vullen. De interviewer laat hiervoor het zogenaamde Vermogensformulier achter. De interviewer noteert in de Huishoudvragenlijst of bij het huishouden een vermogensformulier is achtergelaten. De interviewer noteert op het Vermogensformulier het huishoudensidentificatienummer; met dit nummer kan de data van het Vermogensformulier aan de Huishoudvragenlijst gekoppeld worden.

2.9.7 Ophaalbezoek

De interviewer maakt een afspraak met het huishouden om de schriftelijke vragenlijsten, het Vermogensformulier en het eventueel achtergelaten Navraagformulier ongeveer een week na het afnemen van de huishoudvragenlijst op te komen halen. Tijdens dit ophaalbezoek controleert de interviewer of iedereen de schriftelijke vragenlijst heeft ingevuld. Het meest efficiënt zou zijn als de interviewer de gegevens op het eventuele navraagformulier en het vermogensformulier meteen zou verwerken in de huishoudvragenlijst. Aangezien deze zo snel mogelijk na het face-to-face bezoek wordt teruggestuurd naar het CBS is dit niet mogelijk. Om dit wel mogelijk te maken zou de interviewer de huishoudvragenlijst nog minstens een week na het face-to-face interview moeten vasthouden. Dit betekent echter ook dat de voortgang alleen maar met een vertraging van een week kan worden gemeten. Dit is niet gewenst.

De interviewer neemt voordat het ophaalbezoek wordt afgelegd telefonisch contact op met het huishouden om te vragen of alle vragenlijsten zijn ingevuld. Op deze manier wordt voorkomen dat de interviewer vergeefs op pad gaat voor het ophaalbezoek. Nadeel van deze methode is dat respondenten gemakkelijker weigeren aan de telefoon dan tijdens een aan-huis bezoek, waardoor de hoeveelheid weigeringen om deel te nemen aan het onderzoek kan toenemen. Om telefonisch contact op te kunnen nemen met het huishouden dient het telefoonnummer van het huishouden bekend te zijn. Dit betekent dat de interviewer dit aan de respondent moet vragen.

Indien een interviewer niet zelf het ophaalbezoek kan uitvoeren meldt de interviewer aan de respondent dat het ophaalbezoek door een andere interviewer wordt uitgevoerd. De interviewer meldt aan zijn / haar regiomanager dat het ophaalbezoek door een andere interviewer moet worden uitgevoerd, en geeft daarbij aan welke afspraken met de respondent zijn gemaakt.
De interviewer verantwoordt het ophaalbezoek in een apart formulier. Hierin wordt bijvoorbeeld aangegeven hoeveel bezoeken zijn afgelegd alvorens alle schriftelijke vragenlijsten in een huishouden waren ingevuld en wat de reden van een eventuele weigering is.

De interviewer retourneert de ingevulde schriftelijke vragenlijsten, het Vermogensformulier, het eventuele Navraagformulier en de bijbehorende verantwoording van het ophaalbezoek zo snel mogelijk naar het CBS.

2.9.8 Vertoetsen

De ingevulde schriftelijke vragenlijsten, Vermogensformulieren en Navraagformulieren en de verantwoordingen van de ophaalbezoeken worden vertoetst op het CBS. Alle teruggestuurde vragenlijsten worden vertoetst, dus ook van huishoudens waarvoor niet van ieder lid van het huishouden van 6 jaar of ouder een schriftelijke vragenlijst is ontvangen.

2.10 Koppeling met registers

In CBS-onderzoeken worden in de regel de vragen naar inkomensgegevens niet meer in de vragenlijst opgenomen omdat deze gegevens door koppeling met registraties van de Belastingdienst kunnen worden verkregen. Om de koppeling met de inkomensgegevens tot stand te brengen worden de gegevens van alle personen uit het huishouden voorzien van het bijbehorende Sofinummer. Op basis van de sofi-nummers (te bepalen op basis van adres, geslacht en geboortedatum) van deze personen kunnen voor iedere persoon de inkomensgegevens worden bepaald. Daarmee worden vervolgens de inkomensgegevens op huishoudniveau afgeleid. De huishoudsamensteling komt op deze manier per definitie overeen met de situatie op de enquêtedatum.

De volgende afspraak is overeengekomen voor wat betreft de levering van de Inkomensgegevens:

1) De Inkomensgegevens worden separaat aangeleverd in jaarbestanden.
2) De Inkomensgegevens van 2005, 2006, 2007 en 2008 zijn geleverd.
3) De Inkomensgegevens worden in een afgesproken formaat aangeleverd. De gegevens per persoon zijn vastgelegd in te variabelen-lijsten.

3 Steekproef

3.1 Steekproefkader

De steekproefkaders worden samengesteld op basis van persoon- en adresgegevens afkomstig uit de Gemeentelijke Basisadministratie (GBA). Jaarlijks worden nieuwe steekproefkaders aangemaakt, die geen overlap hebben met de steekproefkaders uit het voorafgaande jaar. De steekproefkaders worden voortdurend geactualiseerd met GBA-informatie over verhuizingen, geboortes en sterfte. Steekproeven waar de persoon de trekkings-eenheid is, worden getrokken uit het steekproefkader van personen. Op basis van de persoonsgegevens kan de doelpopulatie worden onderscheiden. Een alternatief voor het steekproefkader van personen is het adressensteekproefkader, dat gebruikt wordt voor steekproeven waar huishoudens worden waargenomen. Dit kader bevat ook gegevens over het aantal postale afgiftepunten afkomstig uit het Geografisch Basisregister (GBR). Meerdere afgiftepunten op een adres kan duiden op meerdere huishoudens op een adres. Hiermee wordt rekening gehouden bij het trekken van steekproeven uit dit steekproefkader.

3.2 Steekproefontwerp

3.2.1 Tweetrapssteekproef-ontwerp

Het steekproefontwerp voor het AVO 2007 is een zelfwegend tweetrapssteekproef-ontwerp met gemeenten als primaire eenheden en adressen als secundaire eenheden. Het steekproefontwerp in de eerste trap is van het gestratificeerde type, waarbij de gemeenten zijn ingedeeld naar de kenmerken coropgebied en interviewerregio. In ieder stratum zijn via een systematische steekproef gemeenten getrokken met kansen evenredig aan het aantal adressen per gemeente. Tevens is voor elke geselecteerde gemeente het aantal te trekken adressen berekend. De tweede trap is een random steekproef van adressen in de geselecteerde gemeenten, met omvangen zoals vastgesteld in de eerste trap. De belangrijkste kenmerken van het steekproefontwerp zijn in Staat 3.1 weergegeven.

Type steekproef	gestratificeerde tweetrapssteekproef
Frequentie	4 perioden: sep, okt, nov en dec
Kenmerken 1ste trap	
Stratumindeling	interviewerregio × coropgebied
wijze van trekking steekproefeenheden 1ste trap	systematische steekproef met ongelijke kansen gemeenten
Kenmerken 2de trap	
wijze van trekking	aselecte steekproef
steekproefeenheden 2de trap	adressen
Te trekken steekproefomvang	3068 (sep), 3068 (okt), 3068 (nov), 2498 (dec)
Uit te zetten steekproefomvang	2851 (sep), 2851 (okt), 2851 (nov), 2308 (dec)
Clusteromvang	12 adressen per gemeente

Staat 3.1 Kenmerken steekproefontwerp

3.2.2 Cluster

De aantallen te trekken adressen per stratum zijn evenredig met de aantallen adressen volgens de GBA per stratum. Het gewenst aantal te trekken adressen per gemeente is vooraf bepaald en voor iedere gemeente gelijk. Deze zogenoemde clusteromvang is voor het AVO gelijk aan twaalf. Voor gemeenten die met kans 1 zijn geselecteerd, de zogenoemde zelfselecterende gemeenten, is het aantal te trekken adressen aangepast tot het product van de steekproeffractie (quotiënt van steekproefomvang en populatieomvang) en het aantal adressen in de betreffende gemeente. Voor alle andere gemeenten kan het aantal te trekken adressen iets afwijken van de clusteromvang. Dit komt door afrondingen en de keuze om in elke maand ten minste 1 gemeente per stratum te selecteren. Op deze manier ontstaat een zelfwegende steekproef. Dit betekent dat de insluitkans, dat is de kans dat een adres wordt geselecteerd, voor elk adres hetzelfde is. Voor elk adres is de insluitkans gelijk aan de steekproeffractie.
Dat de clusteromvang (*c*) op twaalf is gezet, is een compromis tussen kosten en precisie. Als *c* klein is, dan wordt het aantal te trekken gemeenten groot met als gevolg hogere reiskosten voor de interviewers. Bij een grote *c* kan er een clustereffect optreden, waardoor de foutenmarges groter worden. Dit is vooral het geval als er een samenhang bestaat tussen de antwoordpatronen van de bewoners van eenzelfde gemeente. De ervaring leert dat bij waarden van *c* tussen tien en twintig het clustereffect acceptabel is en de reiskosten niet te hoog zijn. Het clustereffect hangt overigens ook af van de variabelen die gemeten worden en de reiskosten hangen af van de spreiding van het enquêtecorps. Voor telefonische, schriftelijke en internet-enquêtes speelt het punt van de reiskosten niet en kan *c* op 1 worden gesteld waardoor het clustereffect verdwijnt.

3.2.3 Spreiding

De responsverplichting voor het AVO stelt hoge eisen aan de uitvoering van de dataverzameling. Een permanente bewaking van de voortgang en een goede rappelstrategie zijn daarvoor een vereiste. Vanwege de sluiting van het CBS van 24 tot en met 31 december zijn de mogelijkheden om in deze periode de voortgang te bewaken en de rappelstrategie toe te passen beperkt. Daarom is ervoor gekozen om in december een kleinere AVO-steekproef uit te zetten dan in september, oktober en november. De AVO-steekproef is gespreid over september, oktober, november en december met verhouding 105 : 105 : 105 : 85. Bij deze spreiding zijn de gevolgen voor de spreiding van de steekproeven van overige CBS-onderzoeken beperkt.

3.2.4 Steekproefomvang

De te trekken omvangen van de steekproeven zoals vermeld in Staat 3.1 zijn als volgt vastgesteld. Er moeten 10.850 adressen aan huis worden benaderd om van ten minste 6.000 huishoudens ingevulde huishoudens- én persoonsvragenlijsten te verkrijgen. Hierbij is verondersteld dat de responskans op de capi-vragenlijst 63% is en dat 88% van de huishoudens die face-to-face gerespondeerd hebben ook alle persoonsvragenlijsten invullen. Bij de uitvoering van het onderzoek zijn 11 adressen extra uitgezet.

De uitgezette steekproefomvang in december is (10.861 / 4) × 0,85 = 2.308 en in de overige maanden (10.861 / 4) × 1,05 = 2.851. Onder de aanname dat maximaal 7% van een getrokken steekproef verloren gaat bij de steekproefbewerking (zie paragraaf 3.2), zouden voor december ten minste 2.308 / 0,93 = 2.482 adressen moeten worden getrokken en voor de overige maanden 2.851 / 0,93 = 3.066 adressen. De daadwerkelijk getrokken aantallen staan vermeld in Staat 3.1 en zijn iets groter dan laatstgenoemde getallen.

3.3 Steekproefbewerking

De steekproef waarvoor de dataverzameling in zekere maand wordt uitgevoerd, is ongeveer zes weken van tevoren getrokken. Dus eind juli, augustus, september respectievelijk oktober zijn de steekproeven voor dataverzameling in september, oktober, november respectievelijk december getrokken. Aansluitend op de trekking zijn de steekproeven bewerkt. Deze bewerking begon met het verwijderen van steekproefadressen:

- met onvolledige of onbekende adresinformatie,
- die tot de institutionele bevolking behoren,
- die de afgelopen 12 maanden in een andere CBS-steekproef voorkwamen,
- op de Waddeneilanden,
- in de postcodegebieden 1102, 1103 en 1104 van De Bijlmer.

Vervolgens zijn de overgebleven steekproefadressen in de zogenoemde screeningsbakken geplaatst. Hierdoor worden zij de komende twaalf maanden niet benaderd voor andere CBS-onderzoeken. Daarna zijn de steekproeven uitgedund tot de gewenste uit te zetten omvangen zoals opgenomen in Staat 3.2.

Bij de uit te zetten steekproefadressen zijn telefoonnummers gezocht via de externe partij Cendris opdat de interviewers de benaderingsstrategie optimaal kunnen uitvoeren. In Staat 3.2 is de steekproefverantwoording opgenomen.

	sep	okt	nov	dec	totaal
Getrokken	3.068	3.068	3.068	2.498	11.702
Niet beschikbaar voor uitzet	64	68	53	50	235
In eerdere CBS steekproef	30	40	26	27	123
Institutionele bevolking	16	14	10	11	51
Gebied buiten waarneming	7	9	9	8	33
Adresgegevens onjuist	11	5	8	4	28
Beschikbaar voor uitzet	3.004	3.000	3.015	2.448	11.467
Uitgedund	153	149	164	140	606
Uitgezet	2.851	2.851	2.851	2.308	10.861
Met telefoonnummer	1.129	1.126	1.069	852	4.176
Zonder telefoonnummer	1.722	1.725	1.782	1.456	6.685

Staat 3.2 Steekproefverantwoording per waarneemperiode

Bij de uitvoering van het veldwerk zijn twee verschillende versies van de schriftelijke vragenlijst gebruikt. De uitgezette steekproef is willekeurig verdeeld in twee deelsteekproeven. Huishoudens in de ene deelsteekproef hebben de ‘blauwe’ vragenlijsten gekregen (Donkerblauw (volgnummer A) voor de volwassenen; Lichtblauw (volgnummer C) voor de jeugd), huishoudens in de andere deelsteekproef zijn benaderd met ‘groene vragenlijsten gekregen (Donkergroen (volgnummer B) voor de volwassenen; Lichtblauw (volgnummer D) voor de jeugd).

4 Weging

In een ophoging of weging is het wenselijk om achtergrondkenmerken te gebruiken die samenhangen met de doelvariabelen. Dit verkleint de steekproefvariantie. Hoe sterker de samenhang des te groter is de variantiereductie. Nog belangrijker zijn achtergrondkenmerken die ook samenhangen met de non-respons. Het opnemen van deze kenmerken in het weegmodel zorgt voor een reductie van de non-responsvertekening. Omdat de steekproef zelfwegend is, dat wil zeggen dat alle huishoudens dezelfde insluitkans hebben, krijgen alle huishoudens en personen hetzelfde startgewicht.

4.1 Weegmodel

Het weegmodel van het AVO 2003 is:

*Leeftijd*12 × *Geslacht2* × *BurgerlijkeStaat4*
Geslacht2 × *Provincie12*
Leeftijd12 × *Provincie12*
Leeftijd12 × *Stedelijkheidsgraad5*.

Dit weegmodel heeft als uitgangspunt gediend voor de weging van AVO 2007. Het is op drie punten aangepast:

- Vanwege kleine celvulling is de weegterm *Leeftijd*12 × *Geslacht2* × *BurgerlijkeStaat4* gereduceerd tot *Leeftijd*12 × *Geslacht2* + *Geslacht2* × *BurgerlijkeStaat4*;

- Om dezelfde reden is de leeftijdsvariabele in de weegterm *Leeftijd12* × *Provincie12* ingedikt tot zeven categorieën;

- een extra weegterm *Herkomst3* is toegevoegd om eventuele vertekening door ondervertegenwoordiging van vooral niet-westerse allochtonen te verkleinen.

Het uiteindelijke weegmodel voor AVO 2007 is:

*Leeftijd*12 × *Geslacht2*
Geslacht2 × *BurgerlijkeStaat4*
Geslacht2 × *Provincie12*
Leeftijd7 × *Provincie12*
Leeftijd12 × *Stedelijkheidsgraad5*
Herkomst3.

Huishoudweging

De uitgevoerde weging is een huishoudweging, dat wil zeggen dat alle personen binnen een huishouden hetzelfde gewicht krijgen. Deze strategie wordt ook wel lineair consistent wegen genoemd (Lemaître en Dufour, 1987). De belangrijkste reden voor één gewicht per huishouden is dat deze zowel voor personen als huishoudens gebruikt kan worden. Schattingen voor personen en huishoudens zijn dan consistent. Bovendien kan het huishouden worden beschouwd als de steekproefeenheid.

4.2 Variabelen

De variabelen in het weegmodel hebben de volgende categorieën:

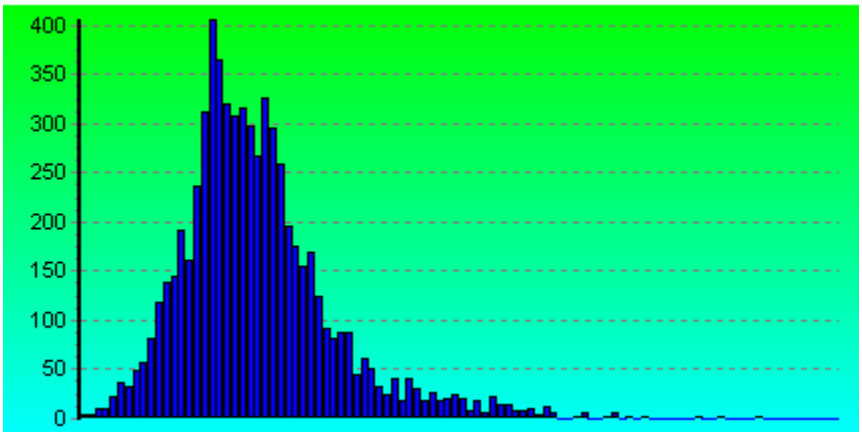
- *Leeftijd12*: 6-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60-69, 70-79, 80+;
- *Leeftijd7*: 6-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70+;
- *Geslacht2*: man, vrouw;
- *BurgerlijkeStaat4*: gehuwd of geregistreerd partnerschap, gescheiden, weduwstaat, nooit gehuwd geweest;
- *Provincie12*: Groningen, Friesland, Drenthe, Overijssel, Flevoland, Gelderland, Utrecht, Noord-Holland, Zuid-Holland, Zeeland, Noord-Brabant, Limburg;
- *Stedelijkheidsgraad5*: zeer sterk, sterk, matig, weinig, niet stedelijk;
- *Herkomst3*: autochtoon, westers allochtoon, niet- westers allochtoon.

4.3 Bascula

De weging van het AVO is uitgevoerd met het weegprogramma Bascula. Voor een handleiding zie Nieuwenbroek en Boonstra (2002). De lineaire weegmethode kan negatieve eindgewichten geven. Om deze weg te werken zonder dat de weging naar populatietotalen wordt verstoord, worden de gewichten begrensd volgens het Huang-Fuller algoritme (Huang en Fuller, 1978). Enkele iteraties zijn normaal gesproken voldoende om positieve eindgewichten te verkrijgen. De populatietotalen die nodig zijn voor de weging zijn afkomstig uit de Gemeentelijke Basisadministratie (GBA). De populatietotalen geven de stand van 1 januari 2008.

4.4 Resultaten van de weging voor de ‘blauwe’ vragenlijsten

Voor de blauwe vragenlijsten zijn er gegevens van 6.594 personen in 3.095 huishoudens. Zonder toepassing van het begrenzingsalgoritme zijn er 4 negatieve gewichten. Het begrenzen van de gewichten verliep in één iteratie. De eindgewichten zijn het product van insluitgewichten en correctiegewichten. Omdat alle personen dezelfde insluitkans hebben, is de verdeling van de correctiegewichten hetzelfde als die van de eindgewichten. In figuur 4.4 is de verdeling van de gewichten na begrenzen te zien.

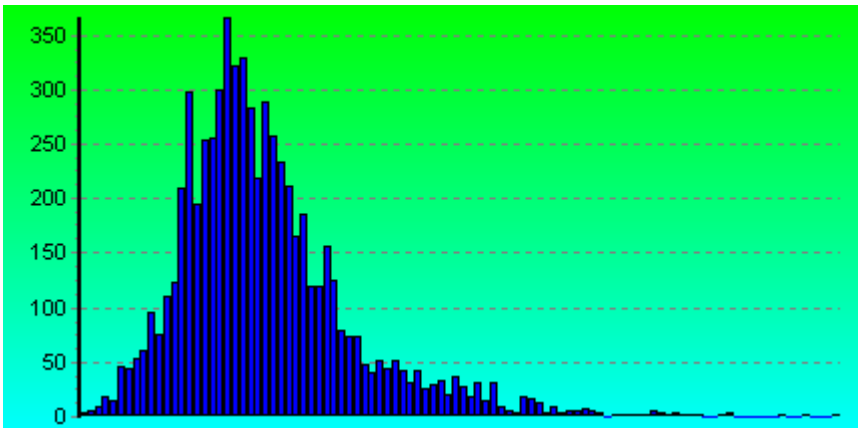


Figuur 4.4 Verdeling van gewichten bij de blauwe vragenlijsten.

De correctiegewichten liggen tussen 0,11 en 3,88. Het eerste kwartiel is 0,76, de mediaan is 0,94 en het derde kwartiel is 1,16. De eindgewichten liggen tussen 256 en 8.979. Het eerste kwartiel is 1.748, de mediaan 2.165 en het derde kwartiel 2.684.

4.5 Resultaten van de weging voor de 'groene' vragenlijsten

Voor de groene vragenlijsten zijn er gegevens van 6.550 personen in 3.041 huishoudens. Zonder toepassing van het begrenzingsalgoritme zijn er 12 negatieve gewichten. Het begrenzen van de gewichten verliep in één iteratie. De eindgewichten zijn het product van insluitgewichten en correctiegewichten. In figuur 2 is de verdeling van de gewichten na begrenzen te zien.



Figuur 4.5 Verdeling van gewichten bij groene vragenlijsten.

De correctiegewichten liggen tussen 0,064 en 3,88. Het eerste kwartiel is 0,72, de mediaan is 0,91 en het derde kwartiel is 1,18. De eindgewichten liggen tussen 150 en 9.039. Het eerste kwartiel is 1.667, de mediaan 2.127 en het derde kwartiel 2.756.

5 Respons

De netto respons van AVO 2007 bedraagt 62,9%van de uitgezette steekproef minus de kaderfouten en de administratieve non-respons (9892 adressen). Bij bezoek van 21,8% van de gevallen werd er geweigerd om een vragenlijst te afnemen. Bij 4,9% is er geen contact geweest met de onderzoeks persoon (OP) en bij 4,2% is er tijdens de veld-werkperiode geen mogelijkheid geweest om een afspraak te maken met de OP.

Van de 6790 huishoudens die de een volledig CAPI-vragenlijst heeft ingevuld, hebben 6579 huishoudens gezegd de papieren vragenlijsten in te vullen. 211 huishoudens hebben na het CAPI-interview aan de enquêtrice verteld niet meer te willen meewerken aan het onderzoek. Van 360 huishoudens zijn niet alle papieren vragenlijsten geretour-neerd.

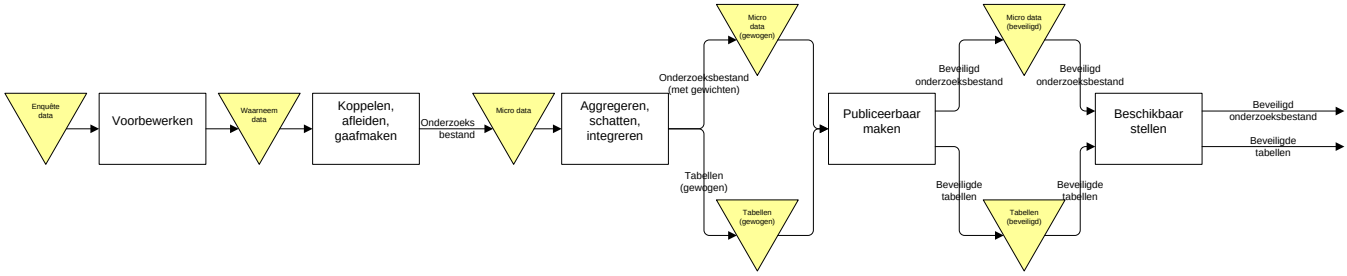
Responsoverzicht AVO totaal Veldwerk en Afboeken

	Aantal	%
Uitgezette steekproef	**10861**	
In aanbouw/afgebroken/sloopwoning	53	
Geen woonadres	65	
Leegstaand/onbewoond	74	
OP overleden	328	
OP verhuisd naar buitenland	0	
Niemand komt in aanmerking als OP	0	
OP onbekend	0	
Instelling/tehuis	0	
Kaderfouten	**520**	
Steekproef exclusief kaderfouten	**10341**	
Taalbarrière	328	
Onbewerkt retour	50	
Onvolledig bewerkt retour	93	
OP verhuisd in Nederland	0	
Anders (bv. laptop gestolen)	3	
Administratieve non-respons	**474**	
Extra huishoudens vanwege meervoudige bewoning	25	
Bezocht met correcte informatie	**9892**	**100%**
Weigering	2159	21,8%
Geen contact	487	4,9%
Geen gelegenheid tijdens veldwerkperiode	420	4,2%
Respons CAPI	6826	
Partieel	0	
Afgebroken	36	
Volledig	**6790**	
Werving voor vragenlijsten (PAPI is Ja)	**6579**	
Complete huishoudens (papi's volledig)	6219	**62,9%**
Incomplete huishoudens (papi's niet volledig)	360	3,6%
Werving voor vragenlijsten (PAPI is Nee)	211	2,1%

Staat 5.1 Responsoverzicht Veldwerk en afboeken

6 Verwerking

Het onderstaande figuur 6 geeft het verwerkingsproces van de onderzoeksdata in hoofdlijnen aan. De vijf deelprocessen (weergegeven in de vierkanten blokken) worden in hierna in detail beschreven.



Figuur 6: Procesmodel Verwerkingsproces

6.1 Voorbewerken: Van enquêtedata naar waarneemdata

Voordat enquêtedata geschikt is om te verwerken zijn er een aantal activiteiten nodig als voorbewerking. Kort gezegd gaat het erom de juiste data (qua cases en variabelen) in het juiste formaat beschikbaar te maken. Dit resulteert in waarneemdata. De waarneemdata is de grondstof voor het 'echte' verwerken.

Binnen het subproces "Voorbewerken" kunnen de volgende processtappen worden onderkend:

6.1.1 Controleren enquêtedata: range & routing controle

Met de range- en routingcontrole wordt gekeken of het enquêtebestand voldoet aan de eisen voor de verdere verwerking van het bestand.

Het enquêtebestand wordt hier gedefinieerd als:
- alle data uit een vragenlijst;
- van een of meerdere respondenten;
- van een specifieke mode.

De range- en routingcontrole is specifiek:
- per onderzoek;
- per mode;
- per versie van de vragenlijst.

De eisen van de controles zijn overwegend gebaseerd op de definitie van de vragenlijsten zoals vastgelegd in de gecompileerde vragenlijsten en op het ontwerp van het uniforme datamodel.

Range controles zijn o.a.:
- formaat van variabele;
- variabele heeft juiste naam;
- waardebereik van variabele conform afspraak;
- dubbele sleutels.

Routing controles zijn o.a.:
- als een blok niet op de route ligt, mogen de variabelen ook geen waarde hebben;
- als een blok wel op de route ligt moeten deze in principe (tenzij veld niet verplicht) ook een waarde hebben.

De volgende standaardcoderingen zijn gebruikt:

- 7 Niet van toepassing, vanwege een routing in de vragenlijst;

- 5 Niet van toepassing;

- 3 Weet niet;

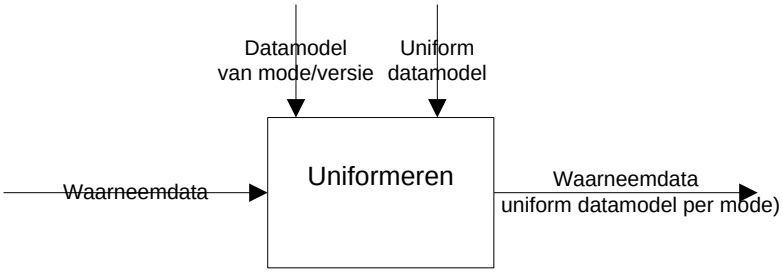
- 2 Weigert.

6.1.2 Uniformeren

De recordstructuur van enquêtedata kan mode specifiek zijn. Dit komt omdat de vragenlijst mode specifiek is. Daarmee is ook de metadata mode specifiek.
Het kan ook voorkomen dat gedurende het onderzoek er voor een mode meerdere versies van een vragenlijst worden gebruikt (eventueel ook met een verschil in variabelen). Dus ook per versie kunnen er verschillen qua recordstructuur zijn.
Doel bij uniformeren is om tot één uniforme recordstructuur te komen. De uniforme recordstructuur is onderzoekspecifiek en geldt in principe voor de duur van het hele onderzoek. Het bepalen van de uniforme recordstructuur[1] en het beschrijven van de bijbehorende metadata is een ontwerpactiviteit. Dit betekent dat de metadata van het uniforme datamodel vooraf gedefinieerd is; dus voordat het uniformeren daadwerkelijk plaats vindt. Bij het uniformeren van de data hoeft er in principe dus geen metadata meer te worden aangepast; als je de data in het uniforme model zet is de metadata automatisch correct.

Pre-conditie:

- het mode-specifieke (en vragenlijstversie specifieke) datamodel moet bekend zijn;
- het uniforme datamodel moet bekend zijn.

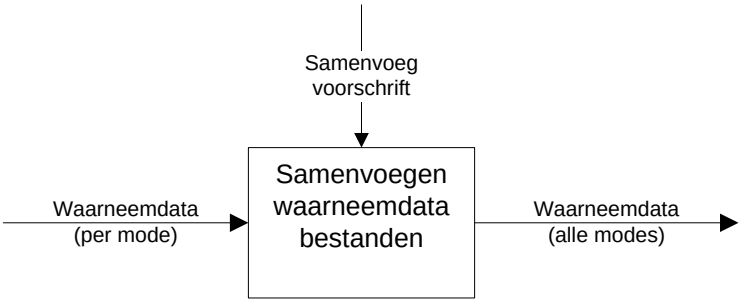


Figuur 6.1.2. : Procesmodel Uniformeren

[1] en trouwens ook de mode-specifieke datamodellen

6.1.3 Samenvoegen waarneemdatabestanden

De waarneemdatabestanden van de diverse modes worden samengevoegd tot één fysiek bestand. Per case moet wel duidelijk blijven uit welke mode het record komt.



Figuur 6.1.3. : Procesmodel Samenvoegen waarneemdata

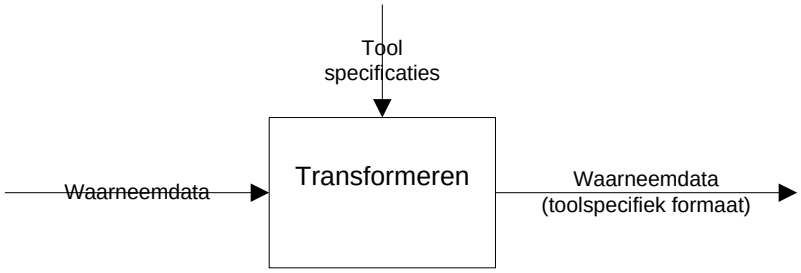
6.2 Koppelen, afleiden, gaafmaken

Bij het koppelen, afleiden en gaafmaken wordt de data verrijkt met o.a. data uit de steekproef, registerdata en andere bronnen. Tevens wordt de respons afgebakend, vindt imputeren en gaafmaken plaats en worden variabelen afgeleid (inclusief coderen). Dit resulteert in een onderzoeksbestand.

6.2.1 Transformeren

Doel van deze procesactiviteit is de waarneemdata transformeren naar een vorm die gebruikt kan worden bij de vervolg activiteiten. De waarneemdata is in een bepaald technisch formaat, in dit geval Blaise. Bij de vervolgstappen van het verwerkingsproces wordt SPSS als verwerkingstool gebruikt. Daartoe moet het technische formaat van de data worden aangepast (van Blaise via ASCII naar SPSS).

T.b.v. SPSS is transformeren bijvoorbeeld:
- Dichitomiseren (men werkt in het onderzoeksbestand niet met meervoudige antwoorden).
- Labels aanbrengen: Op basis van de metadata uit de vragenlijst worden variabelen en value labels gegenereerd, die worden gecombineerd met dit SPSS-systeembestand. Deze labels vormen de beschrijving van de data.
- Imputatie routing: Respondenten hoeven in de vragenlijst alleen die vragen te beantwoorden die op hun situatie van toepassing zijn. Vragen die door de respondent niet beantwoord hoefden te worden, gaan als blanco naar ASCII en krijgen vervolgens in SPSS de waarde SYSMIS. Dit betekent dat het SPSS-systeem de waarden als “n.v.t.” beschouwd bij de uitvoering van statistische analyses.



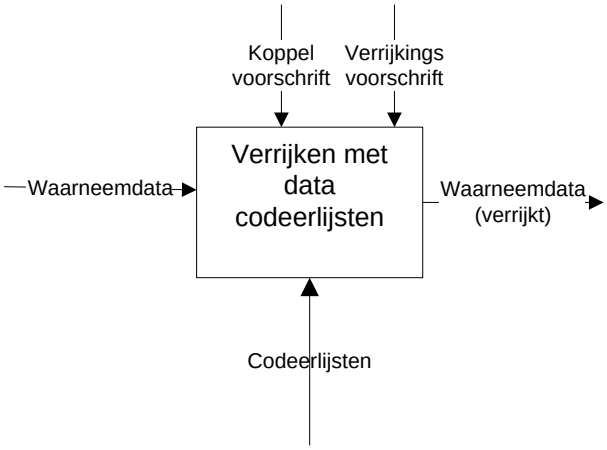
Figuur 6.2.1. : Procesmodel Transformeren

6.2.2 Verrinnen

Binnen de verwerkingsprocessen gebeurt “het verrinnen” alleen t.b.v. het koppelen met registerdata en dus niet om met geanonimiseerde data in het verwerkingsproces te werken. Om waarneemdata te kunnen verrijken met registerdata dient de waarneemdata eerst verrind te worden. Dit betekent dat voor iedere persoon in de waarneemdata een betekenisloze identificerende variabele wordt bepaald (genaamd “RINPersoon”). Dit nummer is gebaseerd op data uit de GBA.
De te koppelen waarneemdata wordt daartoe geleverd aan CBK. Deze koppelen de data aan het Centrale Koppelbestand Personen (CKP). CBK levert de verrinde data vervolgens terug aan het verwerkingsproces. Voor het leggen van een koppeling zijn het Burger Service Nummer (BSN) en/of de combinatie van geboortedatum, geslacht en adres nodig. Een geslaagde koppeling betekent in concreto dat aan het originele record RINPersoon en RINPersoonVolgNr uit het CKP worden toegevoegd. Hiermee is de desbetreffende persoon in het CKP identificeerbaar. Naast de CKP persoonidentificatie wordt nog informatie over de koppeling aan het record toegevoegd.

6.2.3 Verrijken met data codeerlijsten

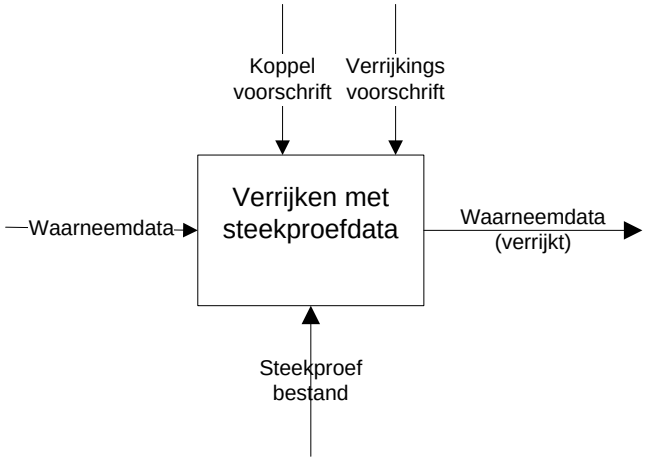
De waarneemdata wordt hier gekoppeld aan verschillende codeerlijsten. Hier wordt bijvoorbeeld op basis van de viercijferige postcode de gemeentecode bepaald. Op basis van de gemeentecode worden de bovengemeentelijke regionale indelingen bepaald.



Figuur 6.2.3. : Procesmodel Verrijken met data codeerlijsten

6.2.4 Verrijken met steekproefdata

T.b.v. non-respons analyses kan de waarneemdata worden verrijkt met de complete steekproef met voor elk element de voor uitdunning van de adressensteekproef gebruikte variabelen (alle adresgegevens bijvoorbeeld), het startgewicht en een eindresultaat (bijvoorbeeld: uitgedunde GBA65plus, niet uitgezet door regiomanager, geen woonadres, leegstand, niemand thuis, taalbarrière, weigering, enzovoorts). Op deze manier kan een betere (uitgebreidere) non-respons analyse naar allerlei achtergrondkenmerken gemaakt worden. Bij het koppelen kan het voorkomen dat er een steekproefeenheid is waarvoor (nog) geen waarneemdata is. En tevens waarneemdata waarvoor geen steekproefeenheid is. In het laatste geval is blijkbaar een verkeerde respondent bevraagd.



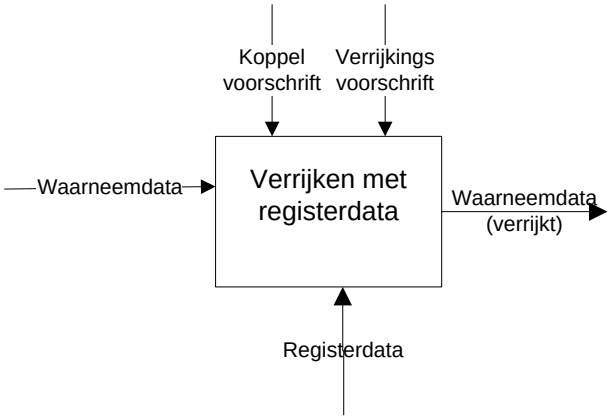
Figuur 6.2.4: Procesmodel Verrijken met steekproefdata

6.2.5 Verrijken met registerdata

Het koppelen met de registerdata gebeurt op basis van het RIN-nummer[2]. Veel gebruikte registers zijn de GBA, de Polisadministratie en het UWV-Werkbedrijf.
De waarneemdata wordt uit de registers verrijkt met o.a. type huishouden, geboorteland persoon en van diens vader en moeder en afleidingen daarop (GBA), bron inkomen en hoogte inkomen (Polisadministratie), provincie, inschrijfduur (is een afleiding), werkend (UWV-Werkbedrijf).

Het verrijken gebeurt niet alleen voor de OP maar ook voor alle vastgestelde personen in het huishouden. Als er een koppeling is, wordt de waarneemdata vervolgens verrijkt met een aantal variabelen uit de registers. Als er geen koppeling is blijven de betreffende register-variabelen leeg.

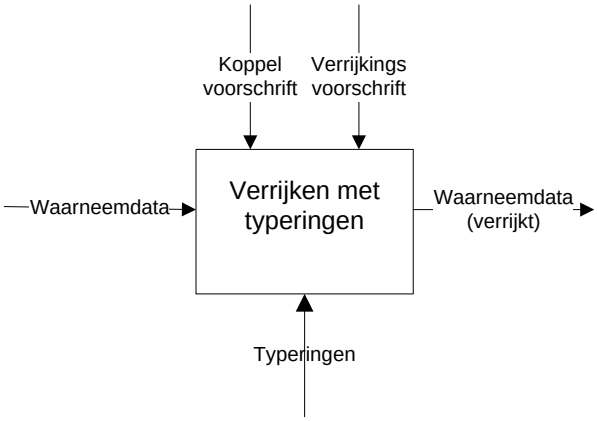
[2] Combinatie van RINPersoon en RINPersoonVolgNr



Figuur 6.2.5: Procesmodel Verrijken met registerdata

6.2.6 Verrijken met typeringen

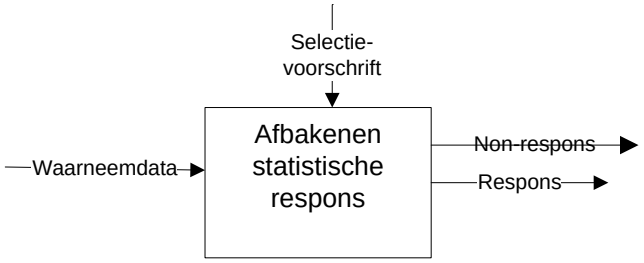
Deze stap kan pas worden uitgevoerd als de typeringen beschikbaar zijn. Afhankelijk van het specifieke onderzoek heeft het typeren een bepaalde doorlooptijd, waardoor deze data (meestal) niet meteen beschikbaar is.



Figuur 6.2.6: Procesmodel Verrijken met typeringen

6.2.7 Afbakenen statistische respons

Alleen statistische respons wordt meegenomen in de verdere verwerking. Wat wel/niet tot respons behoort, staat in een voorschrift. In deze activiteit wordt o.b.v. het voorschrift de respons bepaald. Hóe de non-respons gescheiden wordt van de respons is vanuit logisch oogpunt niet relevant; dit kan bijvoorbeeld door de records fysiek van elkaar te scheiden, maar kan bv ook door met indicatoren te werken. Voorwaarde is dat de non-respons op een gegeven moment beschikbaar is ten bate van non-respons analyses.

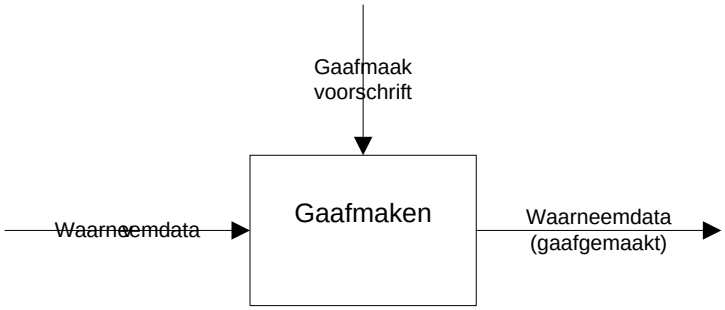


Figuur 6.2.7: Procesmodel Afbakenen statistische respons

6.2.8 Gaafmaken (micro)

Gaafmaken is het opsporen en corrigeren van foutieve gegevens in de waarneemdata.

Bij micro gaafmaken vinden zowel de controles als de correcties plaats op microniveau. Voorbeelden van voorkomende fouten zijn: het geboortejaar klopt niet of is onwaarschijnlijk, een respondent rapporteert in euro's in plaats van in duizenden euro's (of omgekeerd), of de winst van een bedrijf is niet gelijk aan het verschil tussen baten en lasten. Gaafmaken gebeurt op basis van, bij het ontwerp bepaalde, voorschriften.



Figuur 6.2.8: Procesmodel Gaafmaken (micro)

6.2.9 Imputeren

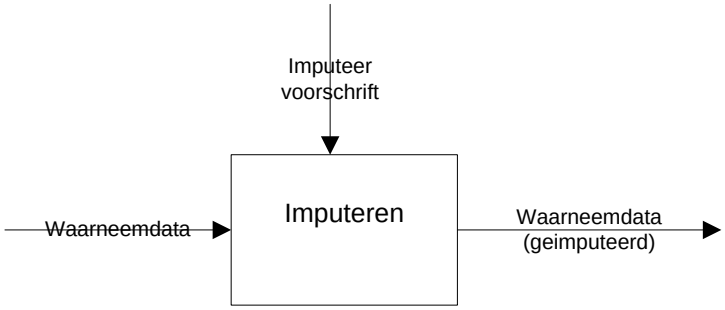
Imputeren is het bepalen en introduceren van een (nieuwe) waarde op een plaats waar een waarde ontbreekt of op ‘onbekend’ (ontbrekend) is gezet.

Bij enquêtes komt het voor dat respondenten op één of meer vragen geen antwoord geven, terwijl dit wel van ze wordt verlangd. Men spreekt dan van item-nonrespons (of partiële nonrespons) en van (ten onrechte) ontbrekende waarden (missing values). Redenen om een vraag niet te beantwoorden zijn het niet kunnen of willen geven van het antwoord. Op ingewikkelde of moeilijk te begrijpen vragen kan men vaak geen antwoord geven, op gevoelige vragen wil men het dikwijls niet. Ook bij registers kunnen gegevens ontbreken die het CBS wel had willen hebben.
Er zijn een aantal manieren om met ontbrekende waarden om te gaan. Eén daarvan is imputeren van een geldige waarde voor de ontbrekende waarde in het databestand.

Een alternatief voor imputeren is om het achterwege te laten. De ontbrekende waarden blijven dan onbekend. Redenen om te imputeren, in plaats van het veld leeg te laten, zijn:

1. het verkrijgen van een ‘volledig’ (geheel gevuld) databestand;
2. verhoging van de kwaliteit van het micro-bestand en/of van de parameterschattingen.

We maken verder onderscheid tussen imputeren en afleiden. Bij het afleiden van variabelen worden nieuwe variabelen gecreëerd als functie van in het bestand reeds bestaande variabelen. Bij imputeren worden ontbrekende waarden op een bestaande variabele gecreëerd. Imputeren gebeurt op basis van, bij het ontwerp bepaalde, voorschriften. Steeds dient te worden vastgelegd dat een waarde is geïmputeerd.



Figuur 6.2.9: Procesmodel Imputeren

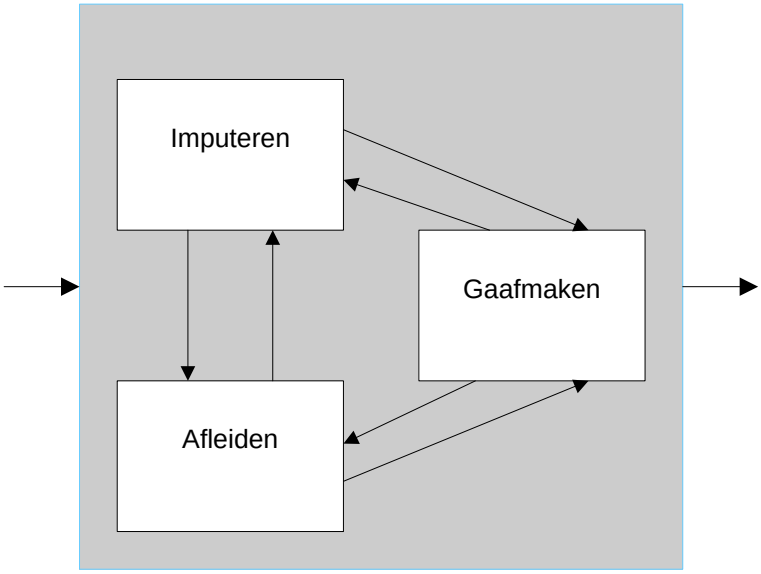
6.2.10 Afleiden

Met afleiden wordt hier bedoeld het creëren van nieuwe variabelen als functie van in het bestand reeds bestaande variabelen.

Coderen is ook een vorm van afleiden. Het coderen van een vraag is het (keuze)proces waarbij een beslissing wordt genomen om een antwoord te interpreteren in termen van een voorgedefinieerde verzameling mogelijke antwoorden. Een dergelijke keuze wordt, tijdens een interview of bij het invullen van een vragenformulier, vaak gedaan door respondenten, al of niet met hulp van een interviewer. Soms echter wordt deze keuze achteraf gedaan, op het CBS en zonder de aanwezigheid van een respondent, door codeurs of typeurs.

6.2.11 Gaafmaken, imputeren en afleiden: samenhang

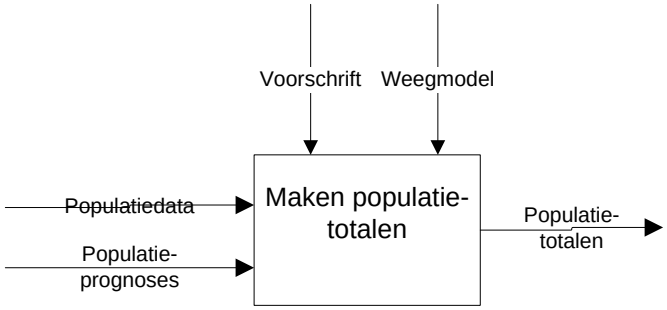
De logische volgorde is eerst gaafmaken, dan imputeren en dan variabelen afleiden. Echter, gaafmaken, imputeren en afleiden zijn geen activiteiten die voor een dataset en individuele case sequentieel verlopen. De activiteiten worden per variabele of set van variabele doorlopen, waarna de activiteiten voor andere variabelen worden doorlopen. Eventueel kan de volgorde van de activiteiten ook anders zijn: bv eerst imputeren dan gaafmaken. Dit is afhankelijk van de specifieke regels die van toepassing zijn binnen het onderzoek.



Figuur 6.2.11: Procesmodel Gaafmaken, imputere en afleiden: Samenhang

6.2.12 Maken populatietotalen

Voor het wegen zijn populatietotalen nodig. Het kan zijn dat de populatietotalen zelf geschat moeten worden aangezien de data bij de taakgroep Demografie niet altijd voldoende gedetailleerd en soms onvoldoende actueel zijn. Populatietotalen worden bepaald op het totale register, niet op de met register verrijkte waarneemdata.

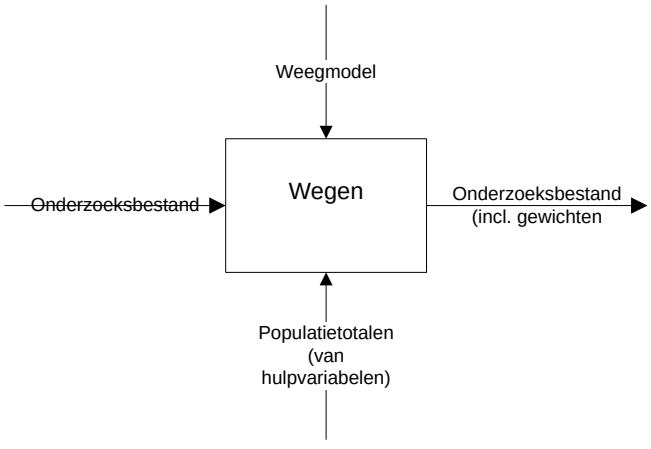


Figuur 6.2.12: Procesmode l makenpopulatietotalen

6.2.13 Wegen

Het wegen is de activiteit waarbij weegfactoren[3] worden bepaald. In het weegmodel staat beschreven hoe het wegen moet plaatsvinden.

Bij het wegen wordt de verdeling van variabelen in de steekproef in overeenstemming gebracht met de verdeling daarvan in de populatie. Daartoe wordt aan iedere case een gewicht toegekend. Ten bate van het wegen zijn populatietotalen (hulpvariabelen) nodig op persoonsniveau. Tevens kan gewogen worden naar verdelingen waaraan het responsproces idealiter moet voldoen, bijvoorbeeld: elke dag evenveel respons of een gelijk responspercentage per mode. De weging resulteert in één of meerdere ophoogfactoren (afhankelijk van het aantal entiteiten waarvoor gewogen wordt).



Figuur 6.2.13: Procesmodel Wegen

[3] Ook wel gewichten of ophoogfactoren genoemd

6.3 Publiceerbaar maken

Bij het publiceerbaar maken worden de tabellen en het gewogen onderzoeksbestand statistisch beveiligd. Binnen het subproces “publiceerbaar maken” kunnen de hieronder genoemde processtappen plaatsvinden.

6.3.1 Maken micro output

Op basis van het onderzoeksbestand wordt de output (microbestanden) voor bijvoorbeeld de externe klant, SAH-ADM/LIS, CvB, DANS en/of Eurostat gemaakt.
De output voor de diverse afnemers bevatten meestal slechts een deelverzameling van de variabelen uit het onderzoeksbestand. Deze deelverzamelingen voor de verschillende afnemers worden in deze stap gemaakt.
SAH-ADM/LIS krijgen meestal het gehele onderzoeksbestand ten behoeve van het maken van publicaties. Hiervoor hoeven geen extra stappen gezet te worden.

6.3.2 Statistisch beveiligen microdata

De microbestanden voor de verschillende afnemers worden vaak nog statistisch beveiligd. De wijze van beveiliging kan wel verschillen.

Onder statistische beveiliging verstaan we hier het voorkómen dat er inhoudelijke conclusies over herkenbare eenheden kunnen worden getrokken op basis van gepubliceerd of anderszins beschikbaar gesteld CBS-materiaal. Uit de statistische publicaties van het CBS (StatLine-tabellen, web-artikelen, persberichten, wetenschappelijke artikelen) mogen zulke conclusies niet getrokken kunnen worden. Maar ook als het CBS microdata beschikbaar stelt voor wetenschappelijke analyse, moet deze grondregel van de statistiek overeind blijven.

6.3.3 Statistisch beveiligen standaardtabellen

De tabellen (t.b.v. de Externe klant, Ministeries,Statline, CvB, DANS en/of Eurostat) dienen ook statistisch beveiligd te worden.

6.4 Beschikbaar stellen

De statistisch beveiligde producten worden vervolgens als output geleverd (beschikbaar gesteld) aan diverse belanghebbenden, waaronder externe opdrachtgevers, verschillende ministeries, Eurostat, DANS, Centrum voor Beleidsstatistiek. Binnen het subproces “Beschikbaar stellen” kunnen de hieronder genoemde processtappen plaatsvinden.

6.4.1 Leveren microbestanden

Dit is de activiteit waarbij microdata daadwerkelijk wordt geleverd. Aan welke partijen geleverd wordt is afhankelijk van het betreffende onderzoek.

Er worden voor het AVO 2 bestanden opgeleverd:

- Volwassenen splithalf Oud en Jeugd splithalf Oud worden geïntegreerd, hier wordt op persoonsniveau de CAPI-vragenlijst aan gekoppeld;
- Volwassenen splithalf Nieuw en Jeugd splithalf Nieuw worden geïntegreerd, hier wordt op persoonsniveau de CAPI-vragenlijst aan gekoppeld.

6.4.2 Leveren standaardtabellen

Dit is de activiteit waarbij de tabellen daadwerkelijk worden geleverd. Aan welke partijen geleverd wordt is afhankelijk van het betreffende onderzoek.

7 AVO-Specifiek ‘Koppelstappen’.

7.1 Capi-vragenlijst: Huishoudensbestand naar Personenbestand

Het AVO-onderzoek is een adressensteekproef waarbij de op het steekproefadres wonende huishouden ondervraagd is. In de eerste plaats gebeurde dit aan de hand van een persoonlijk interview (CAPI) met één van de huishoudleden (de onderzoekspersoon: OP). De OP geeft aan uit hoeveel personen het huishouden bestaat (tot maximaal 8 personen) en geeft vervolgens nog een aantal persoonskenmerken van deze huishoudleden. In eerste instantie staan de gegevens van alle personen in het huishouden in één rij. Om de koppeling met de later door een huishoudlid individueel (of proxy) ingevulde papieren vragenlijsten mogelijk te maken wordt er van dit ‘Huishoudensbestand’ een personenbestand gemaakt worden. De onderstaande regels maken van het huishoudensbestand een personenbestand.

```
VARSTOCASES
 /MAKE WE_ID          FROM WE_ID [hh]
 /MAKE ADRESIDE       FROM adreside [hh]
 /MAKE NRSCHRIF       FROM Nrschrif [pp 1..8]
 /MAKE NR_HH          FROM NR [pp 1..8]
 /MAKE AANTALPP       FROM AANTALPP [hh]
 /MAKE HHKERN         FROM HHKERN [hh]
 /MAKE REGIO_ID       FROM Regio_ID [hh]
 /MAKE VERMOGEN       FROM vermogen [hh]
 /MAKE NRNAVRG        FROM Nrnavrg  [hh]
 /MAKE EVP            FROM EVP [pp 1..8]
 /MAKE Kind           FROM Kind [pp 1..8]
 /MAKE Ander          FROM Ander [pp 1..8]
 /MAKE GESLACHT       FROM M_V [pp 1..8]
 /MAKE GEB_DAT        FROM Geboren [pp 1..8]
 /MAKE Burgst         FROM BurgSt [pp 1..8]
 /MAKE PLHH           FROM PLHH [1..8]
```

Staat 7.1: van huishoudensbestand naar personenbestand

7.2 Koppeling met papieren-vragenlijsten

De papieren vragenlijsten: het Navraagformulier (E-formulier), het Vermogensformulier (F-formulier), de Volwassenvragenlijsten (2 versies: V_A, V_B) en de Jeugdvragenlijsten (2 versies: J_C, J_D) zijn na het ophalen bij de respondenten vertoets via een Data-Entry-module.

Om een goede koppeling tot stand te brengen tussen ingetoetste vragenlijsten en de CAPI-vragenlijst is het van groot belang dat de ID_nummers van de papieren vragenlijsten, de ID van het Huishouden (We_ID) en de geboortedatum van de persoon in het huishouden juist wordt overgenomen van de vragenlijst en juist wordt ingetoetst. Echter, een foutje bij het vertoetsen is zo gemaakt en daarom wordt om een zo goed mogelijke koppeling van de papieren vragenlijsten met de CAPI-vragenlijst tot stand te brengen gekoppeld aan de hand van een combinatie van meerdere koppelsleutels in plaats van met één. Met hoe meer variabelen is gekoppeld hoe beter (‘sterker’) de koppeling is en hoe groter de kans dat er gekoppeld is met de juiste persoon. Na elke koppeling met een koppelsleutel worden de cases die matchen/koppelen opgeslagen

en ‘verwijderd' uit het werkbestand. Vervolgens dient het werkbestand, met records die niet gekoppeld zijn met de vorige koppelsleutel, als invoer voor de volgende sleutel-koppeling De volgorde van de koppeling is gebaseerd op koppelsterkte. Deze koppel-procedure begint bij de ‘sterkste’ koppelsleutel en eindigt bij de ‘slechtste’ koppelsleu-tel. De overgebleven dus niet gekoppelde vragenlijsten worden gecontroleerd op fouten en zo nodig hersteld en vervolgens weer aangeboden voor de koppeling. De boven omschreven werkwijze is toegepast op alle papieren-vragenlijsten.

7.3.1 E-Formulier (Navraag vragenlijst Autogebruik)

Voordat het koppelen met het E-formulier worden de variabelen gerenamed omdat anders de inhoud van originele variabelen (ingevuld in de Capi-vragenlijst) overschre-ven wordt. Ook is zo een onderscheid te maken welke variabelen afkomstig zijn van het E-formulier of de CAPI-vragenlijst.
De Naamwijziging van het E-vragenlijst vind op de volgende manier plaats: ‘E’ + varia-belenaam (Zie onder).

CAPI-NAAM	PAPI-naam	NIEUWE-naam
TypeAuto	A01c01	Etypauto
TypeAut2	A01c02	Etypaut2
TypeAut3	A01c03	Etypaut3
Jraansch	A01d01	Ejaara
Jraansc2	A01d02	Ejaara2
Jraansc3	A01d03	Ejaara3
Bouwjr	A01e01	Ebouwjr
Bouwjr2	A01e02	Ebouwjr2
Bouwjr3	A01e03	Ebouwjr3
Aankpr	A01f01	Eaankpr
Aankpr2	A01f02	Eaankpr2
Aankpr3	A01f03	Eaankpr3
TypeVerp	A01g01	Etypver
TypeVer2	A01g02	Etypver2
TypeVer3	A01g03	Etypver3
KmJr	A01h01	E_KmJr
KmJr2	A01h02	E_KmJr2
KmJr3	A01h03	E_KmJr3
Kmwk	A01i01	E_Kmwk
Kmwk2	A01i02	E_Kmwk2
Kmwk3	A01i03	E_Kmwk3
Naamreg	X_Enaamreg	Enaamreg
Naamreg2	X_Enaamrg2	Enaamrg2
Naamreg3	X_Enaamrg3	Enaamrg3
NrAndLid	X_EnrAlid1	EnrAlid1
NrAndLi2	X_EnrAlid2	EnrAlid2
NrAndLi3	X_EnrAlid3	EnrAlid3

Staat 7.3.1: renamen E-vragenlijst

7.3.2 Koppeling met E-vragenlijst

Omdat in de E-vragenlijsten de geboortedatums ingevuld moesten worden van de huishoudleden zijn de geboortedatums (van maximaal 3 huishoudleden) gebruikt om de koppeling tot stand te brengen.

De koppeling is gestart met koppelsleutel 1 en geboortedatum van huishoudlid 1 en geëindigd met koppelsleutel 4 en geboortedatum van huishoudlid 3. Het invoerbestand van deze koppeling is het CAPI-'personenbestand'.

De volgende 4 koppelsleutels zijn gebruikt:

Koppelsleutel_1 : NRNAVRG We_ID Geb_dat [1..3] Regio_id.
Koppelsleutel_2 : NRNAVRG We_ID Geb_dat [1..3].
Koppelsleutel_3 : NRNAVRG We_ID.
Koppelsleutel_4 : NRNAVRG Geb_dat [1..3].

NRNAVRG: ID_nummer E-vragenlijst (dit nummer werd ook in de CAPI ingevoerd)
We_ID: ID_nummer Huishouden
Geb_Dat [1..3]: Geboortedatum van persoon in huishouden [max 3 personen].
REGIO_ID: CBS-gebiedsindeling

7.3.3 Koppeling met F-vragenlijst (Navraag vragenlijst Vermogensvragenlijst)

De F-vragenlijst is een op een zelfde manier gekoppeld als de E-vragenlijst echter nu zijn de onderstaande koppelsleutels gebruikt en is gebruikt gemaakt van één geboortedatum. Dit is de geboortedatum van het huishoudlid die de E-vragenlijst heeft ingevuld. Het invoerbestand van deze koppeling is het CAPI-personenbestand gekoppeld met de F-vragenlijst.

De volgende 4 koppelsleutels zijn gebruikt:

Koppelsleutel_1 : VERMOGEN We_ID Geb_dat Regio_id.
Koppelsleutel_2 : VERMOGEN We_ID Geb_dat .
Koppelsleutel_3 : VERMOGEN We_ID.
Koppelsleutel_4 : VERMOGEN Geb_dat .

VERMOGEN: ID_nummer F-vragenlijst (dit nummer werd ook in de CAPI ingevoerd)
We_ID: ID_nummer Huishouden
Geb_Dat : Geboortedatum van persoon in huishouden [max 3 personen].
REGIO_ID: CBS-gebiedsindeling

7.3.4 Koppeling met Volwassen-vragenlijsten (V_A en V_B) en Jeugd-vragenlijsten (J_C en J_D)

De koppeling met de Volwassen-vragenlijsten en de Jeugd-vragenlijsten is op een zelfde manier gekoppeld als de E-vragenlijst en de F-vragenlijst echter nu zijn de onderstaande koppelsleutels gebruikt en is gebruikt gemaakt van één geboortedatum. Dit is de geboortedatum van het huishoudlid die één van de vragenlijsten heeft ingevuld. Het invoerbestand van deze koppeling is het CAPI-personenbestand gekoppeld met de F-vragenlijst en E-vragenlijst.

De volgende 4 koppelsleutels zijn gebruikt:

Koppelsleutel_1 : Nrschrif We_ID Geb_dat Regio_id Ref_dat
Koppelsleutel_2 : Nrschrif We_ID Geb_dat Regio_id
Koppelsleutel_3 : We_ID Geb_dat Regio_id
Koppelsleutel_4 : Nrschrif We_ID

Nrschrif: ID_nummer (V_A, V_B, J_C, J_D) vragenlijsten (dit nummer werd ook in de CAPI ingevoerd).
We_ID: ID_nummer Huishouden.
Geb_Dat : Geboortedatum van persoon in huishouden.
REGIO_ID: CBS-gebiedsindeling.
Ref_dat: Enquete-datum

Om de Volwassenvragenlijsten (V_A en V_B) te kunnen koppelen met de Jeugdvragenlijsten (J_C en J_D) moet er een 'rename'-slag plaatsvinden van de vragenlijsten. De volgende naamgeving van de variabelen is gebruikt:

Situatie	**Naamgeving**
Een vraag uit de Jeugdvragenlijst is identiek aan een vraag uit de Volwassenenvragenlijst.	De Jeugdvariabele krijgt de naam van de Volwassenenvariabele.
Een vraag uit de Volwassenenvragenlijst komt niet voor in de Jeugdvragenlijst.	De naam van de Volwassenenvariabele krijgt de toevoeging 'V'.
Een vraag uit de Jeugdvragenlijst komt niet voor in de Volwassenenvragenlijst.	De naam van de Jeugdvariabele krijgt de toevoeging 'J'.

7.3.5. Responsvariabelen.

Na de koppeling van alle papierenvragenlijsten aan het CAPI-personenbestand wordt met de volgende variabelen aangegeven of de respectievelijke de persoon (Kpapi) en of het hele Huishouden (hh_respons) heeft gerespondeerd.

Kpapi
- *0 '(vragenlijst) niet binnen/Geen koppeling (met vragenlijst mogelijk'*
- *1 'Leeg (=lege vragenlijst)'*
- *2 'Respons (= ingevulde vragenlijst gekoppeld)'*

HH_respons
- *-1 'Lege vragenlijsten'*
- *0 'Geen koppeling'*
- *2 'Volledige HHRespons'*

7.3.6. Statistische beveiliging

In de procedure waarin microdata statistisch wordt beveiligd zijn de volgende wijzigingen (onbekend codes) aangebracht om statistische onthulling te voorkomen.

	Onbekend door statistische beveiliging	
Variabelenaam	**V_A + J_D**	**V_B + J_D**
Geslacht	1	1
Plaatshh	7	3
Typehh	3	3
Aantalpp	1	1
Lft01	4	8
LftOp	4	8
Burgst	1	2
Aantpp_6Plus	1	0
Landd5	256	27
Herk3	15	16
Totaal	**293**	**312**

Staat 7.3.6. Onbekendcodes ten gevolge van statistische beveiliging naar bestand en variabelen

8 Referenties

- Van Berkel, C.A.M. (2008), Steekproefontwerp Aanvullend Voorzieningengebruik Onderzoek 2007, BPA nr. SOO-2008-H086, CBS, Heerlen.
- Van den Brakel, J.A. (2008), Design-based analysis of embedded experiments with applications in the Dutch Labour Force Survey. Journal of the Royal Statistical Society (A series) vol. 171, 3, in press.
- Huang, E.T. en W.A. Fuller (1978), Nonnegative regression estimation for survey data, in Proceedings of the Social Statistics Session, American Statistical Association, 1978, 300-303.
- Lemaître, G. en J. Dufour (1987), An integrated method for weighting persons and families, Survey Methodology 13, 199-207.
- Nieuwenbroek, N.J. en H.J. Boonstra (2002). Bascula 4.0 Reference Manual, BPA nr. 279-02-TMO, CBS, Heerlen.
- Cuppen M. en B. Janssen (2007). Onderzoeksdesign AVO 2007, BPA nr. SOO-2007-H0125, CBS, Heerlen, 30 mei 2007.