



Centraal Bureau voor de Statistiek

Centrum voor Beleidsstatistiek

**Documentatierapport Permanent Onderzoek
Leefsituatie basis (POLSbasis) 2009V1**

Datum: 27 december 2010

Bronvermelding

Publicatie van uitkomsten geschiedt door het onderzoeksbureau of de opdrachtgever op eigen titel. Verwijzing naar het CBS betreft uitsluitend het gebruik van de microdatabestanden. Dat wordt als volgt geformuleerd:

“Eigen berekening [*naam onderzoeksbureau, c.q. opdrachtgever*] op basis van bij het CBS beschikbaar gestelde bestanden betreffende het Permanent Onderzoek LeefSituatie basis (POLSbasis) 2009, welke afkomstig zijn van het CBS”

Versiegeschiedenis

Versie 1

101227 POLSbasis 2009V1

101227 GBA-selectiebestand POLSbasis 2009V1

De gebruiker dient rekening te houden met de volgende *do's & don'ts*:

- Dit microdatabestand kan gekoppeld worden aan de POLS module GEZOND (POLSmogEZOND2008V1) via de variabele ENCRYPT_WE_ID in combinatie met de variabele WPERIODE.
 - Uitsluitend de combinatie van RIN met SRTNUM 'R' of 'S' identificeert een persoon. Zie hoofdstuk 3 voor een beschrijving.
-

Inhoudsopgave

Bronvermelding.....	2
Versiegeschiedenis	3
1. Inleiding	8
2. Toelichting op de samenstelling van de bestanden.....	9
Microdatabestand.....	9
Doel onderzoek.....	9
Onderzoeksopzet	9
2 ^e en 3 ^e cyclus POLS	12
Opdrachtgever	12
Historie	13
Steekproef.....	14
Doelpopulatie	14
Steekproefkader.....	14
Steekproeftrekking	15
Steekproefomvang.....	17
Veldwerk	18
Screening en uitzet van de steekproef	18
Instructie	19
Respons	20
Evaluatie.....	21
Verwerking.....	21
Dataprocessing	21
Instroom van gegevens	22
Afleidingen.....	23
Regionale indelingen	23
Afgeleide variabelen.....	24
Ophoging.....	25
Methodiek.....	25
Hulpvariabelen en weegschema	26
Kwaliteit	29
Soorten fouten	29
Nauwkeurigheid van de uitkomsten	29
Betrouwbaarheid	30
PRAT	30
Oude vraagstelling en vraagstelling PRAT	31
Vraagstelling PRAT	31

GBA-selectiebestand	33
3. Bestandsopbouw en toelichting.....	35
Bestandsopbouw van het microdatabestand.....	35
Kwaliteitsaanduiding	35
Toelichting op de variabelen.....	42
SRTNUMOP, SRTNUM1-8, RINOP, RIN1-8	42
GEMCODE	42
Toelichting ISCED-variabelen	43
Voorkomende waarden van de variabelen.....	47

1. Inleiding

Bij het Centrum voor Beleidsstatistiek (CvB) worden microdatabestanden van het Centraal Bureau voor de Statistiek (CBS) beschikbaar gesteld voor gebruik door externe onderzoekers. Dit documentatierapport beschrijft de inhoud en structuur van het microdatabestand

- ‘101227 POLSbasis 2009V1’.

Van alle personen die in het hier beschreven microdatabestand met SRTNUM ‘R’ voorkomen, zijn gegevens uit de Gemeentelijke basisadministratie (GBA) geselecteerd. Deze selectie kunt u vinden in het bestand:

- ‘101227 GBA-selectiebestand POLSbasis 2009V1’

Het complete GBA-bestand is in een apart [documentatierapport](#) beschreven. Voor Remote Access en On-site gebruikers is het documentatierapport over de GBA te vinden op \2-Metadata\GBA.

Hoofdstuk 2 beschrijft in het algemeen de microdatabestanden zoals populatieafbakening, methodologische bijzonderheden, kwaliteit en herkomst van de gegevens. In hoofdstuk 3 wordt de bestandsindeling per microdatabestand gegeven. Daarnaast wordt per bestand per variabele meer informatie gegeven. Voor de categoriale variabelen worden alle mogelijke scores en hun betekenis opgesomd.

Bij het rapport zijn de volgende bijlagen beschikbaar:

- ‘101227 Bijlage 1 Frequentietellingen POLSbasis 2009V1’.

In deze bijlage worden van de (semi-) continue variabelen de frequenties van de eerste en laatste vijf waarden gegeven. Met deze frequentietellingen kan tevens een beeld verkregen worden over de vullingsgraad van de verschillende variabelen, wat van belang kan zijn omdat niet altijd alle variabelen (volledig) gevuld zijn.

- ‘101227 Bijlage 2 Beschrijving variabelen POLSbasis 2008V1’

In deze bijlage worden de variabelen beschreven en de voorkomen de waarden van de variabelen weergegeven.

2. Toelichting op de samenstelling van de bestanden

Microdatabestand

Een record in het microdatabestand geeft informatie over demografische en sociaaleconomische kenmerken voor de gehele POLSsteekproef.

Doel onderzoek

Het doel van het Permanent Onderzoek Leefsituatie (POLS) is - kort samengevat - het verstrekken van een platform om op permanente basis hoogwaardige en coherente informatie over de leefsituatie van de Nederlandse bevolking te verzamelen.

Daarbij is het onderzoek zo opgezet dat adequaat kan worden ingespeeld op de informatiebehoefte van gebruikers en klanten.

Waar informatie over de leefsituatie vóór 1997 uit verschillende afzonderlijke onderzoeken naar delen van de leefsituatie afkomstig was, komt deze informatie vanaf 1997 nagenoeg geheel uit POLS. Doordat POLS een geïntegreerd onderzoek is waarin de oude leefsituatieonderzoeken een plaats hebben gevonden, levert POLS een meer samenhangend beeld van de leefsituatie en wordt meer recht gedaan aan de ééncijfergedachte die luidt dat er slechts één cijfer over één onderwerp wordt gepubliceerd. De maximale afstemming van uitkomsten van leefsituatieonderzoeken ten aanzien van demografische en sociaal-economische kenmerken van huishoudens en van personen in huishoudens is in POLS bewerkstelligd door uit te gaan van één identieke basisvragenlijst voor alle respondenten. Door de hele massa van respondenten die deze basisvragenlijst heeft beantwoord als één grote steekproef te beschouwen wordt slechts één uitkomst, één cijfer, over een onderwerp uit de basisvragenlijst gegenereerd.

Behalve vragen naar de demografische en sociaal-economische kenmerken bevat de basisvragenlijst ook vragen naar de meest belangrijke aspecten van de leefsituatie, kernindicatoren genoemd. Ook voor deze kernindicatoren wordt de ééncijfergedachte waargemaakt.

Voor bepaalde variabelen die in POLS worden gemeten zijn andere CBS- of externe bronnen dan POLS meer geschikt als het gaat om schattingen voor of verdelingen in de populatie. Zo is b.v. de Enquête Beroepsbevolking (EBB) de officiële bron voor een weergave van het onderwijsniveau van de populatie.

Onderzoeksopzet

POLS is een survey onder personen waarbij informatie verzameld wordt die voor de gehele bevolking van toepassing wordt geacht. Voor de onderzoeksopzet van het integratieve onderzoek POLS is een schillenmodel gekozen, waarbij de vragen van de “afzonderlijke” onderzoeken zijn verdeeld over vier zogenaamde schillen. Hiervan

worden de vragen uit schil 1 en 2 gesteld aan alle respondenten die voor de schillen 3, de thematische vragenlijstmodules, worden benaderd.

Schil 1 omvat de gecoördineerde CBS-vragen naar demografische en sociaal-economische kenmerken voor de gehele POLS-steekproef. In beginsel heeft de steekproef betrekking op de gehele bevolking. De demografische kenmerken betreffen leeftijd, geslacht, burgerlijke staat, samenstelling huishouden, nationaliteit en woongemeente. De sociaal-economische kenmerken betreffen opleiding en sociaal-economische positie (onder meer beroep en bedrijfsklasse). De inkomensvariabele uit het IPS wordt eveneens gerekend tot schil 1. Deze metingen blijven in principe als een vast en ongewijzigd onderdeel permanent in POLS meelopen. Aanpassingen worden alleen doorgevoerd als het CBS de definities en/of meetinstrumenten van deze standaardconcepten wijzigt.

Deze variabelen in schil 1 zijn in principe bedoeld als achtergrondvariabelen om de samenhang met andere variabelen te kunnen bepalen en relevante groepen in de bevolking te kunnen onderscheiden. Als het gaat om puntschattingen van de betreffende variabelen of om verdelingen daarvan zijn andere bronnen vaak beter geschikt. Dit laatste geldt overigens ook voor bepaalde variabelen waarbij niet andere bronnen maar andere variabelen (met andere vraagstellingen) binnen POLS tot betere schatters leiden.

Schil 2 bevat voor de gehele POLS-steekproef de kernindicatoren van de verschillende leefsituatie-onderwerpen. Het betreft een vaste set vragen over wonen, gezondheid, leefstijlen, kwaliteit van de arbeid, politieke betrokkenheid, milieu, maatschappelijke participatie, tijdsbesteding, slachtofferschap (van criminaliteit) en onveiligheidsgevoelens. Het gaat om een serie vragen over de verschillende onderwerpen waar het CBS met een zekere regelmaat informatie over wil verzamelen. Vanwege het feit dat de steekproefomvang in de tweede schil groot is kunnen kernvariabelen worden gebruikt voor bij voorbeeld kwartaalcijfers of voor regionale verdieping. Daarnaast is er voorzien in ruimte voor variabele vragenblokken, die moeten worden gemeten met een grote steekproef en voor (in principe) variabele screeningsvragen. Met de screeningsvragen kunnen voor de gehele steekproefmassa die personen worden geselecteerd waarop een bepaalde eigenschap van toepassing is. In een later stadium (schil 4) krijgt de aldus geselecteerde specifieke categorie bovenop het reguliere programma nog additionele vragen voorgelegd. Vanwege de variabele mogelijkheden is de samenstelling van de reeks schil 2-vragen minder stabiel dan die van schil 1. Binnen POLS geldt de richtlijn dat de gezamenlijke interviewduur van de schillen 1 en 2 (gemiddeld) de 15 minuten niet mag overschrijden.

Schil 3 bevat de diverse onderwerpgerichte modules. De gemiddelde interviewduur van de modules bedraagt ongeveer 30 minuten. Derden kunnen in POLS participeren via het principe van horizontale modulaire uitbreiding. Dit houdt simpelweg in dat de gehele POLS-steekproef wordt uitgebreid met het gewenste aantal personen van de nieuwe inhoudelijke module en in schil 2, indien mogelijk, de nieuwe kernindicatoren worden toegevoegd. Ook is het mogelijk dat de steekproefomvang van een bestaande module op verzoek van derden wordt vergroot. Door de wisselende frequenties van de onderzoeken en de fluctuerende betrokkenheid van derden zal het aantal onderwerpgerichte modules van jaar op jaar variëren.

Het beschreven design biedt grotere mogelijkheden bij een geringere steekproef. Grotere mogelijkheden door de onderwerpen die voorheen in afzonderlijke onderzoeken werden verzameld, nu op het niveau van de respondent al gelijktijdig waar te nemen. Hiermee wordt tegemoetgekomen aan de vraag naar gecombineerde informatie over verschillende onderwerpen bij subgroepen van de populatie. Een geringere steekproef vanwege de mogelijkheid vanuit de basisenquête via geavanceerde ophoog- en wegingstechnieken betrouwbare uitspraken over deelonderwerpen bij een kleinere steekproef te verzamelen. Het is evident dat de vaak als belastend voor de respondent ervaren enquêtedruk hierdoor ook kan verminderen.

Met ingang van 2005 is de onderzoeksopzet gehandhaafd hoewel het aantal modules teruggebracht is tot één : de gezondheidsmodule. Voor 2007 ziet het design er als volgt uit (schema 1):

Schema 1. Onderzoeksopzet POLS 2007

SCHIL 1	
BASISVRAGENLIJST	* gecoördineerde vragen m.b.t. demografische en sociaal-economische kenmerken (vast)
CAPI	

SCHIL 2	
BASISVRAGENLIJST	* vragen over kernindicatoren leefsituatie (vast) * lichamelijke- en psychische belemmeringen (variabel)
CAPI	

SCHIL 3	
ONDERWERPGERICHTE MODULES	
GEZONDHEID	
GEZO	
CAPI	
PAPI	

Afkortingen:

CAPI Computer assisted personal interviewing / Computergestuurd face-to-face interview

CATI Computer assisted telephone interviewing / Computergestuurd telefonisch interview

PAPI Paper and pencil interview / Interview met papieren vragenlijst

GEZO Gezondheid en Arbeid

Verdere bijzonderheden:

- Bij de module GEZO behoort nog een schriftelijke vragenlijst.

2^e en 3^e cyclus POLS

Om methodologische redenen vinden bepaalde onderdelen van de enquête plaats door middel van zelfrapportage. Dit gebeurt zowel computergestuurd (de respondent vult op de computer van de interviewer een vragenlijst in) als schriftelijk (de vragenlijst wordt achtergelaten en moet na invulling door de respondent naar het CBS worden opgestuurd).

Opdrachtgever

Opdrachtgever voor het Permanent Onderzoek Leefsituatie is het Centraal Bureau voor de Statistiek zelf, waarbij de Centrale Commissie van de Statistiek het werkprogramma waaruit de opdracht voortvloeit vaststelt. Door het CBS zelf moeten in elk geval nagenoeg alle aspecten van de leefsituatie kunnen worden beschreven in het kader van de statistische informatie over de leefsituatie. Deze aspecten worden voor het merendeel behandeld in de basisvragenlijst en in de vragenlijst van de module GEZO. Deze module

is permanent, van jaar op jaar, in het veld. Bij eventuele overige modules is dit afhankelijk van de wensen van de andere opdrachtgever(s).

Daar waar gebruikers een belangrijke detaillering van de informatie over de leefsituatie wensen, in vraagstelling of in niveau, kunnen zij het CBS verzoeken bestaande vraagstellingen uit te breiden of aan te passen, nieuwe vragenblokken of modules te introduceren en/of de steekproef te verhogen.

Historie

In 1971 werd er in Nederland voor het laatst een volkstelling gehouden. Een van de redenen van stopzetting was dat dit instrument niet meer kon voorzien in de toegenomen maatschappelijke behoefte aan sociaal-culturele en economische gegevens. Om de informatievoorziening op sociaal-cultureel terrein toch sluitend te kunnen dekken, is het CBS medio jaren zeventig van start gegaan met de uitvoering van een reeks sociaal-culturele surveys. Aanvankelijk kon men hiermee goed inspelen op de groeiende vraag naar (trend)gegevens over leefsituatie-indicatoren. Tevens ontwikkelden zich in het begin van de jaren tachtig nieuwe, afzonderlijke CBS-enquêtes rondom onderwerpen als wonen, gezondheid, tijdsbesteding en rechtsbescherming en veiligheid. Deze onderzoeken bestreken samen een groot deel van de leefsituatie, maar waren onderling niet goed op elkaar afgestemd. Behalve ten aanzien van specifieke leefsituatie-aspecten verschilden de onderzoeken soms ook ten aanzien van de te beschrijven populaties (vaak de Nederlandse bevolking als totaal, maar soms doelgroepen daaruit: jongeren, ouderen). In een aantal opzichten kwamen de onderzoeken echter met elkaar overeen. Zo werden in alle onderzoeken een aantal dezelfde achtergrondkenmerken over personen en huishoudens verzameld. Ook in de bewerking van de gegevens en de samenstelling van de bestanden bestonden veel overeenkomsten.

Maatschappelijke ontwikkelingen hebben geleid tot een verandering in de behoefte aan informatie: meer vragen over welvaart- en welzijnsverschillen tussen verschillende groepen in de samenleving, meer behoefte aan aansluiting bij nationale statistieken en aan gedetailleerde regionale informatie, meer vragen over specifieke groepen in de samenleving (ouderen, allochtonen, gehandicapten, bewoners van achterstandswijken) en een grotere behoefte aan informatie over combinaties van leefsituatieaspecten (meerdimensionale informatie).

Het beëindigen van de onderzoeksmatige versnippering en het bewerkstelligen van een algehele kwaliteitsverbetering vormden de stuwende krachten voor de herinrichting van de bestaande CBS-leefsituatie-onderzoeken. Het streven naar meer kwaliteit richtte zich onder meer op de verhoging van de betrouwbaarheid en de validiteit van de cijfers, de vergroting van de mogelijkheden tot detaillering en de uitbanning van meervoudige, niet identieke metingen van hetzelfde kenmerk. Dit aan de hand van een tegelijkertijd flexibel en uniform onderzoeksontwerp waarmee reguliere monitoring van de leefsituatie mogelijk is en gericht geanticipeerd kan worden op de (wisselende) maatschappelijke vraag naar leefsituatiegegevens. Door aansluiting te zoeken bij de nieuwste technologieën op het terrein van dataverzameling, controle, consistentietoetsing, typering en

documentatie moest bovendien worden voorzien in een verbetering van de interne bedrijfsprocessen rondom de verwerking van persoonsinformatie.

In 1995 werd besloten de leefsituatie-onderzoeken en de processen daarvan te herinrichten. De fundamentele integratie van de verschillende onderzoeken moest plaatsvinden door het ontwerp van één onderzoek waarin door middel van modules aandacht aan specifieke onderdelen van de leefsituatie wordt gegeven. Deze integratie moest op den duur leiden tot een aanzienlijke efficiencyverhoging bij de werkzaamheden die liggen op het terrein van dataverzameling, controle en consistentie, gaafmaken en publicaties over de basisinformatie. Deze integratie van de afzonderlijke leefsituatieonderzoeken is een eerste stap in de richting van een verdergaande herinrichting van de door het CBS verzamelde persoonsinformatie.

Aan het herontwerp van de persoonsenquêtes lag een aantal doelstellingen ten grondslag:

- het terugdringen van de totale kosten van verzameling- en bewerkingswerkzaamheden;
- verhoging van de effectiviteit van vraaggesprekken (het zoeken naar een evenwicht tussen respons, gespreksduur en interviewkosten);
- verlaging van de enquêtedruk door vermindering van het aantal enquêtes en verhoging van de respons;
- verbetering van het steekproefontwerp;
- maximale afstemming van de uitkomsten van onderzoeken ten aanzien van de demografische en sociaal-economische kenmerken van huishoudens;
- betere aansluiting bij de informatiebehoefte van klanten, grotere flexibiliteit;
- verhoging van de permanente financiële participatie van ministeries en planbureaus in de onderzoekskosten;
- verbetering van de geldigheid en de betrouwbaarheid van de gegevens;
- het op kwartaalbasis en op regionale basis beschikbaar stellen van relevante kerngegevens over al de aspecten van de leefsituatie van de Nederlandse bevolking.

Steekproef

Doelpopulatie

De doelpopulatie van POLS is de in Nederland woonachtige bevolking in particuliere huishoudens. Bij sommige deelonderzoeken zijn er leeftijdsgrenzen aan de doelpopulatie gesteld.

Steekproefkader

Het steekproefkader dat gebruikt wordt is afgeleid uit de structuurtelling afkomstig uit de Gemeentelijke Basis Administratie (GBA). Dit bestand wordt binnen het CBS beheerd door de Sector Registers Personen (SRP). Elke Nederlander heeft binnen dit bestand zijn

eigen nummer, het zogenaamde A-nummer, bestaande uit 10 cijfers. Verder zijn in dit bestand een aantal persoonsgegevens, namelijk de geboortedatum, de burgerlijke staat, het geslacht, het geboorteland, de nationaliteit en de gemeente waarin de persoon woont opgenomen. Het GBA-bestand bevat evenveel records als er inwoners van Nederland zijn. Om de omvang van het steekproefkader enigszins binnen de perken te houden is eerst een zogenaamde ‘moedersteekproef’ uit dit bestand getrokken. Dit is gedaan door alle personen te selecteren waarvoor het zevende cijfer gelijk is aan een van tevoren bepaald cijfer. Er wordt hierbij verondersteld dat het zevende cijfer in het A-nummer aselekt is en niet een bepaalde betekenis heeft. Om er voor te zorgen dat elk jaar een andere groep personen deel uit maakt van het kader is het cijfer waarmee vergeleken wordt steeds het laatste cijfer van het jaartal, dus 6 voor 2006 en 7 voor 2007.

Aan elk record wordt bij het genereren van het kader een random getal toegekend dat wordt gebruikt bij het trekken van de tweede trap van de steekproef waarin de steekproefpersonen worden getrokken. In de eerste trap worden gemeenten geselecteerd.

Het steekproefkader wordt elk jaar ververs. Ongeveer negen maanden na het begin van een jaar is de GBA voor dat jaar gereed. Dit betekent dat de steekproeven voor 2007 tot en met ongeveer september uit het kader van 2006 worden getrokken. Het kader wordt wel elke maand aangepast met behulp van nieuw binnengekomen informatie betreffende geboorte, sterfte en verhuizing tussen gemeenten.

Het gebruik van de GBA als steekproefkader houdt een verandering in ten opzichte van de werkwijze in eerdere jaren, toen voornamelijk adressensteekproeven werden getrokken. In POLS is om drie redenen gekozen voor de GBA:

1. Gebruik van de GBA maakt het mogelijk om eenvoudig vragenlijsten uit te zetten onder specifieke groepen, onderscheiden naar de in de GBA beschikbare kenmerken. Bij een adressensteekproef moeten deze groepen eerst worden opgespoord (screening).
2. Gebruik van de GBA houdt in dat de non-respondenten onder de steekproefpersonen beter dan voorheen kunnen worden beschreven omdat vanuit de GBA enkele achtergrondkenmerken van deze personen bekend zijn.
3. Gebruik van de GBA betekent dat de veldwerkdienst van het CBS, Sector Data Verzameling (SDV) vooraf kan beschikken over deze demografische kenmerken en daarvan bij de benadering van de steekproefpersonen gebruik kan maken, bij voorbeeld door de tijden van het eerste bezoek aan te passen, waardoor de trefkans van de steekproefpersoon en de kans op een respons groter is.

Steekproeftrekking

De steekproeftrekking vindt plaats in twee trappen: in de eerste trap worden gemeenten geselecteerd, in de tweede trap worden personen getrokken.

Trap 1: trekking van gemeenten

In deze fase van het ontwerpen van de steekproef worden gemeenten getrokken en wordt bepaald hoeveel personen per gemeente geënquêteerd moeten worden. Bovendien wordt

het totaal aantal personen dat getrokken moet worden in een gemeente verdeeld over de verschillende deelonderzoeken. De eerste trap van de steekproef is voor alle deelonderzoeken gelijk en kan ruim voor de start van het veldwerk plaatsvinden.

Bij de trekking van de gemeenten wordt de populatie verdeeld in 40 COROP-gebieden (strata). Uit elk COROP-gebied worden in de eerste trap een of meer gemeenten getrokken. De trekkingskans voor een gemeente is evenredig aan de omvang van die gemeente.

Van tevoren wordt het gewenste minimum aantal te enquêteren personen in een gemeente bepaald. De keuze van dit aantal hangt af van de grootte van het onderzoeksbudget en de capaciteit en de opleidingsmogelijkheden van het enquêtecorps. Dit aantal wordt in één keer vastgesteld voor alle strata en bedraagt voor POLS 18 personen per gemeente. Het uiteindelijke minimum aantal te enquêteren personen in een gemeente is meestal niet precies gelijk aan 18, maar schommelt rond deze waarde.

De berekening van de aantallen te trekken gemeenten per stratum en personen per gemeente verloopt als volgt. Eerst worden de zelfselecterende gemeenten vastgesteld. Dit zijn de grotere gemeenten die met kans 1 worden getrokken en waarvoor de steekproefomvang rechtstreeks wordt afgeleid uit het aantal inwoners. Daarna wordt het aantal te trekken niet-zelfselecterende gemeenten en het aantal te trekken inwoners binnen de geselecteerde gemeenten bepaald. Dit is een itererend proces dat stopt zodra er geen nieuwe zelfselecterende gemeenten bijkomen.

Nu de gemeenten getrokken zijn en de aantallen te trekken personen per gemeente bekend zijn, kan de steekproef verdeeld worden over de deelonderzoeken. Er wordt hier rekening gehouden met mogelijke uitval. Verwacht wordt dat van 1,6 procent van de personen in de steekproef geen adres wordt teruggeleverd door de gemeenten en dat 14,4 procent uitvalt bij de sector SDV door verwijdering van adressen die recent getrokken zijn voor andere onderzoeken. Er wordt getrokken uit disjuncte, niet-overlappende kaders voor adressen- en persoonssteekproeven.

Om te corrigeren voor de verwachte uitval van in totaal 16 procent wordt de steekproefomvang per gemeente vermenigvuldigd met een factor 1,16.

Per gemeente wordt nu het aantal te trekken personen per onderzoek bepaald door de eerder genoemde factoren, te vermenigvuldigen met het berekende aantal per gemeente. Het totale aantal te trekken personen wordt vermenigvuldigd met de desbetreffende factor. Dit levert niet-gehele aantallen personen op. Deze worden zodanig afgerond dat per onderzoek de juiste totale aantallen personen resulteren. De som van de factoren is gelijk aan 1,16.

Trap 2: trekking van personen

De uitkomst van trap 1 in het steekproefdesign geeft voor elk van de deelonderzoeken aan hoeveel personen er per gemeente getrokken moeten worden. In trap 2 worden personen getrokken uit het steekproefkader. Zoals bij de beschrijving van de populatie al is opgemerkt, worden voor de deelonderzoeken leeftijdsgrenzen in acht genomen. De trekking uit het kader verloopt aselekt. Het kader is gesorteerd op gemeente en binnen de

gemeente op het random getal dat aan elk van de records is toegekend bij het samenstellen van het kader. Dit gebeurt elke maand ongeveer 8 weken voor de start van het veldwerk.

Bij het trekken van personen in een gemeente wordt voor het eerste deelonderzoek begonnen bij de persoon die in het steekproefkader volgt op de laatste persoon die voor de vorige maand is getrokken. Er wordt nagegaan of die persoon voldoet aan de leeftijdseisen van het deelonderzoek. Als dat zo is, valt deze in de steekproef en anders niet. Vervolgens wordt de volgende persoon in het kader onder de loep genomen, net zolang tot er zoveel personen getrokken zijn als het design voorschrijft. Voor het volgende deelonderzoek wordt er begonnen bij de persoon die daarna volgt in het steekproefkader en wordt verder dezelfde procedure gevolgd. Dit wordt herhaald voor alle deelonderzoeken.

Na het trekken van de personen is een bestand ontstaan met daarin het A-nummer van de persoon en persoonsgegevens als geboortedatum, geslacht, burgerlijke staat, geboorteland en nationaliteit. Voor het uitvoeren van het onderzoek is het natuurlijk nodig om ook de beschikking te hebben over adresgegevens en namen van de getrokken personen. Deze worden aangevraagd bij de gemeenten. De teruggeleverde adressen met de namen worden gekoppeld aan het steekproefbestand. Uit het resulterende bestand worden de A-nummers verwijderd en dit aangepaste bestand gaat naar SDV. In deze fase treedt uitval op doordat records niet voorzien kunnen worden van de gewenste adresgegevens, bij voorbeeld bij een verhuizing of door problemen in de (automatische) communicatie met gemeenten. Voor de uitval door verhuizingen wordt in het design gecompenseerd, voor de uitval door niet-responderende gemeenten niet.

Steekproefomvang

De steekproefomvang voor 2007 is bepaald op basis van een totaal gewenst aantal respondenten van 10 000. Op basis van de aanvankelijke responsverwachting en de verwachte uitval bij de steekproefcontrole is de bruto steekproefomvang bepaald en de steekproef getrokken (zie ook hoofdstuk 3). Zoals uit overzicht 1 blijkt heeft geen volledige uitzet van de steekproef plaatsgevonden. Dit vloeit voort uit een tekort aan interviewers. Ongeveer 23 procent van de steekproefadressen die voor het POLS zijn getrokken kon niet worden uitgezet. Mede daardoor is de gewenste respons niet gehaald.

Overzicht 1. Gewenste respons, steekproef en uitzet van POLS 2007

Onderzoek	Gewenste respons	Steekproef	Uitzet	in % van steekproef
BASIS/GEZO	10 000	17 855	13 660	76,5

Veldwerk

Het veldwerk van POLS wordt uitgevoerd met behulp van computergestuurde vragenlijsten die met de op het CBS ontwikkelde programmataal Blaise zijn gebouwd. De veldinterviewers hebben deze vragenlijsten via hun modem in hun laptop geladen en ontvangen via dezelfde weg hun steekproefadressen.

Screening en uitzet van de steekproef

De steekproefadressen worden vóór uitzet nog op drie punten gescreend:

1. Reeds eerder benaderde adressen
2. Meervoudige bewoning
3. Institutionele adressen

Ad 1. Een uitgangspunt bij de benaderingsstrategie van het CBS is dat een adres maximaal één keer per twee jaar mag worden benaderd. Daartoe worden de adressen, dus ook die voor POLS, gescreend om doublures met andere steekproeven te elimineren. Zolang de adressen met gelijke kansen zijn getrokken (wat doorgaans het geval is) wordt de representativiteit van de steekproef door de verwijdering niet aangetast. Wel neemt uiteraard de omvang van de steekproef af. Hierdoor moet de bruto steekproef met ongeveer 1,6% worden verhoogd. De adressen van POLS worden gescreend met de (circa 30.000 maandelijkse) adressen van de steekproef voor de Enquête Beroepsbevolking.

Ad 2. Vanwege beperkingen van het enquête-administratiesysteem kan slechts één persoon per adres worden geïnterviewd. Wanneer de steekproef twee of meer personen op hetzelfde adres bevat, wordt willekeurig één persoon gehandhaafd. De andere perso(o)n(en) op dit adres worden verwijderd. De uitval tengevolge hiervan bedraagt ongeveer 0,2%.

Ad 3. Adressen die voorkomen in het register van inrichtingen, instellingen en tehuizen (IIT) worden uit de steekproef verwijderd. De doelpopulatie betreft immers personen in particuliere huishoudens. De uitval door IIT-adressen komt neer op circa 1,2%.

De adressen worden via het systeem voor enquêteadministratie (SEA) van SDV toebedeeld aan de in principe inzetbare interviewers: de fictieve uitzet. Deze wordt definitief gemaakt zodra is vastgesteld dat alle ingeplande interviewers daadwerkelijk beschikbaar zijn. In voorkomende gevallen (ziekte e.d.) kan de fictieve uitzet worden gemuteerd.

Het bleek problematisch het interviewercorps op sterkte te brengen. De reductie van de steekproef vanwege het genoemde capaciteitstekort bij het interviewercorps (uitdunning) is zodanig ingericht dat voor POLS per interviewregio de steekproef is uitgedund naar gelang de beschikbare capaciteit.

Omdat het tekort aan interviewers niet evenredig over het land gespreid is zit er een zekere regionale selectiviteit in de uitzet en dus ook in de respons.

Instructie

Voor de verzameling van de gegevens van POLS zijn in 2007 gemiddeld ca. 400 door het CBS opgeleide interviewers ingeschakeld. Alvorens ingezet te worden hebben deze interviewers een vijfdaagse basistraining doorlopen, waarin aandacht wordt besteed aan algemene achtergrondinformatie, interviewregels en laptoptraining. In deze algemene basistraining is de training voor POLS, die ongeveer de helft van de trainingstijd beslaat, geïntegreerd.

Voor de interviewers is een handleiding voor POLS samengesteld waarin wijzigingen en uitbreidingen eenvoudig kunnen worden toegevoegd. Naast de schriftelijke handleiding hebben alle interviewers voor POLS ook een mondelinge instructie ontvangen. Deze mondelinge instructies voor de interviewers van POLS bevatten algemene informatie over het onderzoek en het veldwerk en specifieke informatie over de modules. In principe kan iedere interviewer elke module van POLS afnemen.

Bij de algemene informatie wordt ingegaan op onder andere de volgende punten:

- Het CBS-beleid om de enquêtedruk te verlagen en de mogelijkheden die het POLS-design daartoe biedt
- Het feit dat voor POLS een personensteekproef wordt getrokken waarbij alleen de getrokken persoon mag worden ondervraagd: proxy-interviews zijn niet toegestaan! Hierop zijn twee uitzonderingen: bij kinderen jonger dan 12 jaar moet het interview worden afgenomen met een ouder of verzorger van het kind en enkele vragen voor de hoofdkostwinner over de situatie van het huishouden mogen eveneens door iemand anders worden beantwoord.
- Het gebruik van de antwoordkaarten bij de modules. In de basisvragenlijst zijn geen vragen opgenomen waarvoor antwoordkaarten gebruikt moeten worden omdat deze, vóór 2007, ook telefonisch werden afgenomen
- De afhandeling van de declaraties via het Enquête Declaratie Systeem
- De verdeling van de modules over de interviewers op basis van een selectie die vooraf is bepaald. Na de basisvragenlijst wordt automatisch 'ingeplugd' op een vervolgmodule; deze module is voor de start van het interview bekend bij de interviewer maar nog niet bij de respondent
- De veldwerkperiode bedraagt een maand
- Het gebruik van de demografische informatie over de steekproefpersoon bij de benaderingen; bijvoorbeeld door ouderen zoveel mogelijk overdag te benaderen
- Het opnieuw benaderen van aanvankelijke weigeraars
- De procedure bij verhuisde steekproefpersonen
- De regelmatig (minimaal twee maal per week) te plegen datacommunicatie met het CBS zodat de adressen zo snel mogelijk op het CBS afgehandeld kunnen worden

Respons

Voordat een persoon in de steekproef wordt benaderd ontvangt hij/zij een op naam gestelde aanschrijfbrief. In deze brief wordt niet gespecificeerd voor welke module de persoon is geselecteerd. Dit is besloten naar aanleiding van uitkomsten uit een proef waaruit blijkt dat er een selectie optreedt bij aankondiging van het onderwerp van het onderzoek. Die selectie bestaat bijvoorbeeld uit een grotere deelname van personen met een zwakkere gezondheid aan de module GEZO en van slachtoffers aan de modules REM en REP. Andere personen (in het voorbeeld gezonde mensen en personen die geen slachtoffer van criminaliteit zijn geworden) zouden sneller aannemen dat het onderwerp niet voor hen bedoeld is omdat zij geen bijzonderheden of voorvallen kunnen melden.

Zoals reeds uit overzicht 1 bleek heeft er geen volledige uitzet van de steekproeven plaatsgehad. In overzicht 2 wordt de verhouding tussen de steekproef en respons weergegeven.

Overzicht 2. Gewenste en gerealiseerde respons van POLS 2007

Onderzoek	Gewenste respons	Bruikbare respons		Uitzet van Steekproef	Respons / uitzet (%)	
		basis	module		basis	module
BASIS/GEZO	10 000	8 741	8 741	13 660	64,0	64,0

Het gaat hier om de respons zoals die uiteindelijk kan worden afgeleid uit de in de bestanden aanwezige records, in de tabel bruikbare respons genoemd. Voor de bepaling van de interviewerrespons en de daarmee gemoeide betalingen aan interviewers hanteert SDV andere criteria dan SAH, waardoor de respons voor SDV meestal iets hoger is. Bij SDV gaat het om de invulling van een bepaalde variabele aan het begin van de vragenlijst (is het gesprek gestart), bij SAH gaat het om een zodanige invulling van de vragenlijst dat het record bruikbaar is. Daarom wordt bij SAH meer naar de vulling van variabelen aan het einde van de vragenlijst gekeken. Als de respondent daar is aanbeland is er een grote kans dat het record goed is ingevuld.

Overzicht 3. Responspercentages module, Pols 2001 - 2007

	Responspercentage module		2003	2004	2005	2006	2007
	2001	2002					
BASIS/GEZO	62	61	58	61	65	66	64

In de loop der jaren zijn er forse verschillen opgetreden in de responspercentages van de modules (zie overzicht 3). De respons in overzicht 3 wordt berekend door de bruikbare

respons te relateren aan de uitgezette steekproef (zie overzicht 2). Volgens deze methode van responsberekening is de respons in 2001 en 2002 ten opzichte van 2000 met ongeveer 5 procent toegenomen. In 2007 is het responspercentage 64%.

Overzicht 3a. Realisatiepercentages ten opzichte van netto target (100), POLS 2001-2007

Onderzoek	Realisatie						
	2001	2002	2003	2004	2005	2006	2007
BASIS/GEZO	97	97	99	111	104	96	87

In 2001 werd door het CBS een reorganisatie van de interviewsector doorgevoerd, met als doel een hogere respons te bewerkstelligen. Uit overzicht 3a blijkt dat de realisatie van de bruikbare respons van het gehele POLS in 2007 nagenoeg niet voldoet aan de doelstelling.

De veldwerksector heeft de mogelijkheid de uitzet over de maanden te laten variëren van 80 tot 120 procent om op die manier de in te zetten enquêt capaciteit per maand te optimaliseren (in het steekproefdesign vastgestelde aantal = 100). Deze uitzet is, op jaarbasis, niet in alle regio's evenredig uitgezet. Vanwege een regionaal gebonden tekort aan interviewer capaciteit varieert de uitzet van de steekproef per regio ongeveer tussen de 15 en 40 procent. In de weging wordt hiervoor middels berekende insluitgewichten gecorrigeerd.

Evaluatie

Interviewers hebben de mogelijkheid om tijdens of na afloop van een interview elektronisch een opmerking te maken. Deze opmerkingen worden samen met de interviewgegevens naar het CBS verzonden. Daarnaast kunnen interviewers telefonisch of schriftelijk contact opnemen met de enquêtebegeleiders van het CBS. Aan de hand van deze reacties wordt een indruk verkregen van de problemen die interviewers in het veld ervaren. In enkele gevallen vormen deze opmerkingen aanleiding om de interviewers te voorzien van additionele instructies.

In 2007 zijn geen evaluatiebijeenkomsten met interviewers gehouden en zijn er ook geen aparte schriftelijke of telefonische evaluaties geweest.

Verwerking

Dataprocessing

Vanaf 2002 is er voor de dataprocessing van POLS een nieuw dataprocessingsysteem ontwikkeld dat het traject vanaf de ontvangst van gegevens (elektronische data afkomstig van CAPI- en CATI-vragenlijsten enerzijds en ingevulde papieren vragenlijsten anderzijds) tot aan de beschikbaarstelling van de bewerkte gegevens voor statistische analyse omvat. Het gaat hierbij om alle handelingen op de ontvangen gegevens vanaf de

instroom in het systeem die een algemeen karakter hebben en dus niet zijn toegesneden op het vormen van een bepaald statistisch product voor een bepaalde klant. Bewerkingen als specifieke afleidingen, typeren, imputeren en beveiligen vallen buiten dit proces. Kortom, dataprocessing is het geheel van bewerkingen op de ruwe data dat leidt tot het ontstaan van een gemeenschappelijke, statistische basis van waaruit specifieke, klantgerichte producten worden gevormd.

Door deze scheiding tussen algemene en specifieke bewerkingen aan te brengen krijgt het dataprocessingtraject een belangrijke eigenschap: *reproduceerbaarheid*. Dat wil zeggen dat op ieder moment alle bewerkingen die na de instroom tot aan de vorming van de statistische basis hebben plaatsgevonden op exact dezelfde wijze herhaald kunnen worden. Hierdoor wordt vooral het proces van het oplossen van fouten betrouwbaarder. Een fout in de data wordt niet tussentijds verholpen maar wordt juist in de oorspronkelijk ontvangen data verbeterd. Vervolgens wordt het hele verwerkingsproces voor de betreffende interviewperiodes opnieuw uitgevoerd. Aldus veroorzaakt het achtereenvolgens corrigeren van fouten niet een in complexiteit toenemend verwerkingstraject. Dezelfde werkwijze wordt ook toegepast bij het aan het licht komen van fouten in een verwerkingsstap.

Het dataprocessingtraject bestaat uit twee onderdelen:

1. de instroom van de gegevens;
2. het bewerken van de gegevens;

De steekprofeenheden van POLS worden periodiek (maandelijks) voor een veldwerkperiode van een maand uitgezet.

Er worden drie soorten gegevens bijgehouden:

- statistische data: interviewdata, afleidingen, ophoogfactoren, etc., per interviewperiode;
- responsdata: gegevens over responderende steekprofeenheden, per veldwerkperiode;
- non-respons data: gegevens over niet-responderende steekprofeenheden, per veldwerkperiode.

Instroom van gegevens

Algemeen geldt tijdens de instroom dat als een record niet verwerkt kan worden, alle desbetreffende invoerrecords van de steekprofeenheid gekopieerd worden naar de overeenkomstige foutbestanden en dus niet in de uitvoerbestanden terechtkomen.

De overige criteria die bij de instroom gehanteerd worden zijn:

- responsafbakening voor de vaste basis (schil 1);
- responsafbakening voor de variabele basis (schil 2);
- responsafbakening voor de onderwerpschil (schil 3 en 4);
- controle op dubbele sleutels.

Indien er een bijbehorende papieren vragenlijst is dan moet synchronisatie plaatsvinden met de PAPI-instroom. Het instroomprogramma gaat met behulp van het nummer van de papieren vragenlijst, die bij het CAPI-onderzoek door de interviewer wordt geregistreerd, na of de steekproefeenheid al in het koppelbestand aanwezig is. Voor instroom van de papieren vragenlijsten moeten de bijbehorende CAPI-gegevens namelijk al zijn ingestroomd. De voorbereiding van de PAPI-instroom bestaat uit het intoetsen of inlezen van de gegevens die op papier zijn ingevuld. Is sprake van inlezen dan worden de gegevens alvorens te kunnen instromen, omgevormd naar het vereiste blaise-formaat.

Gegevensbewerking

Tot de gegevensbewerkingen behoren de volgende onderdelen:

- Het maken van regionale indelingen (op basis van postcode en gemeentecode van het responsadres)
- Het maken van afleidingen
- Het samenvoegen van maandbestanden tot periodebestanden
- Het typeren van bedrijf, beroep en onderwijs (PRAT zie pag. 30)
- Het aanmaken van ophoogfactoren
- Het koppelen van gegevens uit de Inkomensstatistiek 2006
- Het aanmaken van analysebestanden voor de verschillende inhoudelijke thema's
- Het aanmaken van een beveiligd DANS-bestand
- Het produceren van tabellen t.b.v. plausibiliteitscontroles
- Het documenteren van het proces

Afleidingen

In de bewerkingsfase van het dataprocessingtraject worden afgeleide variabelen aan elk record toegevoegd. Deze afgeleide variabelen zijn te onderscheiden in twee hoofdgroepen: enerzijds de variabelen die worden afgeleid van de adresgegevens en die dus onafhankelijk van de antwoorden van de respondent worden gegenereerd en anderzijds de afleidingen waarvoor de antwoorden van de respondent op vragen in de vragenlijst bepalend zijn voor de waarde op de afgeleide variabele.

Regionale indelingen

De regionale indelingen worden direct afgeleid van de, eventueel gecorrigeerde, adresinformatie van de respondent. Voor de meeste regionale (en categoriale) indelingen is de gemeentecode voldoende om te kunnen bepalen in welke gebiedseenheid de respondent woont, voor enkele regionale indelingen is de postcodeaanduiding noodzakelijk.

De volgende indelingen worden afgeleid:

- gemeentegrootte

- stedelijkheid
- landsdelen
- provincie
- COROP-gebieden
- politieregio's
- rechtsgebieden (arrondissementsrechtbanken en gerechtshoven)
- toeristengebieden

Afgeleide variabelen

De afgeleide variabelen waarvoor de antwoorden van de respondent nodig zijn worden ofwel in de bewerkingsfase van het dataprocessingtraject aangemaakt, ofwel ze zijn al in een eerdere fase gedraaid voordat de bestanden het dataprocessingtraject ingaan. In het eerste geval betreft het afleidingen die niet gecoördineerd zijn voor het gehele CBS, in het laatste geval gaat het om voor het gehele CBS gecoördineerde afleidingen. De afgeleide variabelen worden behalve voor de ondervraagde persoon ook voor de kostwinner aangemaakt. Tot de afgeleide variabelen behoren, op het gebied van huishoudens:

- plaats huishouden
- samenstelling huishouden (kort en uitgebreid)
- huishoudensituatie
- basisindeling bevolkingsgroepen
- aantal personen in huishouden
- totaal aantal personen van 18 jaar of ouder in huishouden
- leeftijden van kinderen
- diverse leeftijdsaanduidingen (in jaren en in groepen)
- burgerlijke staat

op het gebied van inkomen, beroep en opleiding:

- inkomensbron (verkorte en standaard indeling)
- inkomen van het huishouden (zie opm. over IPO)
- maatschappelijke participatie
- arbeidspositie
- dagelijkse bezigheden
- sociale groep 1997
- Ultee-Sixma huidig beroep (1992-Schaal naar beroepsprestige, huidig beroep OP)
- EGP huidig beroep (Indeling naar sociale klasse, huidig beroep OP)

- opleidingsniveau (voltooid en gevolgd/volgend)

Ophoging

Methodiek

Hoewel er in POLS verschillende maatregelen zijn getroffen ter verbetering van de non-respons is daarmee het risico van selectiviteit die voortkomt uit de non-respons niet uitgebannen. In POLS worden daarom individuele wegingcoëfficiënten berekend. Eenvoudig gesteld is de berekening hiervan gebaseerd op het bewerkstelligen van volledige correspondentie tussen de verdelingen van (combinaties van) steekproefkenmerken en de overeenkomstige bekende verdelingen in de populatie. Het is de verwachting dat de weging ook eventuele vertekeningen in andere steekproefkenmerken rechtrekt dan wel op zijn minst verbetert.

Hoewel de kern van iedere weging stoelt op dit basisprincipe is de feitelijke weging van POLS iets ingewikkelder. POLS is te beschouwen als een doorlopende enquête waarvoor maandelijks een steekproef wordt samengesteld. Eerst wordt er voor iedere inhoudelijke module een enkelvoudige aselechte steekproef getrokken uit het GBA. Vervolgens wordt de basissteekproef gevormd door alle onafhankelijk van elkaar getrokken steekproeven van de inhoudelijke modules samen te voegen. Daarmee ontstaat er een moeilijkheid. De basissteekproef dient representatief te zijn voor de bevolking van 0 jaar en ouder. Het leeftijdsbereik verschilt evenwel tussen de modules. Zo hadden voorheen de modules REP en REM alleen betrekking op personen van 12 jaar en ouder, terwijl bijvoorbeeld de module GEZO wel de gehele doelpopulatie omvatte. Vanwege deze verschillen zijn in de steekproef sommige leeftijdsgroepen over- en andere ondervertegenwoordigd. Om de gegevens van de basissteekproef evenwichtig te maken voor de gehele bevolking, worden er voorafgaande aan de weging leeftijdsspecifieke correctiegewichten bepaald. Daarnaast is er voor een juiste berekening van de insluitgewichten een correctie nodig om de ongelijke trekkingskansen tussen de veldwerkregio's, die ontstaan door de uitdunning van de steekproef in de regio's, te elimineren.

Voordelen van het POLS-design voor de uitkomsten zijn de mogelijkheden om:

- schattingen van overeenkomstige variabelen uit verschillende enquêtes eenvoudig consistent te maken; en
- door middel van de schattingsprocedure voor een aantal variabelen, in het bijzonder voor de doelvariabelen uit de vervolgenquêtes, een aanzienlijke variantiereductie te behalen.

De schattingsmethodiek voor POLS komt overeen met het toepassen van de aangepaste algemene regressieschatter. De regressieschatter maakt gebruik van hulpinformatie. Allereerst wordt de gehele (door correctiegewichten feitelijk reeds voorgewogen) basissteekproef consistent gemaakt met de uit registraties bekende populatieverdelingen van geslacht, leeftijd, burgerlijke staat, stedelijkheidsgraad en provincie. Dat levert voor de gehele basissteekproef individuele wegingcoëfficiënten op die ervoor zorgen dat de

steekproefverdelingen op voornoemde kenmerken nagenoeg identiek zijn aan die in de populatie.

De variantie van deze schatter kan benaderd worden door uit te gaan van een steekproef met teruglegging. Omdat de basissteekproef klein is ten opzichte van de doelpopulatie is dat een acceptabele veronderstelling.

Het gebruik van de algemene regressieschatter kan eveneens vertekeningen ten gevolge van non-respons verkleinen. Bovendien wordt door het gebruik van hulpvariabelen consistentie bereikt met de bekende populatietotalen. Dit wil zeggen dat de uit de steekproef geschatte totalen van de hulpvariabelen gelijk zijn aan de bekende populatietotalen.

Evenals bij de basissteekproef kan bij het ophogen van een vervolgstekproef reductie van variantie behaald worden indien gebruik gemaakt wordt van hulpvariabelen die correleren met de doelvariabelen van de vervolgenquête. In de basisenquête van POLS zijn vragen (kernvariabelen) opgenomen die gericht zijn op specifieke onderwerpen van de vervolgenquêtes. Deze kernvariabelen zullen vaak in hoge mate correleren met de doelvariabelen van de vervolgenquêtes. De bevolkingstotalen van deze variabelen zijn in eerste instantie onbekend, maar kunnen na ophoging van de basissteekproef geschat worden. Vervolgens kunnen zij gebruikt worden in de schattingsprocedure van de vervolgstekproeven.

Naast hulpvariabelen waarvan de bevolkingstotalen geschat worden kan bij de ophoging van een vervolgstekproef uiteraard ook gebruik gemaakt worden van hulpvariabelen waarvan bevolkingstotalen via andere bron bekend zijn. In de praktijk is het vaak gewenst om consistentie te bereiken met die bekende bevolkingstotalen. Bij het gebruik van twee of meer vervolgstekproeven (of doorsnedes van vervolgstekproeven met verschillende doelpopulaties) zal naar verwachting sprake zijn van een hogere variatiereductie voor het te schatten totaal dan bij het gebruik van slechts één vervolgstekproef. Bij een combinatie van steekproeven voor het schatten van een totaal gebruikt men immers meer beschikbare hulpinformatie (in de vorm van meer respondenten).

Naast de persoonsgewichten worden in POLS ook gewichten opgenomen die kunnen worden gebruikt om gegevens op huishoudensniveau te kunnen vaststellen. Het huishoudensgewicht wordt verkregen door het persoonsgewicht te delen door het aantal personen in het huishouden. Vervolgens wordt het berekende gewicht op nationaal niveau nog gecorrigeerd voor de huishoudens naar grootte uit de huishoudensstatistiek.

Hulpvariabelen en weegschema

Met POLS wordt beoogd jaarcijfers te geven over de verschillende aspecten van de leefsituatie.

In deze documentatie wordt het schema voor de jaarophoging van de basis vermeld.

Wat de notatie van de weegschema's betreft betekent een plusteken (+) dat de verdeling van de betrokken variabele(n) in de ophoging wordt meegenomen, maar dat - in tegenstelling tot bij een vermenigvuldigingsteken (×) - deze aansluiting met de

populatie-totaal alleen geldt voor deze variabele(n) afzonderlijk en niet voor de kruising van deze variabele(n) met andere variabelen.

Voor de jaarophoging van het vaste deel van de *basissteekproef* in 2007 is het volgende weegmodel gehanteerd:

$(\text{Geslacht} \times \text{Leeftijd-2}) + \text{Burgerlijke staat-2} + \text{Stedelijkheidsgraad} + \text{Provincie-plus} + \text{Huishoudgrootte} + (\text{Geslacht} \times \text{Leeftijd-1} \times \text{Burgerlijke staat-1}) + (\text{Landsdeel} \times \text{Leeftijd-1}) + \text{Enquêtemaand}$

Hierbij zijn de volgende indelingen van de variabelen gebruikt:

- *Geslacht (twee categorieën)*

Man

vrouw

- *Leeftijd-1 (drie categorieën)*

0–34 jaar

35–54 jaar

55 jaar en ouder

- *Leeftijd-2 (zeventien categorieën)*

0 - 3 jaar	20 - 24 jaar	45 - 49 jaar	70 – 74 jaar
4 - 11 jaar	25 - 29 jaar	50 - 54 jaar	75 jaar e.o.
12 - 14 jaar	30 - 34 jaar	55 – 59 jaar	
15 - 17 jaar	35 - 39 jaar	60 – 64 jaar	
18 - 19 jaar	40 - 44 jaar	65 – 69 jaar	

- *Burgerlijke staat-1 (twee categorieën)*

gehuwd

ongehuwd

- *Burgerlijke staat-2 (vier categorieën)*

gehuwd

gescheiden

verweduwd

nooit gehuwd geweest

- *Landsdeel (vier categorieën)*

Noord (Groningen, Friesland, Drenthe)

Oost (Overijssel, Gelderland, Flevoland)

West (Utrecht, Noord-Holland, Zuid-Holland, Zeeland)

Zuid (Noord-Brabant, Limburg)

- *Provincie-plus (zestien categorieën)*

Groningen	Zeeland
Friesland	Noord-Brabant
Drenthe	Limburg
Overijssel	Flevoland
Gelderland	Amsterdam
Utrecht, exclusief de stad Utrecht	Rotterdam
Noord-Holland, exclusief Amsterdam	Den Haag
Zuid-Holland, exclusief Rotterdam en Den Haag	Utrecht

- *Stedelijkheidsgraad (vijf categorieën)*

zeer sterk stedelijk
 sterk stedelijk
 matig stedelijk
 weinig stedelijk
 niet stedelijk

- *Huishoudgrootte (vijf categorieën)*

éénpersoonshuishouden
 tweepersoonshuishouden
 huishouden bestaande uit drie personen
 huishouden bestaande uit vier personen
 huishouden bestaande uit vijf of meer personen

- *Enquête maand (twaalf categorieën)*

Januari	Mei	September
Februari	Juni	Oktober
Maart	Juli	November
April	Augustus	December

Kwaliteit

Soorten fouten

Uitkomsten van enquêtes op steekproefbasis komen in het algemeen niet overeen met de werkelijke waarden van de kenmerken in de bevolking. De afwijking (de totale fout) kan zowel het gevolg zijn van steekproeffouten als van niet-steekproeffouten.

Steekproeffouten ontstaan doordat bij slechts een gedeelte van de elementen in de populatie waarnemingen worden verricht. Deze fouten treden niet op als alle elementen in de populatie worden ondervraagd. Steekproeffouten worden onderscheiden in:

- *trekkingsfouten*: deze treden op als de kans op trekking van een element uit de doelpopulatie niet overeenkomt met de in het steekproefontwerp gespecificeerde trekkingskansen, bij voorbeeld als in het steekproefkader elementen zijn opgenomen die niet tot de doelpopulatie behoren.
- *schattingsfouten*: de afwijkingen tussen de met behulp van de schatter in de steekproef gevonden waarde en de waarde die in een integrale telling verkregen zou zijn.

Niet-steekproeffouten hebben betrekking op dat gedeelte van de totale fout dat ook zou zijn opgetreden indien alle elementen in de doelpopulatie worden ondervraagd.

Niet-steekproeffouten zijn te verdelen in:

- *waarnemingsfouten*: fouten die in het veld worden gemaakt bij het verzamelen en het vastleggen van de gegevens:
 - *overdekkingfouten*: in het steekproefkader zitten elementen die niet tot de doelpopulatie behoren;
 - *meetfouten*: de interviewer vult per abuis een verkeerde code in;
 - *verwerkingsfouten*: de fouten die worden gemaakt in het bewerkingstraject van controleren, corrigeren en afleiden.
- *niet-waarnemingsfouten*: fouten die ontstaan doordat de gewenste metingen in het geheel niet kunnen worden verricht:
 - *onderdekkingfouten*: de doelgroep is niet in het steekproefkader opgenomen;
 - *non-respons*: bij de steekproefpersoon is, om welke reden dan ook, geen geslaagd interview afgenomen.

Nauwkeurigheid van de uitkomsten

De kwaliteit van de uitkomsten wordt bepaald door het veldwerk, de steekproefopzet, de schattingsmethode en door het verwerkingsproces.

De nauwkeurigheid van steekproefonderzoek heeft twee aspecten: precisie en zuiverheid. De precisie geeft de mate aan waarin de verdeling van de schatter geconcentreerd ligt om de verwachtingswaarde ervan. Met de verwachtingswaarde wordt het gemiddelde bedoeld van alle mogelijke uitkomsten die de schatter kan aannemen bij herhaald toepassen van

de steekproefprocedure. Is de verwachtingswaarde gelijk aan de waarde van de te onderzoeken grootte dan wordt de schatter zuiver genoemd. Variaties in de mogelijke uitkomsten worden dan uitsluitend toegeschreven aan fluctuaties in de samenstelling van de steekproef. De orde van grootte van deze variaties is te bepalen aan de hand van variantieberekeningen. Wanneer de verwachtingswaarde afwijkt van de werkelijke waarde is er sprake van een onzuivere schatter. Deze afwijking wordt onzuiverheid of vertekening genoemd. Bij steekproefuitkomsten wordt meestal slechts een uitspraak gedaan over de precisie omdat informatie over de onzuiverheid meestal ontbreekt.

Betrouwbaarheid

Een grootte waarmee de precisie van schattingen kan worden aangegeven is het zogenaamde 95%-betrouwbaarheidsinterval. Met behulp van de waarschijnlijkheidsrekening kan bij elk van de uitkomsten van POLS een dergelijk interval worden berekend. De betekenis van zo'n interval is dat, indien de steekproef en het onderzoek vele malen zouden worden herhaald, en steeds een 95%-betrouwbaarheidsinterval zou worden bepaald, het in minstens 95 van de 100 gevallen de werkelijke waarde zal bevatten. Hoe smaller het interval, des te preciezer de uitkomst.

Uit het begrip 95%-betrouwbaarheidsinterval kunnen twee andere begrippen worden afgeleid: de absolute en de relatieve 95%-betrouwbaarheidsmarge. De absolute marge is de afstand van de schatting tot één der grenzen van het betrouwbaarheidsinterval; de relatieve marge is het quotiënt van de absolute marge en de schatting. Opgemerkt wordt dat het 95%-betrouwbaarheidsinterval en de daaruit afgeleide marges alleen aangeven hoe groot de afwijking van de geschatte waarde met de werkelijke waarde zou kunnen zijn ten gevolge van het toeval. Afwijkingen die niet het gevolg zijn van het toeval, zoals fouten bij de steekproeftrekking, fouten in het veldwerk of fouten in de verwerking, zijn niet in deze marge begrepen.

PRAT

Het **Programma Anders Typeren (PRAT)** staat voor nieuwe typeermethoden voor bedrijf, beroep, en onderwijs. Het resultaat is een vernieuwde vragenlijst waarbij de interviewer – ondersteund door software op de laptop – typeercodes voor bedrijf, beroep, en opleidingen kan vaststellen. De ontwikkeling en beoordeling van de nieuwe vragenlijst is ondersteund door een uitgebreid testtraject.

Anders typeren is het toepassen van andere typeertechnieken dan handmatig typeren. Omdat bij deze typeertechnieken de toekenning van de codes geheel of gedeeltelijk is geautomatiseerd wordt ook wel gesproken van automatisch typeren. Een belangrijke consequentie van anders typeren is dat de uitvoering van het proces verschuift van typeurs naar veldinterviewers en respondenten of naar geheel geautomatiseerde processen. In het geval van typeren door interviewers worden hulpmiddelen beschikbaar gesteld opdat interviewers in het veld nauwkeurig en snel codes kunnen toekennen. In het geval van geheel automatisch typeren wordt met behulp van speciale software aan informatie over bedrijf, beroep en opleiding automatisch een typeercode toegekend. Ook

het trainen van deze systemen is geautomatiseerd. Het geheel automatisch typen vindt plaats op het CBS. Een belangrijke consequentie van bovengenoemde methoden van anders typen is de noodzaak om de vraagstelling voor bedrijf, beroep en onderwijs te standaardiseren. Deze aanpak past binnen de SSB werkwijze om verschillende onderzoeken zoveel mogelijk te integreren.

Oude vraagstelling en vraagstelling PRAT

Hieronder volgt een kort overzicht van de oude vraagstelling voor POLS voor zover deze betrekking heeft op de codering van bedrijf, beroep en onderwijs.

- **Bedrijf**

Voor POLS is de vraagstelling voor bedrijf eenvoudig. Er wordt hier alleen gevraagd naar de bedrijfsactiviteit.

De verzamelde gegevens worden op het CBS gebruikt om aan de werkring een code voor de economische activiteit van het bedrijf toe te kennen (SBI). Deze toekenning gebeurt handmatig. Is de bedrijfsnaam en het adres beschikbaar dan wordt het bedrijf en de bijbehorende SBI-code door de typeur opgezocht in het ABR(Algemeen Bedrijfsregister). Is alleen de bedrijfsactiviteit gegeven dan wordt op basis hiervan door de typeur de SBI-code vastgesteld.

- **Beroep**

Er wordt gevraagd naar het beroep of de functie die de respondent in het bedrijf uitoefent, of men leiding geeft en zo ja aan hoeveel personen, en de belangrijkste werkzaamheden en leidinggevende werkzaamheden. Deze informatie wordt door de typeur samengevoegd met de in de bedrijvenmodule vastgestelde bedrijfsactiviteit. Op basis hiervan wordt een voorlopige beroepencode vastgesteld (bewerkingscode). In geval van twijfel kan ook van extra informatie –bijvoorbeeld opleiding of arbeidspositie- gebruik worden gemaakt. Uitgaande van de bewerkingscode en informatie over opleidingsniveau, opleidingsrichting, aantal personen waaraan men leiding geeft, en arbeidspositie kunnen de beroepenclassificaties worden afgeleid (SBC, ISCO).

- **Onderwijs**

Er wordt voor het coderen van de opleidingen gevraagd naar opleidingsnaam, soort school of instelling, en studierichting (of leerweg, vak, module). De verzamelde gegevens worden op het CBS door typeurs worden voorzien van een opleidingsnummer. Met dit opleidingsnummer kunnen de coderingen binnen de diverse onderwijsclassificaties (SOI, ISCED) worden bepaald en kunnen relevante kenmerken als hoogst gevolgde en hoogst behaalde opleiding worden afgeleid.

Vraagstelling PRAT

De PRAT-vragenlijsten voor POLS bestaan uit de volgende componenten: de Blaise-vragenlijst, de zoekbestanden voor bedrijf, beroep en onderwijs, en de zoekmachines. De zoekmachines zijn aparte stukjes programmatuur waarmee tekstantwoorden van respondenten aan bedrijven, beroepen en opleidingen in de zoekbestanden kunnen

worden gekoppeld. Er zijn twee typen zoekmachines; één voor bedrijf en beroep en één voor onderwijs. Hieronder volgt een beknopte uitleg van de werking van de diverse componenten.

- Bedrijf

De Blaise vragenlijst begint met enkele vragen naar omvang van de vestiging/het gehele bedrijf voor werknemers en omvang van het eigen bedrijf (of dat van de partner) voor zelfstandigen. Vervolgens wordt de zoekmachine voor bedrijf gestart. De vraagstelling binnen de add-in wordt bepaald door de bedrijfsomvang: in geval het gehele bedrijf 100 of meer werknemers heeft wordt gestart met de vraag naar de naam van het bedrijf en de zoekmachine selecteert in het zoekbestand met bedrijfsnamen de meest lijkende bedrijfsnamen van grote bedrijven (dit zijn juridische en handelsnamen uit het ABR). Bij selectie van een groot aantal bedrijfsnamen (≥ 7) wordt gevraagd om de bedrijfsnaam verder uit te schrijven om het aantal mogelijkheden in te perken. Is vervolgens het aantal overblijvende bedrijfsnamen niet te groot (< 7) dan worden deze aan de interviewer getoond in de vorm van een keuzelijst, waarbij de namen moeten worden teruggekoppeld naar de respondent. Uit deze keuzelijst kunnen interviewer en respondent een keuze maken voor de meest toepasselijke bedrijfsnaam (en SBI). Als in het zoekbestand voor bedrijfsnamen niets wordt gevonden gaat de zoekmachine na of de bedrijfsnaam in het bestand voor bedrijfsomschrijvingen voorkomt. Het blijkt namelijk uit historische gegevens van de EBB dat bij grote bedrijven vaak de bedrijfsnaam wordt genoemd bij de vraag naar soort bedrijf. Levert ook dit niets op of als er te veel mogelijkheden over blijven dan wordt in de Blaise-vragenlijst verder gevraagd naar de economische activiteit van het bedrijf. Bij bedrijven met minder dan 100 werknemers en bij zelfstandigen is dit sowieso de eerste vraag. De zoekmachine wordt aangeroepen en deze selecteert in het zoekbestand voor soort bedrijf de meest lijkende bedrijfsomschrijvingen. Is het aantal gevonden omschrijvingen niet te groot (< 7) dan wordt ook hier de selectie voorgelegd aan de respondent in de vorm van een keuzelijst. Keuze van een item uit deze lijst betekent dat een SBI-code is bepaald. Indien in het zoekbestand geen geschikte bedrijfsomschrijving wordt gevonden of als de respondent geen keuze kan maken dan wordt na het afsluiten van de zoekmachine de zogenaamde ‘boomvraag’ gesteld: met behulp van een getrapte vraagstelling wordt een grove SBI-code vastgesteld –meestal op 2 à 3 digits, soms 1 digit. Niet-getypeerde bedrijven worden op het CBS door typeurs van een code voorzien. Zij maken hierbij gebruik van de verzamelde informatie over bedrijfsnamen en bedrijfsomschrijvingen.

- Beroep

De beroepenmodule begint met de vraag naar het ‘beroep of functie’ van de respondent (uitgezonderd zelfstandigen). De zoekmachine wordt vervolgens gestart en deze selecteert in het zoekbestand voor beroepsnamen de meest lijkende beroepsomschrijvingen. Is de selectie voldoende klein (< 7) dan worden de beroepsomschrijvingen aan de interviewer getoond in de vorm van een keuzelijst, waarbij de antwoordmogelijkheden moeten worden teruggekoppeld naar de respondent. Uit deze lijst kunnen interviewer en respondent een keuze maken. Keuze van een item uit deze lijst

betekent dat een bewerkingscode (bwc) wordt bepaald. Wordt er geen keuzelijst aangeboden of kan de respondent geen keuze maken dan wordt vervolgens de vraag naar belangrijkste werkzaamheden gesteld. Hierbij kunnen 2 verschillende situaties optreden:

- als bij de voorgaande vraag naar beroepsnaam een keuzelijst is voorgelegd en de respondent heeft geen keuze kunnen maken dan wordt bij de vraag naar belangrijkste werkzaamheden de zoekmachine niet meer opgestart; dit om te voorkomen dat meerdere malen naar ‘beroep of functie’ wordt gevraagd.
- Is er echter geen keuzelijst voorgelegd (omdat geen of juist te veel beroepsomschrijvingen zijn geselecteerd) dan wordt de zoekmachine wel aangeroepen. In dat geval wordt met behulp van een zoekbestand met verzamelde werkzaamheden alsnog geprobeerd de respondent te laten kiezen uit een beperkte selectie van corresponderende beroepen. Lukt dit niet dan worden net als bij bedrijf de niet-getypeerde beroepen op het CBS door typeurs van een bewerkingscode voorzien. Gecombineerd met informatie over onderwijsniveau, onderwijsrichting, arbeidspositie, en het aantal personen waaraan de respondent leiding geeft, worden in het productieproces vervolgens automatisch de diverse beroepenclassificaties afgeleid (SBC en ISCO).
- Onderwijs

De zoekmachine voor onderwijs is een andere dan die voor bedrijf en beroep. Direct bij aanvang van de vragenmodule voor actueel of gevolgd onderwijs wordt de zoekmachine opgestart. De eerste vraag betreft de naam van de opleiding of cursus. Met dit gegeven selecteert de zoekmachine in het zoekbestand onderwijs de meest lijkende opleidingsnamen. Vervolgens worden een aantal kenmerken van de opleiding gevraagd waarmee deze selectie verder kan worden ingeperkt. Voorbeelden van dergelijke kenmerken zijn: onderwijsinstelling/soort onderwijs (universiteit, hbo), lerarenopleiding, en vak/studierichting. Indien de selectie van opleidingen voldoende klein is geworden (<7 opleidingen) bestaan er twee mogelijkheden: 1) er is slechts één opleiding overgebleven; deze wordt niet getoond aan de respondent, maar wordt direct opgeslagen als typeerresultaat óf 2) er zijn een beperkt aantal opleidingen geselecteerd (minimaal 2, maximaal 6); deze worden voorgelegd aan de interviewer, die de selectie vervolgens terugkoppelt naar de respondent. Uit deze selectie kiest de respondent de meest toepasselijke opleiding.

GBA-selectiebestand

Het bijbehorende GBA-selectiebestand bevat persoons- en huishoudensgegevens van alle personen die in het microdatabestand voorkomen en ook geregistreerd staan in de GBA. De persoonsgegevens bestrijken een periode van 1 januari 1995 tot 1 januari 2010. De huishoudensgegevens worden door het CBS berekend voor de stand op 1 januari. In het GBA-selectiebestand zijn huishoudensgegevens beschikbaar van 1998 tot en met 2010.

Zowel het microdatabestand als het GBA-selectiebestand is gesorteerd op respectievelijk SRTNUM en RIN. Met behulp van deze twee variabelen als koppelsleutel kan de

gebruiker persoons- en huishoudensgegevens uit het GBA-selectiebestand toevoegen aan het microdatabestand, zodat deze gegevens in het onderzoek meegenomen kunnen worden.

Om RIN-nummers toe te kennen aan het microdatabestand is gebruik gemaakt van de meest recente versie van de GBA. Hierin is slechts een beperkt aantal variabelen aanwezig, daarom wordt dit bestand niet gebruikt voor het GBA-selectiebestand. Het GBA-selectiebestand maakt gebruik van gegevens tot 1 januari 2010. Hierdoor kunnen personen die na 31 december 2009 naar Nederland zijn geïmmigreerd, niet worden teruggevonden in het GBA-selectiebestand.

3. Bestandsopbouw en toelichting

Bestandsopbouw van het microdatabestand

Onderstaand volgt een lijst van alle variabelen in het microdatabestand.

Kwaliteitsaanduiding

De kolom 'Kw' geeft, waar nodig, de kwaliteit van de variabelen aan op het moment waarop dit documentatierapport is opgesteld:

- 'kwaliteitsaanduiding is niet nodig; bijvoorbeeld bij hulpvariabelen'
- 1. 'variabele is gebruikt in publicaties en is geanalyseerd';
- 2. 'variabele is gecontroleerd op plausibiliteit en zou in de toekomst nog kunnen veranderen';
- 3. 'naar de variabele is nog niet gekeken'.

De volgende afkortingen komen in de labels voor:

- OP 'Ondervraagd persoon',
- HH 'Huishouden',
- KW 'Kostwinner'.

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
1	SRTNUMOP Soort identificatiecode OP	-	Tekst	1	0
2	RINOP Identificatienummer OP	-	Getal	9	0
3	SRTNUM1 Soort identificatiecode persoon 1	-	Tekst	1	0
4	RIN1 Identificatienummer persoon 1	-	Getal	9	0
5	SRTNUM2 Soort identificatiecode persoon 2	-	Tekst	1	0
6	RIN2 Identificatienummer persoon 2	-	Getal	9	0
7	SRTNUM3 Soort identificatiecode persoon 3	-	Tekst	1	0
8	RIN3 Identificatienummer persoon 3	-	Getal	9	0
9	SRTNUM4 Soort identificatiecode persoon 4	-	Tekst	1	0
10	RIN4 Identificatienummer persoon 4	-	Getal	9	0
11	SRTNUM5 Soort identificatiecode persoon 5	-	Tekst	1	0

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
12	RIN5 Identificatienummer persoon 5	-	Getal	9	0
13	SRTNUM6 Soort identificatiecode persoon 6	-	Tekst	1	0
14	RIN6 Identificatienummer persoon 6	-	Getal	9	0
15	SRTNUM7 Soort identificatiecode persoon 7	-	Tekst	1	0
16	RIN7 Identificatienummer persoon 7	-	Getal	9	0
17	SRTNUM8 Soort identificatiecode persoon 8	-	Tekst	1	0
18	RIN8 Identificatienummer persoon 8	-	Getal	9	0
19	WPERIODE Veldwerkperiode (jjjjmm)	1	Getal	10	0
20	GEWBGEZO Persoonsgewicht	1	Getal	25	5
21	GEWHERKO Persoonsgewicht, weging met oa. var.herkomstgeneratie	1	Getal	25	5
22	GEWHHBAS Gewicht huishouden	1	Getal	25	12
23	GWHHHERK Gewicht huishouden berekend uit gewkherko	1	Getal	25	12
24	SURVEYNR Surveynummer	1	Tekst	2	0
25	RESPMND Enquete maand	1	Getal	10	0
26	REF_DAT Referentiedatum leeftijdsberekening OP	1	Tekst	8	0
27	REGOP Regelnummer huishoudbox respondent	1	Getal	10	0
28	AANTALPP Hhbox aantal personen	1	Getal	10	0
29	HHKERN Hhbox relatie tot hhkern	1	Getal	10	0
30	TOTPP18 Tot.aant.pers. > 17 jr(excl. lft onbek.)	1	Getal	10	0
31	RESDOOR2 Beantwoorder van de enquete	1	Getal	10	0
32	LFTOP Leeftijd respondent-datum interview	1	Getal	10	0
33	LFT01 Leeftijd OP op 1 januari	1	Getal	10	0
34	TLFTOP Leeftijd steekproefpersoon 2 klassen	1	Getal	10	0
35	GSLOP Geslacht respondent	1	Getal	10	0
36	BURGSTOP Burgerlijke staat OP op datum vraaggesprek	1	Getal	10	0

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
37	PLHHOP Plaats in huishouden OP	1	Getal	10	0
38	SAMHHUIT Samenstelling huishouden (uitgebreid)	1	Getal	10	0
39	SAMHH00 Samenstelling huishouden 2000	1	Getal	10	0
40	HHT Huishoudenssamenstelling	1	Getal	10	0
41	POPOPB Basisindeling bevolking CBS	1	Getal	10	0
42	NEDNATOP Nationaliteit stkprfpers	1	Getal	10	0
43	NEDOP OP nederlandse nationaliteit of een andere	1	Getal	10	0
44	GLANDOP Geboorteland stkprfpers	1	Getal	10	0
45	GLANDVA Geboorteland vader van stkprfpers	1	Getal	10	0
46	GLANDMO Geboorteland moeder van stkprfpers	1	Getal	10	0
47	HERK3 Herkomst in drie categoriën	1	Getal	10	0
48	EBB Zevendeling in gebruik bij ebb	1	Getal	10	0
49	HERKOMST Zevendeling in gebruik bij ssb	1	Getal	10	0
50	HERKGNRT Niet westers allochtoniteit met generatie	1	Getal	10	0
51	TYDBESOP Tijdsbesteding- OP	1	Getal	10	0
52	BETWRKOP_AFL Betaaldwerk_afleiding	1	Getal	10	0
53	BETWRKOP Betaald werk, 1 uur pw telt- OP	1	Getal	10	0
54	AUURWKOP Afleiding werk van 12 uur of meer per week	1	Getal	10	0
55	SUURWKOP Schatting aantal uren betaald werk- OP	1	Getal	10	0
56	UITKEROP Ontvangen sociale uitkering- OP	1	Getal	10	0
57	ARBONGOP Ontvangen arbeidsongeschiktheidsuitk.- OP	1	Getal	10	0
58	UURHWOP Aantal uren huishoudelijk werk- OP	1	Getal	10	0
59	SUURHWOP Schatting aantal uur huish. werk- OP	1	Getal	10	0
60	MAPARTOP Maatschappelijke participatie OP	1	Getal	10	0

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
61	SOI98AOP Soi'98 huidige (actuele) opleiding OP	1	Getal	10	0
62	SOIAC_OP Soi'78 actueel onderwijs OP	1	Getal	10	0
63	ISCEDAOP Isced huidige (actuele) opleiding OP	1	Tekst	17	0
64	WRKNEMOP Werkt 12 uur of meer als werknemer- OP	1	Getal	10	0
65	BEDRPROP Werkzaam in eigen bedrijf of dat van partner/ouders- OP	1	Getal	10	0
66	WILWRKOP Zou OP werk willen hebben 12 uur of meer- OP	1	Getal	10	0
67	URENGAOP Werk gevonden: aantal uur per week	1	Getal	10	0
68	AANPWERKOP Aanpassingen op werk dan meer uren mogelijk	1	Getal	10	0
69	WIL_12UROP Wil 12 uur of meer per week werken	1	Getal	10	0
70	WIL_URWKOP Wil werken: aantal uur per week	1	Getal	10	0
71	EIGENBDFOP Wil werken: in eigen bedrijf of praktijk	1	Getal	10	0
72	KUNSTROP Begintermijn werk- OP	1	Getal	10	0
73	VERWSTRT Verwachte startdatum bedrijf of praktijk	1	Getal	10	0
74	AFRONDOP1 Reden afrondingsperiode: huidig werk	1	Getal	10	0
75	AFRONDOP2 Reden afrondingsperiode: vrijwilligerswerk	1	Getal	10	0
76	AFRONDOP3 Reden afrondingsperiode: opleiding of studie	1	Getal	10	0
77	AFRONDOP4 Reden afrondingsperiode: of kinderopvang	1	Getal	10	0
78	OF_IVMOP1 Persoonlijke reden afrondingsperiode: ziekte	1	Getal	10	0
79	OF_IVMOP2 Persoonlijke reden afrondingsperiode: vakantie	1	Getal	10	0
80	OF_IVMOP3 Persoonlijke reden afrondingsperiode: persoonlijke omstandigheden	1	Getal	10	0
81	AFG4WKOP Zoekact. werk afg. 4 weken- OP	1	Getal	10	0
82	WERKZOP aantal uren werkzaam per week	1	Getal	10	0

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
83	WENSOP Wil werken OP	1	Getal	10	0
84	BESCHKOP Beschikbaar voor arbeidsmarkt-OP	1	Getal	10	0
85	ZOEKCT2OP Zoekactiviteiten werk OP	1	Getal	10	0
86	BERBEV2OP Beroepsbevolking OP	1	Getal	10	0
87	POSWRKOP Positie in werkkring OP	1	Getal	10	0
88	ARBPOSOP Arbeidspositie OP	1	Getal	10	0
89	CDBEDROP Categorie bedrijf	1	Getal	10	0
90	CODBEROP Categorie beroep	1	Getal	10	0
91	TK_06 Gestemd voor de 2e kamerverkiezingen in 2006	1	Getal	10	0
92	GELOVIG Kerkelijke gezindte	1	Getal	10	0
93	KERKBEZ Frequentie kerkbezoek	1	Getal	10	0
94	GEMCODE Gemeentecode	1	Getal	10	0
95	ARRON Rechtsgebieden arrondissementsrechtbanken	1	Getal	10	0
96	COROP Corop-gebieden	1	Getal	10	0
97	GEMGRJJ Gemeentegrootte	1	Getal	10	0
98	HOF Rechtsgebieden gerechtshoven	1	Getal	10	0
99	LANDD Landsdelen	1	Getal	10	0
100	POLRG Politieregio's (1991)	1	Getal	10	0
101	STEDGEM Stedelijkheid gemeenten	1	Getal	10	0
102	TOER Toeristengebieden	1	Getal	10	0
103	GG2 Grootteklasse woongem-herc	1	Getal	10	0
104	PROV Provincie geherc.	1	Getal	10	0
105	POSCODN Postcode cijferdeel	1	Getal	10	0
106	STEDAGG Grootstedelijke agglomeraties	1	Getal	10	0
107	STEDGEW Stadsgewesten (functioneel)	1	Getal	10	0
108	GR25 25 grootste gemeenten per 1-1-2009	1	Getal	10	0

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
109	LFTJONG Leeftijd jongste kind	1	Getal	10	0
110	LFTOUD Leeftijd oudste kind	1	Getal	10	0
111	LFTKND3 Leeftijdscat. 3e kind	1	Getal	10	0
112	LFTKND4 Leeftijdscat. 4e kind	1	Getal	10	0
113	LFTKND5 Leeftijdscat. 5e kind	1	Getal	10	0
114	LFTKND6 Leeftijdscat. 6e kind	1	Getal	10	0
115	LFTKND7 Leeftijdscat. 7e kind	1	Getal	10	0
116	GSLJONG Geslacht jongste kind	1	Getal	10	0
117	GSLOUD Geslacht oudste kind	1	Getal	10	0
118	GSLKND3 Geslacht 3e kind	1	Getal	10	0
119	GSLKND4 Geslacht 4e kind	1	Getal	10	0
120	GSLKND5 Geslacht 5e kind	1	Getal	10	0
121	GSLKND6 Geslacht 6e kind	1	Getal	10	0
122	GSLKND7 Geslacht 7e kind	1	Getal	10	0
123	LFTOVR1 Leeftijdscat. 1e niet hh/partner hh/kind	1	Getal	10	0
124	LFTOVR2 Leeftijdscat. 2e niet hh/partner hh/kind	1	Getal	10	0
125	LFTOVR3 Leeftijdscat. 3e niet hh/partner hh/kind	1	Getal	10	0
126	LFTOVR4 Leeftijdscat. 4e niet hh/partner hh/kind	1	Getal	10	0
127	LFTOVR5 Leeftijdscat. 5e niet hh/partner hh/kind	1	Getal	10	0
128	LFTOVR6 Leeftijdscat. 6e niet hh/partner hh/kind	1	Getal	10	0
129	LFTOVR7 Leeftijdscat. 7e niet hh/partner hh/kind	1	Getal	10	0
130	GSLOVR1 Geslacht 1e niet hh/partner hh/kind	1	Getal	10	0
131	GSLOVR2 Geslacht 2e niet hh/partner hh/kind	1	Getal	10	0
132	GSLOVR3 Geslacht 3e niet hh/partner hh/kind	1	Getal	10	0

Nr	Variabele & label	Kw	Format	Lengte	Decimalen
133	GSLOVR4 Geslacht 4e niet hh/partner hh/kind	1	Getal	10	0
134	GSLOVR5 Geslacht 5e niet hh/partner hh/kind	1	Getal	10	0
135	GSLOVR6 Geslacht 6e niet hh/partner hh/kind	1	Getal	10	0
136	GSLOVR7 Geslacht 7e niet hh/partner hh/kind	1	Getal	10	0
137	N12PLUS Aantal personen 12 jaar of ouder in huish	1	Getal	10	0
138	OPLNIVEAU_VRLAB Onderwijsniveau -jan-mei2009-respondenten-	1	Getal	2	0
139	OPLNIVEAU Onderwijsniveau -na-31mei2009-respondenten-	1	Getal	2	0
140	OPLNIV5 Onderwijs in 5 niveaus van -na-31mei2009-respondenten-	1	Getal	2	0
141	LEERJROP Huidig leerjaar	1	Getal	10	0
142	SOI_ACOP Soi98aop volgens automatische typeerproces (prat)- act.ond. op	1	Getal	10	0
143	OPLNALS_VRLAB Opleiding gevolgd na lo	1	Getal	10	0
144	DIPLOMA_VRLAB Opleiding na lo voltooid (diploma)	1	Getal	10	0
145	NIVEAU_VRLAB Niveau voltooide opl. na lo	1	Getal	10	0
146	OPLNALO Gevolgd onderwijs: na lo 2 jr of meer	1	Getal	10	0
147	NIVGEV Gevolgd onderwijs: niveau	1	Getal	10	0
148	SRTMBO Gevolgd onderwijs: niveau mbo	1	Getal	10	0
149	DIPLOM Gevolgd onderwijs: diploma	1	Getal	10	0
150	LASTDIPL Gevolgd onderwijs: laatst behaalde diploma	1	Getal	10	0
151	SRTMBO2 Gevolgd onderwijs: laatste behaald, niveau mbo	1	Getal	10	0
152	KLAS4 Gevolgd onderwijs: overgegaan naar klas 4	1	Getal	10	0
153	GEBJAAR Geboortejaar	1	Getal	4	0
154	GEBMND Geboortemaand	1	Getal	2	0
155	ENCRYPTED_WE_ID Versleutelde WE_ID	1	Tekst	8	0

Toelichting op de variabelen

SRTNUMOP, SRTNUM1-8, RINOP, RINI-8

Bronbestanden die aan het CBS geleverd worden, bevatten doorgaans als persoons-identificator het burgerservicenummer. Het burgerservicenummer wordt als zeer identificerend beschouwd en daarom vervangt het CBS dit nummer door een intern persoonsidentificatienummer, het betekenisloze RIN-nummer. Met dit RIN-nummer kan het bestand worden gekoppeld aan andere bestanden bij het CBS.

Het bronbestand van dit microdatabestand bevatte geen burgerservicenummer. Via andere identificerende gegevens dan het burgerservicenummer kon het RIN-nummer ook aan dit bestand worden toegevoegd. Dit is gebeurd door van ieder record in het bronbestand de combinatie van geslacht, geboortedatum en adres op te zoeken in de GBA. Het CBS heeft namelijk iedere persoon in de GBA voorzien van een RIN-nummer. Als de specifieke combinatie voorkomt in de GBA, dan is het RIN-nummer van die persoon toegevoegd aan het record in het bronbestand. Op deze manier kon het RIN-nummer van personen in het bronbestand worden achterhaald.

Het microdatabestand is voorzien van de door het CBS toegevoegde variabelen soort identificatiecode (SRTNUM) en identificatienummer (RIN). Records in het microdatabestand die konden worden voorzien van een RIN-nummer hebben de waarde 'R' op SRTNUM gekregen. De overige records zijn niet teruggevonden in de GBA. Deze records hebben de waarde 'G' op SRTNUM gekregen, met een lege score in het veld RIN.

Uitsluitend de combinatie van SRTNUM 'R' en RIN identificeert een uniek persoon. Koppelen aan andere bestanden of aggregeren op persoonsniveau moet daarom alleen worden uitgevoerd op records met SRTNUM 'R'. Met behulp van SRTNUM en RIN kan de gebruiker gegevens uit het GBA-selectiebestand toevoegen aan de records in het microdatabestand met SRTNUM 'R'.

GEMCODE

De gemeentecodes op 1 januari van het verslagjaar. Voor een toelichting op het ontstaan, opheffing en/of naamswijziging van gemeenten zie de Statline tabel [Historisch overzicht gebieden](#). Voor Remote Access en On-site gebruikers is de gemeentecodelijst te vinden op \8_Uilities\Code_Listings\Gemeentecodes\.

Met de gemeentecode kunt u aan de hand van dit sleutelveld en het GIN bestand van hetzelfde jaar gegevens samenvoegen, tellingen maken, naar een hoger regionaal niveau zoals landsdeel, provincie, COROP of toeristengebied. U kunt de GIN bestanden vinden op de Onsite/remote access omgeving: \8_Uilities\Code_Listings\Gebieden in Nederland\.

Toelichting ISCED-variabelen

In het bestand komen zes ISCED (onderwijs codering van de UNESCO)-variabelen voor, met een zeer specifieke indeling.

Voorbeeld: 3_A_#_3_#_1_#_141

Deze 17 posities geven in acht variabelen een inzicht in de opleiding. De eerste twee variabelen 'level' en 'destination' geven samen het niveau, en andere specifieke eigenschappen van de opleiding aan.

De variabelenamen, volgens ISCED 1997, de positie(s) en de labels, met Nederlandse toelichting staan hieronder:

Variabele	positie	Codes	Omschrijving
Level en destination	1 en 3	0	Onderwijs aan kleuters
		1	Primair onderwijs
		2	(Lager) secundair onderwijs
		2A	Programma's ontworpen voor directe toegang tot opleidingen op niveau 3A en 3B die op hun beurt toegang verschaffen tot het hoger onderwijs beschreven onder ISCED 5A en 5B
		2B	Programma's die in de eerste plaats leiden naar eind(=beroeps)onderwijs op niveau 3C
		2C	Lager eind(=beroeps)onderwijs
		3	(Hoger) secundair onderwijs
		3A	Programma's die toegang geven tot 5A-programma's, d.w.z. tot het wetenschappelijk onderwijs en tot het 4-jarig hbo
		3B	Programma's die toegang geven tot 5B-programma's
		3C	Beroepsonderwijs: WEB-basisberoepsopleiding en -vakopleiding
		4	Post-sekundair niet tertiair onderwijs
		4A	Programma's die voorbereiden op 5A
		4B	Programma's die voorbereiden op 5B
		4C	Beroepsonderwijs; Afsluitende programma's die primair zijn bedoeld voor intrede in de arbeidsmarkt
		5	Eerste fase tertiair onderwijs (niet direct leidend tot wetenschappelijk diploma)
		5A	Wetenschappelijk onderwijs en (post-)hbo (4-jarig)

		5B	Kort hbo
		6	Tweede fase tertiair educationonderwijs
Exam	5		
		#	Niet relevant voor ISCED-categorie
		1	Tussstijds examen (tertiair onderwijs)
		2	eerste examen (tertiair onderwijs)
		3	Tweede en volgende. Examen (tertiair onderwijs)
Orientation	7		
		#	Niet van toepassing
		1	Geen beroepsopleiding
		2	Vorbereidende beroepsopleiding
		3	Beroepsopleiding
Duration	9		
		#	Niet relevant voor ISCED-categorie
		1	=< 3 jaar cumulatief in fulltime-equivalenten op het betreffende ISCED-niveau
		2	3-<5 jaar
		3	5-<6 jaar
		4	>=6 jaar
BBL	11		
		#	Niet relevant voor ISCED-categorie
		1	Opleiding is een werkend-leren-opleiding op level 3 of 4 (beroepsbegeleidende leerweg)
(Geen) Level completion	13		
		#	Niet relevant voor ISCED-categorie of er is sprake van niveau voltooiing
		1	Geen niveau voltooiing (alleen gebruikt bij 3C)
Field	15-17		(Hier is de oorspronkelijke Engelse text gehanhaafd)
		010	Basic programmes
		080	Literacy and numeracy
		090	Personal development
		141	Teaching and training
		142	Education science
		210	Arts

211	Fine arts
212	Music and performing arts
213	Audio-visual techniques and media production
214	Design
215	Craft skills
220	Humanities
221	Religion and theology
222	Foreign languages and cultures
223	Mother tongue
224	History, philosophy and related subjects
310	Social and behavioural science
321	Journalism and reporting
322	Library, information, archive
340	Business and administration
341	Wholesale and retail sales
342	Marketing and advertising
343	Finance, banking, insurance
344	Accounting and taxation
345	Management and administration
346	Secretarial and office work
347	Working life
380	Law
400	Science
420	Life science
440	Physical science
460	Mathematics and statistics
481	Computer science
482	Computer use
500	Engineering, Manufacturing and Construction
520	Engineering and engineering trades
521	Mechanics and metal work
522	Electricity and energy
523	Electronics and automation
524	Chemical and process

525	Motor vehicles, ships and aircraft
541	Food processing
542	Textiles, clothes, footwear, leather
543	Materials (wood, paper, plastic, glass)
544	Mining and extraction
581	Architecture and town planning
582	Building and civil engineering
620	Agriculture, forestry
621	Crop and livestock production
622	Horticulture
623	Forestry
624	Fisheries
640	Veterinary
720	Health
721	Medicine
722	Medical services
723	Nursing
724	Dental studies
760	Social services
761	Child care and youth services
762	Social work and counselling
811	Hotel, restaurant and catering
812	Travel, tourism and leisure
813	Sports
814	Domestic services
815	Hair and beauty services
840	Transport services
850	Environmental protection
860	Security services
861	Protection of property and persons
862	Occupational health and safety
863	Military
999	Not known or unspecified

Voorkomende waarden van de variabelen

Zie bijlage ‘101227 Bijlage 2 Beschrijving variabelen POLSbasis 2009V1’ voor de extra informatie per variabele. Voor de categoriale variabelen worden alle mogelijke scores en hun betekenis opgesomd.