

Steunpunt Demografie

Point d'Appui Démographie

Interface Demography

IN DE WETENSCHAP HET NIET TE WETEN

Achtergronden van de non-respons
in de Volkstelling van 1991

Reinhard STOOP & Johan SURKYN

STEUNPUNT DEMOGRAFIE, VRIJE UNIVERSITEIT BRUSSEL

WORKING PAPER NR. 1997-1

IN DE WETENSCHAP HET NIET TE WETEN

Achtergronden van de non-respons
in de Volkstelling van 1991

Reinhard STOOP & Johan SURKYN

STEUNPUNT DEMOGRAFIE, VRIJE UNIVERSITEIT BRUSSEL

WORKING PAPER NR. 1997-1

IN DE WETENSCHAP HET NIET TE WETEN

Achtergronden van de non-respons in de Volkstelling van 1991.

REINHARD STOOP & JOHAN SURKYN

1. INLEIDING

De non-respons in volkstellingen en de beoordeling van de kwaliteit van de beschikbare data vormen - zoals dat ook voor grote steekproefenquêtes het geval is - een complex en soms vrij ondoorzichtig studieobject. Bovendien kunnen we deze elementen in de Belgische census maar evalueren aan de hand van de opgekuiste datafile, zodat de originele non-respons op de eigenlijke telformulieren niet meer bekend is. Wat we als vastgestelde non-respons terugvinden is dus een combinatie van ontbrekende of foutieve informatie die op verschillende niveaus ontstond en later mogelijk weer gecorrigeerd of aangevuld werd. De door het N.I.S. toegepaste correctieprocedure liet namelijk enerzijds toe een aantal missings te recupereren, ofwel via informatie die rechtstreeks uit het Rijksregister werd opgevraagd, ofwel via automatische verbeteringen door het computerprogramma, of tenslotte ook door het originele formulier opnieuw ter hand te nemen en daarvan het juiste verhaal trachten af te leiden. Daarbij was het natuurlijk mogelijk dat al deze procedures nog steeds niet toelieten een coherent en geloofwaardig beeld van het betrokken individu of huishouden te verkrijgen en in dat geval werden een aantal variabelen die mogelijk oorspronkelijk een op zichzelf geldige code hadden naar een missing-waarde gehercodeerd. Spijtig genoeg werd daarbij niet consequent voor een nog niet bestaande en originele code geopteerd. Het ontbreken van een antwoord werd voor een aantal variabelen -in de meerderheid der gevallen ongetwijfeld terecht- als een negatief antwoord geïnterpreteerd. Het gevolg hiervan is dat we voor een aantal variabelen dus wel weten in welke "brede" code de ontbrekende antwoorden zijn opgenomen, maar dat we de (geldige of ontbrekende) herkomst van die codes niet meer kunnen achterhalen. In die gevallen is dus ook het aandeel der ontbrekende antwoorden, of de verdeling ervan over een aantal relevante achtergrondvariabelen onbekend. Het is dus duidelijk dat we de non-respons niet kunnen behandelen zonder eerst dieper in te gaan op de toegepaste correctieprocedure.

We kunnen er bovendien niet omheen dat de census in feite per huishouden en niet per individu georganiseerd wordt, wat vrij verre gaande implicaties heeft. De eerste daarvan heeft te maken met de aanduiding van een referentiepersoon, doorgaans het gezinshoofd. De keuze van een referentiepersoon zal in kleine kerngezinnen meestal weinig problemen stellen, wanneer er echter meerdere generaties of gezinskernen samenwonen wordt de keuze moeilijker. Aangezien de verwantschapsrelaties in het gezinsformulier betrekking hebben op de band met de referentiepersoon is de keuze van deze laatste cruciaal voor het identificeren van het huishoudenstype.

Dit huishoudenstype en de verwantschapsrelaties met de referentiepersoon worden in de correctieprocedure zeer vaak als een aanknopingspunt gebruikt.

We kunnen er van uitgaan dat het gezinshoofd uit het Rijksregister doorgaans als referentiepersoon werd aangeduid en zelf het gezinsformulier en het woningformulier invulde, maar vaak ook individuele gegevens voor de andere leden van het huishouden. Ook andere personen kunnen soms in naam van het ganse huishouden optreden, zonder dat ze noodzakelijk het statuut van "referentiepersoon" krijgen. Dit heeft enerzijds voor gevolg dat ontbrekende informatie meer voorkomt naarmate dat het een meer "perifeer" en niet familiaal verwant lid van het huishouden betreft. Anderzijds wordt in grote gezinnen dan weer sneller een (eventueel nieuwe) verantwoordelijke voor de telling bereikbaar gevonden dan in kleine. De informatie over jonge kinderen wordt hoe dan ook door anderen ingevuld, en ontbreekt dus slechts wanneer geen "plaatsvervangers" gevonden kunnen worden. Hoe dan ook brengt het verzamelen van informatie via derden steeds een verhoogd risico op vertekening of onvolledigheid met zich mee.

1.1. De correctieprocedure: ingrepen door de correctoren

Tijdens de correctieprocedure werden met behulp van een reeks geldigheidstests op de computer essentieel drie soorten fouten opgespoord en zo mogelijk verbeterd: (1) ontbrekende (blanco) antwoorden, (2) dubbelcoderingen en (3) onmogelijke, onwaarschijnlijke of inconsistente antwoordpatronen.

De eerste twee gevallen konden te maken hebben met het optisch inlezen van de telformulieren, waarbij verbeteringen naast het antwoordhokje of doorhalingen en met de hand aangebrachte correcties voor leesfouten zorgden. Dergelijke fouten konden na raadpleging van het originele telformulier meestal vrij gemakkelijk gecorrigeerd worden. Een aantal formulieren waren ook helemaal niet ingevuld. In deze gevallen werden in principe alle in het Rijksregister beschikbare gegevens overgenomen, de andere variabelen bleven ontbrekend. Konkreet werden de ontbrekende codes (behalve de "recupereerbare" Rijksregistervariabelen) in de regel op nul gezet voor de variabelen uit het individuele en gezinstelformulier, maar bleven ze blanco voor het woningformulier. Deze "record-wise" ontbrekende informatie zal in onze analyse bijzondere aandacht krijgen. Het is over alle modules van de census-vragenlijst heen de belangrijkste bron van non-respons, en daarom de voornaamste uitzondering op de exhaustiviteit die de volkstelling beoogt.

Het derde foutentype (onmogelijke, onwaarschijnlijke of inconsistente antwoordpatronen) is complexer. Er werd gebruik gemaakt van consistentie- en correctieregels die naargelang de vraag sterk konden verschillen: soms werden alle missings naar een vaste categorie gehercodeerd en soms kreeg één variabele prioriteit boven de andere. Soms ook mocht de corrector ingrijpen door uit het ganse verhaal het meest plausibele antwoord af te leiden, en tenslotte kon het ook zijn dat de corrector besliste dat het "onwaarschijnlijke" antwoordpatroon wel

degelijk het juiste was en de toelating kreeg de oorspronkelijke codering te behouden. Het correctiewerk werd uitgevoerd door ambtenaren die manueel foutenlistings doorliepen, en daarbij vraag per vraag de opgegeven corrigeer-strategie volgden.

Als basisprincipe gold de absolute prioriteit van enkele centrale (Rijksregister-) gegevens boven alle andere: het aantal huishoudens, het aantal leden van elk huishouden en de geboortedatum en het geslacht van de telplichtige moesten tijdens de correctieprocedure als prioritair en onveranderlijk beschouwd worden. Bij het verbeteren mocht desgevallend wel ingegrepen worden in de verwantschap met de referentiepersoon, de burgerlijke staat en de geboortedatum van de echtgenoot(e).

Het corrigeren van een fout verliep, vooral voor het individuele telformulier, doorgaans volgens een aantal stappen met een vaste volgorde, waarbij de vrijheidsgraad van de corrector was vastgelegd, maar stap per stap en ook naargelang de aard van het opgemeten gezins- of individuele kenmerk nogal kon variëren. In haar volle lengte kon een procedure er dan mogelijk als volgt uitzien, al werden enkel de voor het corrigeren noodzakelijke stappen ook daadwerkelijk ondernomen:

Stap 1: Het originele telformulier opzoeken en daarin de fout nakijken

Stap 2: (Indien nodig) de informatie uit het Rijksregister opvragen

Deze eerste twee stappen, waartussen vaak geen duidelijke hiërarchie gesteld werd, lieten waarschijnlijk meestal toe de correctie door te voeren. Ten gevolge van een initiële correctie moesten vaak ook de antwoorden op andere vragen (of gegevens bij andere leden van het huishouden) worden aangepast, om in overeenstemming met het gevonden gecorrigeerde profiel te blijven.

Stap 3: Specifieke correcties: enkele voorbeelden van diverse types:

- A. Indien de burgerlijke staat op geen enkele manier kan worden vastgesteld krijgt de respondent de code 1: Vrijgezel.
- B. Personen jonger dan 18 (mannen) of 15 jaar (vrouwen) kunnen slechts referentiepersoon van het huishouden zijn indien geen enkel ander lid van het huishouden daarvoor in aanmerking komt. Indien bovendien de referentiepersoon jonger is dan 14 jaar moet het huishouden gerangschikt worden onder het type ZXX: "gezin van onbekend type".
- C. Er moet tenminste 15 jaar leeftijdsverschil zijn tussen de leeftijd van de referentiepersoon en de leeftijd van zijn/haar kinderen. Corrigeren door de verwantschapsrelatie "kind van de referentiepersoon" aan te passen.

- D. (Bij dubbelcodering van vraag 2: "Ligging of bestemming van de woning"): Het juiste antwoord bepalen en navenant coderen. Bij twijfel voorrang geven aan codes L,M en N (gemengd gebruik) op code K (uitsluitend huisvesting).

Stap 4: "Forceren" (=behouden) van de code wanneer een onwaarschijnlijk antwoord toch het juiste blijkt te zijn:

- E. Indien het om een kind van de partner van de referentiepersoon (en bovendien erkend door de referentiepersoon) gaat, mag de code "kind" behouden blijven ingeval dan wel tussen de leeftijd van die partner en zijn/haar kind tenminste 15 jaar leeftijdsverschil bestaat.

Naargelang de variabele is een dergelijke optie tot "forceren" van de bestaande code soms wel, maar soms ook niet opgenomen.

In bovenstaande voorbeelden werd de bewegingsruimte van de corrector vrij strikt afgebakend. Soms zijn de aanwijzingen vager en laten ze de corrector meer vrijheid:

- F. Indien de personencategorie onbekend is: rekening houden met de antwoorden op de andere vragen en/of de leeftijd. Slechts in laatste instantie de code "onbekend" toekennen.

1.2. Geldigheidstests en de complexiteit van het consistentieprobleem

De geldigheidstests genereren "foutmeldingen" indien (combinaties van) gegeven antwoorden niet aan enkele vooropgestelde regels voldoen. Deze regels lopen naargelang het onderwerp uiteen van algemene en volstrekt onontkoombare voorwaarden enerzijds, tot wat lossere regelmaten en waarschijnlijkheden anderzijds. Soms moet de foutmelding er dus toe leiden dat de corrector een noodzakelijke aanpassing aanbrengt, maar soms gaat het ook om een simpele controle van een enigszins afwijkende maar verder volstrekt plausibele situatie. De groeiende veelvoud en complexiteit in verschijningsvormen van huishoudens en gezinnen maakt het steeds moeilijker een veralgemeenbare set van dergelijke regelmaten in gezinsposities en relaties tussen leden van hetzelfde huishouden te veronderstellen. De gehanteerde correctieprocedure is dan ook een evenwichtsoefening geweest tussen algemene regels enerzijds, en de op het terrein geobserveerde diversiteit en graad van verandering anderzijds.

Vooral het gezinstelformulier bevat een massa vragen die logisch met elkaar verbonden zijn, en die dus ter verificatie naast elkaar gelegd kunnen worden. Enkele van de (meestal erg evidente) regels die bij overtreding aanleiding zijn voor verdere controle worden hier ter illustratie opgesomd:

- Als een persoon geen vrijgezel is, dan moet hij op 01/03/91 minimum 18 jaar zijn als het een man is, en 15 jaar zijn als het een vrouw is.
- In elk gezin moet en mag er slechts 1 maal de code "referentiepersoon" voorkomen
- Schoonzoon of schoondochter van de referentiepersoon kunnen niet alleenstaand zijn
- Het leeftijdsverschil tussen broers en zussen mag nooit groter zijn dan 25 jaar
- Een particulier gezin mag niet meer dan 30 personen bevatten

1.3. Atypische huishoudens: selectieve missings en implicaties voor wetenschappelijk gebruik van volkstellingsdata

Het is duidelijk dat vooral bij huishoudens met een complexe samenstelling, zoals bij het samenwonen van personen zonder verwantschapsbanden, gezinskernen behorende tot verschillende generaties, of bij wedersamengestelde gezinnen het nagaan van dergelijke consistentieregels geen sinecure is. Vooral omdat de positie van elke persoon binnen een huishouden afgeleid moet worden van de relatie tot de referentiepersoon, zodat de keuze van de referentiepersoon cruciaal is, en de verwantschapsrelaties in atypische gezinnen dreigen moeilijk reconstrueerbaar te worden.

De non-respons is bovendien, net zoals overigens het voorkomen van foutieve codes, hoe dan ook zelden gelijkmatig verdeeld over de diverse huishoudentypes, sociale lagen of opleidingsniveaus. Daarnaast blijven ook regionale factoren steeds weer de kop opsteken als een moeilijk verklaarbare bron van variatie. Het is dus wellicht onvermijdelijk dat de nauwkeurigheid en volledigheid waarmee de census individuele en gezins- en woningkenmerken registreert varieert samen met de gezinsomvang, of de taalkennis of graad van geschooldheid van de respondenten.

We zullen daarom in de volgende paragrafen afzonderlijk voor het individuele telformulier dat door elk lid van het huishouden werd ingevuld, en het woningformulier dat slechts één keer moest ingevuld worden, nagaan door welke achtergrondkenmerken de kans op het ontbreken van een geldig antwoord beïnvloed wordt. Aangezien we met deze analyse in de eerste plaats te weten willen komen hoezeer het exhaustieve karakter van de volkstelling onder ontbrekende antwoorden te lijden heeft, gaan we eerst op zoek naar de totale "record-wise" missings: telformulieren die niet-ingevuld (of helemaal niet?) in de telbureaus toekwamen, en waarvan achteraf dus enkel de Rijksregistergegevens bekend zijn.

Het correctiewerk voor de vragen uit het individuele en gezinstelformulier maakt het bijzonder moeilijk in deze eerste vragenreeksen nog precies de originele non-respons terug te vinden, enerzijds omdat alle blanco's naar een andere (meestal niet unieke) code gehercodeerd werden, maar ook omdat in functie van de leeftijd heel wat vragen soms niet van toepassing waren op grote delen van de bevolking, en het ontbreken van een betekenisvol antwoord dus niet altijd als een missing geïnterpreteerd mag worden.

Zo is het evident dat meisjes beneden de leeftijd van 14 jaar geen vragen i.v.m. huwelijk en fertiliteit moesten beantwoorden, en dus geen inhoudelijke code voor de betreffende vragen kregen.

We zullen daarom de vragen uit het woningtelformulier gebruiken om record-missings op te sporen: de blanco's die op een ontbrekend antwoord op het originele formulier wijzen zijn er grotendeels in behouden gebleven. Op de relatie tussen het voorkomen van missings in de woningvragen en in de rest van de vragenlijst, en dus op de vraag of de woningmissings een goede proxy vormen voor de originele systematische non-respons bij de vergaring van de gegevens (weigerings, telplichtige niet bereikbaar, communicatieproblemen, etc.) wordt later dieper ingegaan.

2. NON-RESPONS IN HET WONING-TELFORMLIER EN HET INDIVIDUELE TELFORMULIER

Een 'record-missing' of 'record non-respons' betekent dat alle gegevens voor alle vragen van een persoon ontbreken. Hierbij gaat het ons om de vragen die de telplichtige zelf diende in te vullen. In de gegevens van de volkstellingen werden immers ook nog gegevens uit het Rijksregister in afzonderlijke variabelen vergenomen. Deze laatste zullen we, precies omdat ze steeds beschikbaar zijn, verder gebruiken om de achtergrondkenmerken van personen met een record-missing te achterhalen. Het gaat daarbij o.a. om de geboortedatum, huidige nationaliteit, aantal personen in het huishouden en de woonplaats. We maken een onderscheid tussen de record non-respons in het woningtelformulier en het individuele formulier. Voor beide formulieren gaan we ook kort in op de non-respons voor enkele afzonderlijke vragen. Op deze manier krijgen we ook een beeld van de non-respons die mogelijk samenhangt met de karakteristieken van de afzonderlijke vragen, wat ons zal toelaten na te gaan of de kwaliteit van gedeeltelijk ingevulde formulieren samenhangt met dezelfde kenmerken die ook het weigeren of niet ter beschikking zijn van respondenten voor het gehele formulier beïnvloeden.

2.1. Non-respons op het woningtelformulier.

Eerst en vooral moet opgemerkt worden dat voor de behandeling van de missings in het woningtelformulier het huishouden als eenheid van analyse verkozen moet worden boven het individuele lid ervan. Er is immers slechts één woningformulier ingevuld per huishouden. Willen we de relatie leggen tussen individuele kenmerken en woningkenmerken, dan kiezen we de referentiepersoon als ankerpunt. Deze laatste zal doorgaans zelf het woningformulier hebben ingevuld.

Als eerste bespreken we de "record-missings", i.e. het volledig ontbreken van alle informatie over alle woningvariabelen heen. Berekend aan de hand van een één op zeventig steekproef van (vervolledigde) huishoudens is er voor 5.2% van alle huishoudens geen enkele informatie over de woonsituatie beschikbaar. Daarmee is deze record-missing de grootste bron van non-respons in het woningtelformulier. Zouden we dit niet op huishoudensbasis maar op individuele basis opmeten, dan was slechts 4.1% van de totale populatie van dit missing-type. In kleine huishoudens ontbreekt dus vaker de woning-informatie dan in grotere. Dit heeft wellicht te maken met het feit dat wanneer in eenpersoonshuishoudens (singles) de referentiepersoon niet bereikt of bevroegd kan worden, geen ander gezinslid diens plaats kan innemen bij het invullen van het formulier. Bovendien zijn in deze kleine huishoudens jonge thuisverlaters zowel als hoogbejaarden veel sterker vertegenwoordigd: twee groepen die o.m. door hun hogere mobiliteit of door hun bijzondere woonsituatie traditioneel moeilijker te bevragen zijn.

De specifieke situatie van vooral bejaarden komt tot uiting in de zeer hoge percentages ontbrekende antwoorden in collectieve huishoudens (religieus, bejaardentehuis, rust- en verzorgingsinstellingen, etc.), namelijk ca. 45%. We zullen deze groep verder buiten beschouwing laten. Hun woningtype valt hoe dan ook vaak buiten de vaste categorieën van de woningvariabelen. Bovendien is er in dit huishoudentype geen referentiepersoon aangeduid, waarvan de kenmerken zouden helpen het voorkomen van missings te begrijpen. Tenslotte nog hebben de collectieve gezinnen in de correctieprocedure een bijzondere behandeling gekregen, die op zich al ten dele voor het ontbreken van geldige antwoorden verantwoordelijk is. Door de collectieve gezinnen uit de analyse weg te halen daalt (nog steeds op huishoudensbasis) het aandeel record-wise ontbrekende antwoorden van 5.2% naar 5.1%.

De particuliere huishoudens met een totale non-respons behoren bijna uitsluitend tot de zgn. ZXX-huishoudens: "Z", "X", "X" zijn de codes voor de eerste drie woningvariabelen die respectievelijk wijzen op (Z) particuliere huishoudens, (X) ontbrekende informatie over het woningtype en (X) ontbrekende situatie/bestemming van het gebouw. De codes ZXX worden gevolgd door een reeks van blanco's over alle later volgende woningvariabelen. Daardoor had er verwarring met andere gevallen van ontbrekende antwoorden kunnen ontstaan, maar gelukkig werd er voor de meeste variabelen ook een specifieke (niet-blanco) categorie van missings voorzien.

Het is nochtans ook mogelijk dat huishoudens met een correct ingevuld woningtelformulier achteraf toch nog de ZXX-codes krijgen. Dat is bijvoorbeeld (zij het uitzonderlijk) het geval wanneer een alleenwonende persoon, die automatisch referentiepersoon wordt, minder dan 15 jaar oud is. Het betekent evenwel ook dat het aandeel ZXX-huishoudens vermoedelijk een lichte overschatting vormt van de originele non-respons. Dat is vrijwel zeker het geval aangezien de woningmodule als laatste vragenreeks ook af en toe werd overgeslagen door respondenten die het individuele formulier eerst wel hadden ingevuld.

Zoals reeds opgemerkt heeft het ZXX-type van missings veruit het belangrijkste aandeel in de ontbrekende antwoorden: voor 22 van de 27 woningvariabelen bedraagt dit aandeel meer dan 80% (voor 16 variabelen zelfs meer dan 95%). Slechts voor 2 woningvragen is er een meerderheid van missings die niet tot het ZXX-type behoren, namelijk WQ5A: "Bouwperiode van de woning" (27%) en WQ10: "Aansluiting op het telefoonnet" (36%). Daarbij is vooral de non-respons op deze laatste, inhoudelijk erg eenvoudige vraag erg frappant. Het gebrek aan mogelijkheden om een origineel ontbrekend antwoord via andere vragen of externe bronnen te vervolledigen speelde hier misschien een rol. Dergelijke procedures bestonden voor een aantal andere vragen wel: zo werd bvb. de oppervlakte van de woning, wanneer deze niet was opgegeven, bij benadering afgeleid uit het aantal en het type van de woonvertrekken, waarnaar al in een andere reeks vragen geïnformeerd werd. Op de non-respons op deze laatste twee afzonderlijke vragen (item non-respons) zullen we later ingaan. Bovendien komen ook de regionale aspecten van de non-respons nog uitvoeriger aan bod.

2.1.1. De originele non-respons, benaderd via de record-wise missing in het woningformulier

Er werd met behulp van een logistische regressie-analyse nagegaan in welke mate de kans op non-respons stijgt of daalt in functie van de positie op een aantal achtergrondvariabelen (Tabel 1). Daartoe kozen we, gezien de zeer hoge onderlinge samenhang tussen het ontbreken van informatie over verschillende variabelen, enkel achtergrondskennmerken die ook in het Rijksregister zijn opgenomen en dus steeds beschikbaar zijn. We gebruikten met name de leeftijd, het geslacht, de nationaliteit, de omvang van het huishouden en de provincie. Voor elke categorie van deze predictoren werd nagegaan in welke mate deze meer of minder aanleiding gaf tot missings dan dat over alle categorieën samen het geval is (i.e. het "Deviation contrast"). Het alternatief had erin kunnen bestaan de effecten van elke categorie te vergelijken met een vaste referentiecategorie ("Indicator contrast").

Kijken we enkel naar de bruto-effecten, d.w.z. zonder rekening te houden met het feit dat de verklaring die één variabele biedt soms ook al in (een combinatie van) de andere vervat is, dan hebben alle predictoren een

Tabel 1: Logistische regressieanalyse van record-wise missings in het woningformulier

| Variabele + categorie | Sig. bruto-effect (5.1%) | R bruto-effect | Exp(B) bruto-effect | Sig. netto-effect | R. netto-effect | Exp(B)netto-effect |
|-----------------------|--------------------------|----------------|---------------------|-------------------|-----------------|--------------------|
| Provincie | .0000 | .1779 | | .0000 | .1340 | |
| Antwerpen | .0000 | -.0325 | .7285 | .0000 | -.0360 | .6994 |
| Brabant | .0000 | .1336 | 2.2400 | .0000 | .0778 | 1.6516 |
| West-Vl. | .0000 | -.0417 | .5990 | .0001 | -.0247 | .7233 |
| Oost-Vl. | .0000 | -.0421 | .6262 | .0000 | -.0282 | .7183 |
| Henegouw. | .0000 | .0904 | 1.9554 | .0000 | .0764 | 1.8117 |
| Luik | .0000 | .0224 | 1.2481 | .6402 | .0000 | 1.0300 |
| Limburg | .0000 | -.0517 | .3898 | .0000 | -.0464 | .4199 |
| Luxemburg | .4311 | .0000 | .8967 | .9633 | .0000 | .9935 |
| Namen | | | 1.9152 | | | 2.1405 |
| Nationalit. | .0000 | .2401 | | .0000 | .1678 | |
| Belg | .0000 | -.2057 | .2927 | .0000 | -.1428 | .3843 |
| Ander Euro | .0001 | .0237 | 1.2077 | .0001 | .0238 | 1.2304 |
| Tur + Mar. | .7835 | .0000 | 1.0230 | .1077 | .0051 | 1.1579 |
| Rest | | | 2.7653 | | | 1.8265 |
| Leeftijd | .0000 | .2027 | | .0000 | .1617 | |
| 1 - 24 | .0000 | .1721 | 3.7338 | .0000 | .1243 | 2.8272 |
| 25 - 29 | .0000 | .0522 | 1.4367 | .0000 | .0531 | 1.4884 |
| 30 - 59 | .0000 | -.0650 | .7348 | .6188 | .0000 | 1.0174 |
| 60 - 69 | .0000 | -.1001 | .4524 | .0000 | -.0851 | .4980 |
| 70 + | | | .5608 | | | .4690 |
| Geslacht | .0000 | -.0952 | | .9984 | .0000 | |
| Man | .0000 | -.0952 | .7492 | .9984 | .0000 | 1.0000 |
| Vrouw | | | 1.3348 | | | 1.0000 |
| Omvang huish | .0000 | .2503 | | .0000 | .2176 | |
| 1 persoon | .0000 | .2358 | 3.4251 | .0000 | .2107 | 3.5820 |
| 2 personen | .0004 | -.0214 | .8522 | .6567 | .0000 | 1.0212 |
| 3 personen | .0000 | -.0382 | .7167 | .0000 | -.0408 | .6973 |
| 4 personen | .0000 | -.0407 | .6780 | .0000 | -.0450 | .6455 |
| 5 en meer | | | .7050 | | | .6074 |

duidelijk significant effect. Vooral de omvang van het huishouden, en meer specifiek het verschil tussen éénpersoonshuishoudens en alle andere types samen, speelt duidelijk een cruciale rol. Vertrekkend van een gemiddelde non-respons van 5.1% heeft het censusformulier van alleenwonenden driemaal meer kans om ontbrekend te zijn. Bovendien verliest dit verband nauwelijks aan kracht wanneer we ook de andere predictoren in rekening brengen.

Verder is ook de leeftijd van invloed: referentiepersonen van minder dan 25 jaar oud vulden bijna vier maal vaker het volledige woningformulier niet in; voor de 25 tot 29-jarigen is dat nog anderhalf maal. De 60 tot 69-jarigen toonden zich het meest bereikbaar en/of plichtsgetrouw: slechts half zo vaak als gemiddeld bleef hun woningformulier oningevuld. Mede omdat we geen collectieve gezinnen in de analyse opnamen blijft de kans op non-respons dalen met de leeftijd, al is er een lichte ommekeer waarneembaar in de allerhoogste leeftijdsgroep.

De vergelijking van diverse nationaliteiten toont dat de Belgische huishoudens een driemaal lagere non-respons hebben. De andere Europeanen dienden minder vaak een correct ingevuld formulier in dan Turkse en Marokkaanse gezinshoofden, al liggen ze allemaal vrij dicht rond het gemiddelde. Vooral de andere niet-Europese nationaliteiten vertonen een sterk verhoogde non-respons (factor 2.8). Die is echter ten dele verklaarbaar door hun leeftijds- en gezinsstructuur, en ook door hun regionale overtegenwoordiging in de meer verstedelijkte provincies .

Het geslacht speelt in onze analyse een enigszins dubieuze rol, wat vooral te maken heeft met het feit dat we enkel de referentiepersonen als vergelijkingspunt gebruiken. In gezinnen waarin een vrouw referentiepersoon is komen vaker missings voor, maar dat is volledig verklaarbaar door de structuur van deze huishoudens: het gaat vooral om kleinere huishoudens die bovendien vaak in een meer stedelijke omgeving verblijven. In Tabel 1 is te zien dat er zonder controle (bruto) een duidelijk geslachtseffect bestaat, maar dat na controle voor de andere achtergrondskennmerken het effect van de geslachtsvariabele volledig wegvalt.

We refereerden al enkele malen naar regionale verschillen en effecten van de verstedelijkingsgraad. Deze zijn zelfs in de nogal rudimentaire provinciale indeling markant aanwezig en kunnen, behalve voor de provincie Brabant, slechts marginaal aan locale verschillen in de bevolkingsstructuur toegeschreven worden. Het hogere aantal missings in Brabant was uit hoofde van de bevolkingsstructuur niet alleen verwacht, de grootteorde van het verschil viel in het licht van de concentratie van "risicofactoren" zelfs nog mee. Vooral de hoge aanwezigheid van niet-europese vreemdelingen en de urbane concentratie van zeer kleine gezinnen in en rond Brussel-hoofdstad duwden de non-respons in Brabant omhoog.

Het valt sterk op hoe vooral in de provincie Limburg beduidend minder, en in de provincies Namen en Henegouwen beduidend meer non-respons genoteerd wordt. Ook deze verschillen zijn niet tot verschillen in de

bevolkingscompositie terug te voeren. Tenzij we moeten aannemen dat er een en ander is misgelopen met de praktische organisatie van de census, lijkt het er sterk op dat er in de twee Waalse provincies een sterk afwijzende houding tegenover de volkstelling bestond.

2.1.2. Vraag-specifieke non-respons in de woningtelling: twee voorbeelden

Twee eerder geciteerde vragen uit het woningformulier hadden bovenop de record-missing nog een belangrijk aandeel eigen vraag-specifieke non-respons: de bouwperiode van de woning (Q5A: 18.5% missing) en de aansluiting op het telefoonnet (Q10: 14% missing). Uit de resultaten van een tweede logistische regressie-analyse (Tabel 2) is op te maken dat deze op een vrij gelijkaardige wijze als de totale non-respons verdeeld is over de achtergrondskennmerken. Toch vallen ook enkele verschillen op. Zo is bvb. de regionale differentiatie sterk afgezwakt. De non-respons in Brabant en vooral in Henegouwen en Namen heeft kennelijk eerder te maken met totale weigering en onbereikbaarheid (misschien ook organisatorische problemen bij de census?), dan met de moeilijkheidsgraad of het gevoelige karakter van specifieke vragen voor de betrokken respondenten. Waar echter de regionale spreiding vooral de systematische non-respons weerspiegelt, is het duidelijk dat specifieke non-respons bij uitstek met de woningkenmerken samenhangt. Zo is de slechtere (oudere) behuizing van Turkse en Marokkaanse huishoudens aanleiding voor een groter aandeel onbekende bouwperiodes van de woning. Een gelijkaardige redenering gaat wellicht ook op voor vrouwelijke gezinshoofden (referentiepersonen). We kunnen dus wel spreken van twee vormen van non-respons: de systematische non-respons die uiting geeft aan de onmogelijkheid tot deelname of mogelijk het gebrek aan bereidwilligheid bij respondenten, en anderzijds de item non-respons die eerder samenhangt met de complexiteit van bepaalde vragen, en die varieert naargelang de positie van de respondent.

2.1.3. De regionale verdeling van de non-respons

Kaart 1 geeft de gemeentelijke verdeling weer van de record-missings over het Belgische grondgebied. Aangezien het (bovendien vaak in rurale gemeenten) meestal om kleine percentages gaat, was in zowat de helft van alle gemeenten de beschikbare steekproef te klein om in de analyse betrokken te kunnen worden. De weggelaten gemeenten zijn op de figuur in het wit gebleven. Het beeld van het Waalse landsgedeelte en vooral de provincie Luxemburg en grote delen van Namen en Luik is daardoor sterk onvolledig. Toch kunnen we uit het overgebleven gedeelte nog de belangrijkste trends aflezen, en op die manier het inzicht in de regionale differentiatie nog enigszins verfijnen.

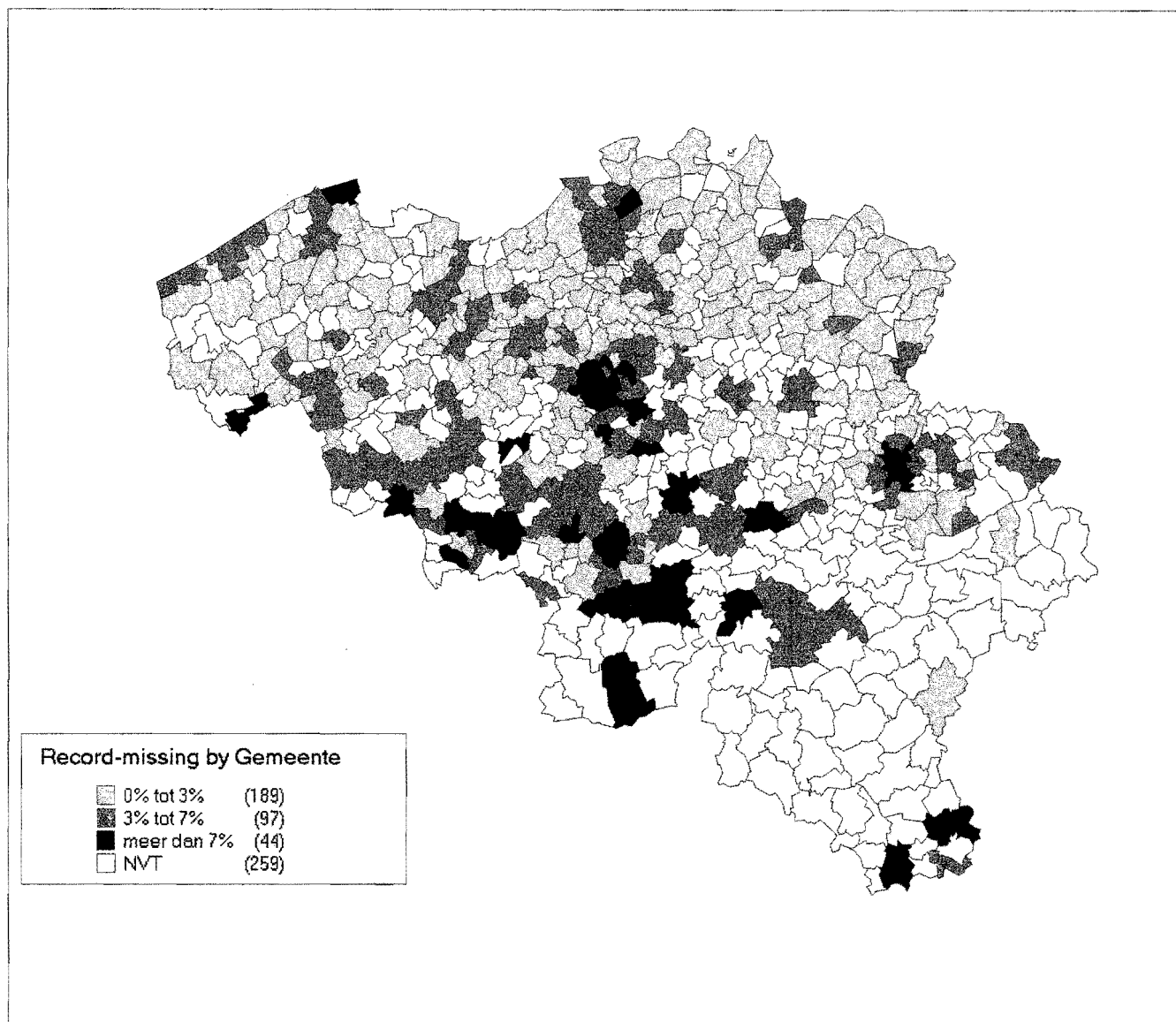
Tabel 2: Logistische regressie: record-wise missing, “bouwperiode woning” (WQ5A) missing en “aansluiting telefoonnet” (WQ10) missing tegenover elkaar geplaatst.

| Variabele + categorie | record-miss. bruto-effect (5.1% mis.) | record-miss. netto-effect | WQ5A bruto-effect (18.5% mis.) | WQ5A netto-effect | WQ10 bruto-effect (14.0% mis.) | WQ10 netto-effect |
|-----------------------|---------------------------------------|---------------------------|--------------------------------|-------------------|--------------------------------|-------------------|
| Provincie | R=.1779** | R=.1340** | R=.1261** | R=.0973** | R=.0872** | R=.0539** |
| Antwerpen | .7285 | .6994 | .8780 | .8517 | .9641 | .9584 |
| Brabant | 2.2400 | 1.6516 | 1.5276 | 1.2720 | 1.4881 | 1.2659 |
| West-Vl. | .5990 | .7233 | .7101 | .7956 | .8973 | .9888 |
| Oost-Vl. | .6262 | .7183 | .8519 | .9298 | .8328 | .9004 |
| Henegouw. | 1.9554 | 1.8117 | 1.5597 | 1.5115 | 1.2509 | 1.1670 |
| Luik | 1.2481 | 1.0300 | 1.2229 | 1.0994 | .9994 | .8985 |
| Limburg | .3898 | .4199 | .5369 | .5297 | .7647 | .8285 |
| Luxemburg | .8967 | .9935 | .8331 | .9058 | .8086 | .8455 |
| Namen | 1.9152 | 2.1405 | 1.4447 | 1.5650 | 1.2066 | 1.2604 |
| Nationalit. | R=.2401** | R=.1678** | R=.1751** | R=.1350** | R=.1259** | R=.1065** |
| Belg | .2927 | .3843 | .3272 | .3905 | .4626 | .4934 |
| Ander Euro | 1.2077 | 1.2304 | .8742 | .8864 | .9871 | 1.0082 |
| Tur + Mar. | 1.0230 | 1.1579 | 1.8072 | 2.0298 | 1.0738 | 1.2254 |
| Rest | 2.7653 | 1.8265 | 1.9345 | 1.4233 | 2.0394 | 1.6405 |
| Leeftijd | R=.2027** | R=.1617** | R=.2107** | R=.2001** | R=.0881** | R=.0459** |
| 1 - 24 | 3.7338 | 2.8272 | 3.7573 | 3.2471 | 1.9076 | 1.5433 |
| 25 - 29 | 1.4367 | 1.4884 | 1.7222 | 1.8864 | .9270 | .9505 |
| 30 - 59 | .7348 | 1.0174 | .7004 | .8713 | .7354 | .9257 |
| 60 - 69 | .4524 | .4980 | .4526 | .4611 | .7632 | .8140 |
| 70 + | .5608 | .4690 | .4875 | .4064 | 1.0076 | .9047 |
| Geslacht | R=.0952** | R=.0000 | R=.0807** | R=.0301** | R=.0804** | R=.0000 |
| Man | .7492 | 1.0000 | .8003 | .9012 | .7974 | 1.0131 |
| Vrouw | 1.3348 | 1.0000 | 1.2495 | 1.1096 | 1.2541 | .9871 |
| Omvang hh | R=.2503** | R=.2176** | R=.1593** | R=.1313** | R=.1786** | R=.1432** |
| 1 persoon | 3.4251 | 3.5820 | 1.9532 | 2.0245 | 2.2011 | 2.1704 |
| 2 personen | .8522 | 1.0212 | .8988 | 1.0568 | 1.0147 | 1.0709 |
| 3 personen | .7167 | .6973 | .8867 | .8585 | .8032 | .8221 |
| 4 personen | .6780 | .6455 | .7130 | .6959 | .7021 | .7100 |
| 5 en meer | .7050 | .6074 | .9010 | .7824 | .7940 | .7371 |

Sig. : < .05 = *

Sig. : < .01 = **

Kaart 1: Percentage record-missing per gemeente (Wit = onvoldoende observaties)



In Vlaanderen kennen de kuststrook en de driehoek Antwerpen - Gent - Brussel een verhoogde non-respons. De Antwerpse en Limburgse Kempen en het binnenland van West-Vlaanderen hebben een zeer lage non-respons. In Vlaanderen hangen de ontbrekende antwoorden vooral met de verstedelijkingsgraad samen. In het herkenbare patroon van geurbaniseerd Vlaanderen komen de centra en directe periferie van de grote steden nog eens extra duidelijk naar voren.

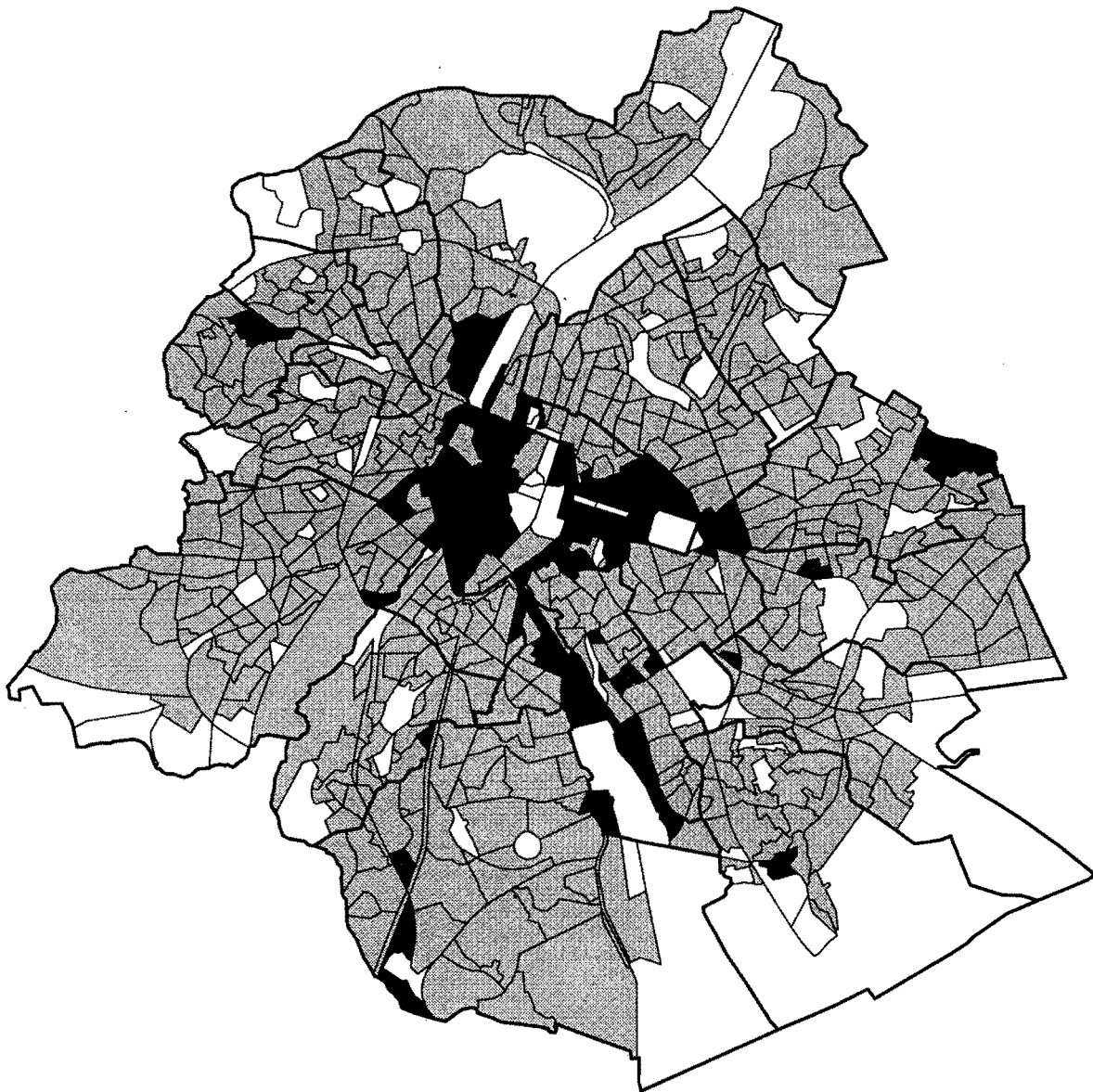
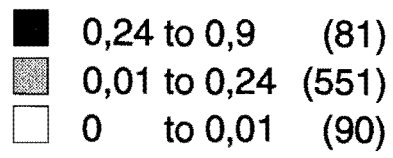
In Wallonië vinden we de non-respons vooral in de oude industrie-as van de Borinage en het Waalse Staalbekken van Samber en Maas terug. Ook gemeenten die niet direct bij de grote stadscentra aansluiten kennen opvallend veel missings. Over het ganse arrondissement Mons bedraagt de systematische non-respons meer dan 11%; in de gemeenten Dour en Saint-Ghislain vinden we extreme waarden van 35 à 40% terug. In de provincie Namen zijn ook enkele aberante gevallen terug te vinden: in Anhee waren 30% van de woningformulieren van onze steekproefhuishoudens ontbrekend, in Gesves zelfs 60%. Ter vergelijking kan worden aangegeven dat de non-respons in geen enkele Brusselse deelgemeente de 20 procent overschreed. Nochtans concentreren zich precies in die Brusselse gemeenten alle kenmerken die zoals uit de regressie-analyses bleek de kansen op non-respons sterk verhogen.

Kaart 2 geeft per statistische sector de verdeling van de non-respons over de 19 Brusselse gemeenten. Weer zijn de weggelaten sectoren (wegens te weinig steekproefgezinnen) in het wit aangegeven. Hier vormt vooral de Brusselse binnenstad de kern van de non-respons, met enkele uitlopers langs Elsene en Etterbeek naar het zuiden. Daarbij steken de sectoren die tot de gemeente Brussel behoren opvallend donkerder af tegen de omringende sectoren (voor zover het om sectoren met een zeker peil van bewoning gaat), in die mate dat de gemeentegrenzen bijna afgebakend zouden kunnen worden a.d.h.v. de percentages missings. Sint-Joost sluit volledig aan bij het Brusselse profiel. Dat geldt ook voor een deel van Molenbeek en (enkel) de kuregemwijk in Anderlecht. Schaarbeek heeft vrij behoorlijke maar toch niet extreme missing-percentages.

De gemeenten Etterbeek en Elsene scoren slechter dan men wellicht zou verwachten. Eigenaardig is ook het hoge aantal missings in vooral het zuidelijke gedeelte van Ukkel. Mogelijk hangt dit ten dele samen met de aanwezigheid van de E.G.-burgers die bijzonder mobiel en moeilijk bereikbaar zijn (cfr. kaart 3). In elk geval gaat het om sectoren waarvan de bewoners tot de allerhoogste inkomensgroepen behoren (cfr. kaart 4), wat het stelling dat de non-respons enkel bij lager geschoolde en kansarme gezinnen gesitueerd zou zijn toch wel enigszins ondergraaft. De buitenwijken van de beide Woluwes lijken, zij het wat minder uitgesproken, in dezelfde richting te wijzen. Zonder een directe koppeling van de hier aangehaalde bronnen is een definitieve konklusie evenwel nog wat voorbarig.

Kaart 2: Percentage record-missing per statistische sector (Wit = onvoldoende observaties)

Statistische sectoren met min. 100 inw.



Kaart 3: Percentage E.G.-burgers per statistische sector (Wit = onvoldoende observaties)



Percentage EG-burgers

| | | |
|---|--------------|-------|
| ■ | 25,9 to 69,6 | (50) |
| ■ | 16,5 to 25,9 | (137) |
| ■ | 9,9 to 16,5 | (201) |
| ■ | 4,6 to 9,9 | (220) |
| □ | NVT | (114) |

Kaart 4: Inkomen per statistische sector (aangiften 1993) (Wit = onvoldoende observaties)



Inkomen per statistische sector (1993)

| | | |
|---|----------------|-------|
| ■ | 1.130 to 2.020 | (80) |
| ■ | 900 to 1.130 | (206) |
| ■ | 730 to 900 | (185) |
| ■ | 520 to 730 | (141) |
| □ | NVT | (110) |

Bron : NIS
Tekening : Jo Buelens

2.2. Non-respons op het Individuele telformulier.

2.2.1. Record-missing op het individuele telformulier

De analyse van de record-missing in het individuele telformulier is minder eenvoudig door te voeren. Er komen immers geen ‘blanco’s’ voor in de dataset van het individuele telformulier. Ontbrekende gegevens staan gecodeerd als ‘nullen’. Voor bepaalde variabelen kan ‘nul’ ook een geldige code zijn. Verder werden er correcties en aanvullingen in de individuele gegevens doorgevoerd. Hierdoor is het niet zomaar mogelijk om personen die hun telformulier in feite niet invulden te identificeren. Zoals in punt 2.2.1.1. zal duidelijk worden kunnen we toch een vrij betrouwbaar beeld krijgen van de record-missings in het individuele telformulier indien we ons beperken tot de personen van 18 tot 60 jaar oud. In punt 2.2.1.2. gaan we na met welke achtergrondkenmerken deze record-missings samenhangen en onderzoeken we de overeenkomsten en verschillen tussen het voorkomen van de record non-respons in het individueel- en het woonformulier.

2.2.1.1. Identificatie van de record-missings in het individuele telformulier.

Een record-missing op het individuele telformulier kan bestaan uit een rij ‘nullen’. We houden ook rekening met het voorkomen van ‘gecorrigeerde record-missings’. Deze vorm van record non-respons betekent dat bepaalde vragen voor deze persoon achteraf door het NIS werden gecorrigeerd of aangevuld vanuit het rijksregister, maar waarvoor bij alle andere vragen geen geldige antwoorden zijn terug te vinden. Als dusdanig zijn zij te beschouwen als personen waarvoor de gegevens vrijwel geheel ontbreken. Op basis van het instructieboek voor de correctoren en de aard van de codes die expliciet op het ontbreken (onbekend) van de gegevens wijzen, konden we deze originele record-missings achterhalen. Hieronder zal echter duidelijk worden dat we niet voor alle gevallen exact kunnen nagaan of er sprake was van een oorspronkelijk blanco telformulier. Personen waarvoor de enige vragen die zij dienden te beantwoorden kunnen worden afgeleid uit het rijksregister zijn niet meer als ‘origineel’ missing te identificeren.

Er zijn twee variabelen waarvan we weten dat een geldige code mogelijk afkomstig is van een correctie aan de hand van het rijksregister. Dit is ten eerste de verblijfplaats van de moeder bij geboorte (missing vervangen door geboorteplaats rijksregister) en ten tweede de nationaliteit bij geboorte (missing vervangen door (1)huidige nationaliteit/ (2)geboorteland/ (3)Belg).

De manier waarop de ontbrekende data voor de personencategorie werden gehercodeerd is minder doorzichtig. De instructie zegt dat dit dient te gebeuren aan de hand van de andere mogelijke antwoorden en/of de leeftijd, en dat men slechts in laatste instantie de code ongekend mocht toekennen. Dit maakt dat voor kinderen, scholieren en ouderen (i.e. vrouw +60, man +65), die geen enkele vraag beantwoordden, de gegevens voor de

personencategorie van kind, schoolgaande of gepensioneerde toch nog kon worden toegewezen. Een man van 35 jaar, bijvoorbeeld, zonder enige antwoorden valt met betrekking tot zijn ontbrekende socio-economische status echter niet te corrigeren. Een blanco telformulier zal dus in het ene geval resulteren in een geldige en plausibele code voor de socio-economische status en in het andere geval ontbrekend blijven. Er is bij onze reconstructie van de originele blanco formulieren nog een bijkomend probleem. In bepaalde mate genereren gereconstrueerde codes op basis van de leeftijd en het rijksregister op hun beurt nieuwe geldige codes voor andere 'origineel' ontbrekende data. We denken hierbij ten eerste aan de sociale uitkering van een pensioen (of overlevingspensioen) voor oudere personen die als gepensionerd gehercodeerd¹ werden, of waarvan het weduwschap uit andere gegevens kon worden afgeleid. Ten tweede worden jongeren tussen de 6 en 18 jaar automatisch als schoolgaand gehercodeerd. Dit alles betekent dat kinderen, scholieren en ouderen waarvan een mogelijk blanco formulier werd ingediend niet als dusdanig zijn te reconstrueren. De variabelen die aan de hand van het rijksregister werden ingevuld zijn immers de enige vragen die zij dienden te beantwoorden.

We vergeleken de individuele record non-respons waarbij aanvullingen vanuit het rijksregister plaatsvonden met diegene waar enkel een rij 'nullen' voorkomt. Daaruit bleek dat het niet steeds duidelijk is waarom bepaalde personen aan de hand van het rijksregister vooralsnog op variabelen bepaalde geldige codes kregen toegewezen en anderen als geheel onbekend (nullen) werden behouden. De huidige nationaliteit is bijvoorbeeld voor iedereen in het rijksregister gekend. In het instructieboek staat vermeld dat bij een ongekende nationaliteit bij geboorte de huidige nationaliteit of de Belgische door de corrector dient te worden ingevuld. Voor sommigen is dit gebeurd, voor anderen dan weer niet. Het enige wat we wel kunnen vaststellen is dat er bij ouderen, kinderen en scholieren bijna geen gehele records met 'nullen' voorkomen. Bij hen werd de nationaliteit bij geboorte en de verblijfplaats van de moeder bij geboorte consequent volgens het rijksregister ingevuld. We vinden verder ook minder vrouwen terug bij diegenen met een ganse range 'nullen'. Voor vrouwen ouder dan 14 die volgens het rijksregister gehuwd zijn, werd immers de huwelijksdatum van het rijksregister gebruikt om de datum van het eerste huwelijk (Q3a) in te vullen².

Voor bepaalde categorieën van personen kon door het N.I.S. al de benodigde informatie uit het rijksregister worden gehaald. De keerzijde hiervan is dat we in onze analyse van de record missing enkel een zicht krijgen op die 'moeilijke' gevallen waarvoor niet de benodigde informatie beschikbaar was om bijvoorbeeld hun professionele status of hun scholingsniveau alsnog te corrigeren of waarbij de mogelijke correcties om één of andere redenen niet werden doorgevoerd. Dit maakt dat resultaten van de analyse over de achtergrondkenmerken van de record missing zowel beïnvloed zijn door de werkelijke kans op een missing als wel de mate waarin het verhaal voor een persoon al dan niet uit het rijksregister en andere gegevens kon worden

¹ Omdat men bijvoorbeeld uit het woningformulier kon opmaken dat de persoon in een bejaardenhoeft woont.

² Dit gebeurde blijkbaar op een automatische wijze met een (fout) computerprogramma. Hierbij werd in veel gevallen de maand ingevuld op de positie van het jaar en de dag op de positie van het maand.

opgemaakt. We beperken ons in de analyse van de record-missing naar achtergrondkenmerken dan ook tot personen met een leeftijd tussen de 18 en 60 jaar. Zij vormen de 'risico'-groep waarvan we effectief kunnen nagaan of de gegevens geheel of gedeeltelijk ontbreken. De personen met een volledige rij 'nullen' en diegenen waarvan we konden vaststellen dat hun geldige gegevens enkel bestaan uit aanvullingen vanuit het rijksregister worden verder samen genomen en beschouwd als de record non-response op het individuele telformulier.

2.2.1.2. Achtergrondkenmerken van de record-missings van het individuele telformulier.

We gingen hierboven reeds in op de aard van de record-missing van de woningen. Dit gebeurde echter op het niveau van de gezinnen. Hieronder gaan we op het niveau van de individuen na wat het profiel is van de personen die een record-missing hebben op het individuele telformulier. We doen dit aan de hand van de nationaliteit, woonplaats, leeftijd, geslacht en het aantal personen in het huishouden. In de analyse naar deze achtergrondkenmerken bakenen we de onderzoekspopulatie op een pragmatische wijze af. Personen waarvan de antwoorden die zij dienden in te vullen geheel uit het rijksregister zijn op te maken werden niet in de analyse opgenomen. Dit zijn zoals we hierboven aantoonde kinderen, scholieren en ouderen. De onderzoekspopulatie in deze analyse bestaat dus uit al de personen met een leeftijd van 18 tot 59 jaar³. We vinden voor deze groep 3,4% terug waarvan antwoorden op de vragen van het individuele telformulier geheel of grotendeels ontbreken⁴. Voor 79% van deze gevallen ontbreken ook de gegevens van de woningen. We vinden dus een 'harde kern' terug, die 2,7% van onze onderzoekspopulatie uitmaakt, waarvoor alle gegevens (woning en individueel) ontbreken⁵. Verder onderzoeken we ook het profiel van de personen die niet tot de harde kern behoren waarvan alle gegevens ontbreken. Dit zijn enerzijds diegenen waarvoor de informatie van het individuele formulier gekend is maar niet dat van de woningen en omgekeerd die personen waarvoor de informatie van de woning gekend is maar niet hun persoonlijke gegevens. Voor 0,08% ontbreken de individuele gegevens en is enkel het woningformulier bekend. Dit komt er in feite op neer dat één (ander) gezinslid het woningtelformulier invulde maar dat er voor één of meerdere personen de individuele gegevens onbekend zijn. Verder zijn er 1,5% personen waarvan wel de individuele gegevens gekend zijn maar de woninggegevens ontbreken.

In de eerste twee kolommen van tabel 3 worden de kansen op non-respons weergegeven van diegenen met een record-missing op het individuele formulier. Deze hebben in bijna 80% van de gevallen ook geheel ontbrekende

³ Om een oververtegenwoordiging van de mannen te vermijden hebben we enkel die mannen opgenomen die jonger zijn dan de wettelijke pensioensleeftijd voor vrouwen. Vrouwen boven de 60 worden immers in meerdere mate als 'op rust' gehercodeerd.

⁴ met uitzondering van de variabelen die vanuit het rijksregister konden worden geïmputeerd.

⁵ Aangezien de analyse van de record non-respons van de woningen (zie boven) op het niveau van de gezinnen gebeurde waren percentages voor de ontbrekende data hoger. We stelden immers vast dat vooral éénpersoonsgezinnen een hogere kans hebben op non-respons. Indien we de record missing voor de woningen op het niveau van de individuen berekenen vinden we uiteraard een lager totaal percentage van non-respons.

Tabel 3: Logistische regressie-analyse van record-wise missings in het individuele telformulier, uitsluitende record-wise missing op het individuele telformulier en uitsluitende record-wise missing op het woontelformulier.

| | Record missing op individueel telformulier 3,4% | | | Enkel record missing op individueel telformulier 0,8% | | | Enkel record missing op woning- telformulier 1,2% | | |
|------------------------|--|--------|--------------|---|--------|--------------|---|--------|--------------|
| | BRUTO NETTO | | | BRUTO NETTO | | | BRUTO NETTO | | |
| | Exp(B) | Exp(B) | sign 0,05 | Exp(B) | Exp(B) | sign 0,05 | Exp(B) | Exp(B) | sign 0,05 |
| PROVINCIE | | | | | | | | | |
| <i>Antwerpen</i> | 0,7291 | 0,7222 | * | 0,6985 | 0,7165 | * | 0,8243 | 0,8079 | * |
| <i>Brabant</i> | 2,7795 | 1,9590 | * | 1,5260 | 1,3747 | * | 1,3018 | 1,1104 | * |
| <i>W-VL</i> | 0,6546 | 0,8165 | * | 0,7862 | 0,8655 | * | 0,7228 | 0,7720 | * |
| <i>O-VL</i> | 0,7395 | 0,9035 | * | 0,6661 | 0,7206 | * | 0,4346 | 0,4601 | * |
| <i>Henegouwen</i> | 1,4228 | 1,1862 | * | 1,6349 | 1,5272 | * | 2,5498 | 2,5386 | * |
| <i>Luik</i> | 1,4895 | 1,1752 | * | 1,0813 | 1,0195 | * | 0,7802 | 0,7242 | * |
| <i>Limburg</i> | 0,3754 | 0,4195 | * | 0,4834 | 0,4719 | * | 0,7168 | 0,7835 | * |
| <i>Luxemburg</i> | 1,1043 | 1,3051 | * | 1,5956 | 1,6384 | * | 0,6698 | 0,6967 | * |
| <i>Namen</i> | 1,1382 | 1,3103 | * | 0,7334 | 0,8307 | * | 1,0470 | 0,9965 | * |
| NATIONALITEIT | | | | | | | | | |
| <i>Belg</i> | 0,2431 | 0,3255 | * | 0,3821 | 0,4592 | * | 0,6410 | 0,8222 | * |
| <i>Eur</i> | 1,3667 | 1,4138 | * | 0,9756 | 0,9693 | * | 0,8693 | 0,8087 | * |
| <i>etn mind</i> | 0,9042 | 0,9734 | * | 1,3721 | 1,223 | * | 1,2459 | 1,3607 | * |
| <i>Rest</i> | 3,3287 | 2,2324 | * | 1,9551 | 1,837 | * | 1,4404 | 1,1053 | * |
| LEEFTIJD | | | | | | | | | |
| <i>18-24</i> | 1,2919 | 1,4828 | * | 1,7776 | 1,7066 | * | 1,3268 | 1,3892 | * |
| <i>25-29</i> | 1,6514 | 1,4359 | * | 1,6511 | 1,6427 | * | 1,2609 | 1,2253 | * |
| <i>30-34</i> | 1,3205 | 1,3168 | * | 0,9632 | 0,9670 | * | 0,9564 | 0,9490 | * |
| <i>35-39</i> | 1,0667 | 1,0872 | * | 0,9767 | 0,9605 | * | 0,9000 | 0,8921 | * |
| <i>40-44</i> | 0,9762 | 1,0317 | * | 0,7609 | 0,7636 | * | 0,8340 | 0,8521 | * |
| <i>45-49</i> | 0,8821 | 0,9289 | * | 0,8154 | 0,8284 | * | 0,8970 | 0,9314 | * |
| <i>50-54</i> | 0,6628 | 0,6342 | * | 0,7559 | 0,7648 | * | 0,9109 | 0,9023 | * |
| <i>55-59</i> | 0,583 | 0,5398 | * | 0,7722 | 0,7938 | * | 1,0191 | 0,9690 | * |
| GESLACHT | | | | | | | | | |
| <i>Man</i> | 1,1764 | 1,1056 | * | 1,2122 | 1,2073 | * | 0,9977 | 0,9629 | * |
| <i>Vrouw</i> | 0,8501 | 0,9045 | * | 0,8249 | 0,8283 | * | 1,0023 | 1,0385 | * |
| AANTAL PERSONEN | | | | | | | | | |
| <i>1</i> | 4,6046 | 3,8638 | * | 0,9524 | 0,7997 | * | 2,9701 | 2,9518 | * |
| <i>2</i> | 0,9482 | 1,1397 | * | 1,0666 | 1,2095 | * | 0,8176 | 0,8463 | * |
| <i>3</i> | 0,5882 | 0,6412 | * | 0,8123 | 0,8713 | * | 0,6310 | 0,6323 | * |
| <i>4</i> | 0,5335 | 0,5615 | * | 0,8593 | 0,9365 | * | 0,6545 | 0,6843 | * |
| <i>5+</i> | 0,7299 | 0,6307 | * | 1,4103 | 1,2670 | * | 0,9971 | 0,9252 | * |

gegevens voor de woningen en ze vormen grotendeels de harde kern waarvoor alle gegevens onbekend zijn. In de tweede en derde kolom vindt men de kansen op een uitsluitende record non-respons van het individuele formulier en in de laatste kolommen de kans op uitsluitende non-respons van de woninggegevens.

Met betrekking tot het effect van de woonplaats kunnen we vaststellen dat alle Vlaamse provincies over de ganse lijn een vrij lage kans op non-respons vertonen, met een extreem lage kans op ontbrekende gegevens in Limburg. De provincie Brabant scoort over het algemeen vrij slecht. Hier vinden we de hoogste kans op een totale non-respons (zowel woning als individueel). Onder controle van de andere indicatoren daalt het voorkomen van geheel ontbrekende gegevens voor personen in Brabant wel enigszins. Dit wijst erop dat de hogere non-respons van deze regio gedeeltelijk te verklaren is door de bevolkingssamenstelling (vreemdelingen, alleenstaanden). Ook voor de provincie Henegouwen en Luik zijn er meer personen waarvan de individuele gegevens geheel ontbreken en kunnen we dit effect ook gedeeltelijk toeschrijven aan de bevolkingssamenstelling. Voor de provincie Henegouwen valt de uitzonderlijk hoge kans (2,5 keer meer) van een uitsluitende non-respons van het woonformulier op. We vinden daar dus een sterk regionaal netto-effect op de kans om wel de informatie voor het individuele formulier terug te vinden maar niet voor de woninggegevens. Dit bevestigt het vermoeden (zie hoger) dat er in bepaalde gemeenten mogelijk één en ander zou kunnen fout gegaan zijn bij het verzamelen van de gegevens over de woningen.

De record non-respons op het individuele formulier is voor belgen tot drie maal lager dan algemeen gemiddeld. Voor Europese vreemdelingen is het zo dat ze de hoogste kans hebben op ontbrekende data op de beide formulieren tegelijkertijd, en dat wanneer het individuele formulier werd ingevuld dit meestal ook gebeurde voor het woonformulier en omgekeerd. De kans op geheel ontbrekende data bij de Turkse en Marokkaanse populatie wijkt met uitzondering van de uitsluitende non-respons op het woonformulier niet significant af van het algemene gemiddelde. Alle overige nationaliteiten scoren beduidend slechter over de ganse lijn.

De mobiliteit van de jongeren (18 tot 30 jarigen) die zich op een andere plaats vestigen om te studeren of samen te wonen of om andere redenen tijdelijk verhuizen zonder dat dit door een wijziging van het domicilie wordt geformaliseerd veroorzaakt hun hoge record non-respons.

De vrij hoge kans op record missings voor alleenstaanden is vrij eenvoudig te verklaren. Er is immers slechts één persoon die kan instaan voor het invullen van de enquête. Het statuut van alleenstaanden is ook vaker slechts tijdelijk of fictief wat maakt dat ze minder makkelijk terug te vinden zijn op hun (voormalige) woonplaats. Waar enkel de individuele gegevens ontbreken en niet die voor de woningen is de hogere kans op ontbrekende data voor alleenstaanden niet aanwezig, omdat indien men het woonformulier als alleenstaande invulde men meestal ook wel het individuele formulier beantwoordde. Het gaat hierbij meestal om personen waarvan andere gezinsleden wel een ingevuld individueel telformulier indienden. Deze non-respons identificeert dan die personen die door hun afwezigheid in het gezin waar zij gedomicilieerd zijn een record

missing op hun individueel formulier vertonen. De ontbrekende informatie situeert zich dan ook voornamelijk bij de jongeren (18 tot 30 jaar) die (tijdelijk) niet aanwezig zijn in het gezin.

Over het algemeen geldt het besluit dat geheel ontbrekende gegevens vooral voorkomen bij alleenstaanden, jongeren, en vreemdelingen. Voor de eerste twee is de voornaamste oorzaak hiervan de moeilijke bereikbaarheid van deze groepen. Dit is ook het geval voor een gedeelte van de Europese vreemdelingen die in Brussel gevestigd zijn. We wijzen er ook op dat de effecten in tabel 4 gecombineerd kunnen worden. Een 50-tig jaar oude Belgische man van een Limburgs vierpersoonshuishouden heeft een kans van 0,14% op een record missing. Een jonge, alleenstaande Belgische man uit Brabant heeft echter 13,5% kans op het ontbreken van alle individuele gegevens en indien deze laatste een allochtoon is kan dit tot 40% oplopen. Dit wijst erop dat we toch met een sterke vertekening van de volkstellingdata in bepaalde bevolkingscategorieën dienen rekening te houden. Wanneer het gaat om een uitsluitende non-respons op het individueel formulier zijn het de jongeren die een hogere score van record missing vertonen. Dit wijst erop dat jongeren niet enkel moeilijker bereikbaar zijn indien ze zelf gezinshoofd en/of alleenstaand zijn. Hun non-respons voor de individuele gegevens is ook te verklaren door hun afwezigheid in hun (ouderlijk) gezin waarvan voor de aanwezige leden de gegevens wel gekend zijn.

De effecten van de regio van woonplaats op geheel ontbrekende individuele data zijn gedeeltelijk te verklaren vanuit de bevolkingsstructuur van de regio's (alleenstaanden, vreemdelingen, jongeren). Toch blijft er een uitzonderlijk lage non-respons (na controle) in Limburg en een uitzonderlijk hoog aantal ontbrekende gegevens van de woningen in Henegouwen.

Tenslotte willen we met een voorbeeld kort ingaan op de problemen die kunnen ontstaan door het niet expliciet opnemen van een code voor ontbrekende data. Onder de vrouwen met een record non-respons bevinden zich een aantal vrouwen waarbij de pariteit (aantal kinderen) de code 'nul' heeft en waarbij alle geboortedata van de kinderen ook de code 'nul' kregen. Hieruit zou men kunnen besluiten dat deze vrouwen nooit kinderen hebben gehad en hen op deze wijze in de vruchtbaarheidsstatistieken opnemen. Een 'nul' op pariteit en geboortedata van kinderen betekent voor deze vrouwen echter dat de informatie niet gekend is en dat, aangezien deze 'nullen' deel uitmaken van een ganse range van 'nullen' die op een record-wise non-respons wijzen, deze vrouwen mogelijk wel en mogelijk ook niet kinderen hebben gehad. Een expliciete, duidelijke en consequente codering en identificatie van ontbrekende data is zeker voor dit soort van variabelen een noodzaak.

Tabel 4: Non-respons en inconsistenties voor de variabelen van het onderwijs in België voor alle 18 tot 60 jarigen (exclusief record non-respons op het individuele telformulier en houders van buitenlandse diploma's).

| leeftijd laatste onderwijs | diploma behaald | een geldige code voor diploma in België | |
|-----------------------------------|------------------------|--|--------------|
| gekend | ja | + | 70,07% |
| <i>gekend</i> | <i>ja</i> | - | 0,10% |
| <i>gekend</i> | <i>nul</i> | + | 0,01% |
| gekend | nul | - | 0,01% |
| <i>gekend</i> | <i>nee</i> | + | 5,89% |
| gekend | nee | - | 17,93% |
| <i>gekend</i> | <i>ongekend</i> | + | 0,13% |
| gekend | ongekend | - | 0,26% |
| <i>ongekend</i> | <i>ja</i> | + | 1,07% |
| ongekend | ja | - | 0,01% |
| ongekend | nul | - | 0,00% |
| ongekend | nee | - | 2,31% |
| <i>ongekend</i> | <i>ongekend</i> | + | 0,03% |
| ongekend | ongekend | - | 2,16% |
| | | | N=73495 100% |

2.2.2. Item non-respons op enkele kernvariabelen van het individuele telformulier na uitzuivering van record non-respons.

We bespraken hierboven de kenmerken van record-missings. Deze laten we verder buiten beschouwing. We gaan in dit deel na wat het aandeel is van de non-respons op enkele kernvragen van het individuele telformulier waarbij de personen met een record non-respons niet in de analyse worden opgenomen. De item non-respons die we uitzuiveren van de record missing vormt in dit deel de afhankelijke variabele. Gezien we de record-missing enkel voor de leeftijdsgroep van 18 tot 60 jarigen met zekerheid konden identificeren gebeurt de analyse van de zuivere item non-respons ook op deze onderzoekspopulatie.

We beperken⁶ ons tot vier thema's: vestigingsjaar, fertiliteit (eerste huwelijk en aantal kinderen), onderwijs en de aard van het beroep. Voor elk van deze thema's geven we zowel aan hoe we de ontbrekende data konden identificeren (en wanneer niet) en leggen we vervolgens het verband met enkele achterliggende kenmerken. Deze achtergrondkenmerken zijn: leeftijd, woonplaats, nationaliteit en geslacht. Ze zijn afkomstig uit het rijksregister en per definitie gekend.

We willen er nogmaals op wijzen dat het niet steeds mogelijk is om de 'originele' non-respons van de variabelen te weten te komen. Er werden soms correcties toegepast waarbij gebruik gemaakt werd van informatie uit andere variabelen. Zo werden bijvoorbeeld de codes voor de graad van het buitenlandse diploma nooit naar een code 'ongekend' gehercodeerd indien dat niet expliciet werd opgegeven door de respondent. Wel werd de code van de vraag -Heeft u een buitenlands diploma?- van 'ja' naar 'nee' gehercodeerd indien de respondent geen specifiek buitenlands diploma opgaf. Verder vinden we soms enige inconsistenties terug wanneer we bepaalde variabelen van de onderwijsmodule combineren (bv.: code 'Nee' geen Belgisch diploma behaald voor een persoon met een de code 'ja' voor het diploma Belgisch lager onderwijs). Deze inconsistenties zijn uiteraard geen non-respons op zich. Ze kunnen er wel toe leiden dat de gegevens voor de onderzoeker onbruikbaar worden indien men niet weet hoe men deze inconsistenties dient te interpreteren of met andere woorden aan welke variabele men de prioriteit mag geven. We zullen voornamelijk in de paragraaf die handelt over de items van het onderwijs ingaan op de correctieprocedure en de inconsistenties die er voor bepaalde variabelen voorkomen.

2.2.2.1. Vestigingsjaar in België.

Voor het jaar van vestiging in België werd een afzonderlijke code '999' voorzien die de non-respons aangeeft op deze vraag. Dit maakt voor al diegenen, zowel Belgen als vreemdelingen, die in het buitenland geboren zijn een

⁶Gezien deze thema's de kernvariabele bevatten waarop de hercoderingen en verbeteringen voor de andere variabelen grotendeels zijn gebaseerd krijgen we met deze vier topics een goed beeld van de ganse 'range' van ontbrekende data in de volkstellingsgegevens.

item non-respons van 10,7%⁷. Uit de logistische regressie-analyse (zie bijlage) bleek dat deze non-respons het hoogst is in de provincies Antwerpen en Brabant. Ook hier stellen we vast dat alleenstaanden een hogere kans hebben op een onbekend vestigingsjaar. De mate waarin een ontbrekend vestigingsjaar voorkomt verschilt niet significant naar leeftijd en geslacht indien we controleren op de andere effecten. Voor Belgen die in het buitenland geboren zijn vinden we een hogere kans op non-respons voor het vestigingsjaar dan voor vreemdelingen. Hier speelt waarschijnlijk het effect van het ‘falende geheugen’ van Belgen die als kind in het buitenland verbleven en reeds op jonge leeftijd naar België kwamen. De mate waarin de 9% non-respons op het vestigingsjaar bij vreemdelingen voor een mogelijke vertekening zorgt is niet duidelijk. We kunnen enerzijds veronderstellen dat bij diegenen met een langere verblijfsduur (oudere vreemdelingen of diegenen die als kind naar België kwamen) er ook een effect van het ‘falende geheugen’ optreedt. Anderzijds kan het zijn dat vreemdelingen met een korte verblijfsduur door taal en andere problemen er niet goed in slagen hun censusformulier in te vullen en er hierdoor een non-respons optreedt op de vraag naar het vestigingsjaar in België. Indien we ook rekening houden met de record non-respons, die bij vreemdelingen meer voorkomt (zie hoger), kan het percentage van vreemdelingen waarvan we de migratiegeschiedenis niet kunnen achterhalen voor bepaalde nationaliteiten tot meer dan 20% oplopen.

2.2.2.2. Fertiliteit

Bij het berekenen van de vruchtbaarheidscijfers wordt gebruik gemaakt van de datum van het eerste huwelijk (voor de huwelijksvruchtbaarheid), de pariteit (aantal kinderen) en de geboortejaren van de kinderen. Deze vragen vinden we in de volkstellingsenquête terug in een aparte module die van toepassing is op alle vrouwen die ouder zijn dan 14 jaar. We beperken ons hier wel tot de analyse van de non-respons voor alle vrouwen van 18 tot 59 jaar. Van hen konden we de record-missings identificeren en uit de analyse verwijderen zodanig dat ze bij deze analyse op de item non-respons voor de fertiliteit niet samenvalt met de record missing (zie hoger). Voor de datum van het eerste huwelijk vonden we drie vormen van non-respons terug. Door gebruik te maken van de huwelijksdatum van het rijksregister die in de datafile van de volkstellingen opgenomen is konden we nagaan waar er mogelijk een datum van het eerste huwelijk ontbrak. Dit is voor 0,2% van de vrouwen het geval. Ten tweede vinden we jaartallen terug voor de datum van het eerste huwelijk die gezien de leeftijd van de vrouw niet mogelijk zijn (1,2%). De reden voor deze fout ligt waarschijnlijk bij een automatische imputering van de datum van het huwelijk uit het rijksregister, wanneer deze van het eerste huwelijk door de vrouw niet werd opgegeven. Hierbij werd in bepaalde gevallen de maand van het huwelijk op de positie van het jaar van het eerste huwelijk gekopieerd. Ten derde vinden we voor 1,1% van de vrouwen de code ‘99’ voor het huwelijksjaar terug wat wijst op een expliciete codering van de non-respons op deze variabele. Voor de pariteit is er een item non-respons van 1,2 %, en voor 0,9% van de vrouwen is minimaal de geboortedatum van één kind ongekend. In

⁷ record-missing niet inbegrepen.

totaal vinden we 2,6% van de vrouwen met minimaal één non-respons op de drie vragen die exclusief voor hen bestemd zijn.

Voor deze ontbrekende gegevens op de variabelen die handelen over de vruchtbaarheid gingen we ook na hoe ze wordt beïnvloed door enkele achtergrondkenmerken (zie bijlage). Weerom blijkt dat we een significant hogere kans op non-respons vinden in de provincie Brabant en een lagere non-respons in de provincie Limburg. De kans op onbekenden voor deze items stijgt met de leeftijd. Vooral bij Europese vreemdelingen vinden we een hoog aantal onbekenden op deze vragen. Opvallend is dat vrouwen in tweepersoonshuishoudens een hogere kans op non-respons vertonen in vergelijking met gezinnen met meerdere personen. Dit is voornamelijk te wijten aan het hoge aantal onbekende codes voor gescheiden vrouwen (5,8%). De kennis van de datum van het eerste huwelijk of de geboortedatum van de kinderen uit het eerste huwelijk is minder gekend door de nieuwe partner of ander persoon, die mogelijk als referentiepersoon instaat voor het invullen van de enquête in het tweepersoonshuishouden.

2.2.2.3. Onderwijs.

De vragen die peilen naar het onderwijsniveau behoren tot de meest cruciale voor onderzoek dat met behulp van de volkstellingsgegevens verricht wordt. Het onderwijsniveau is immers één van de belangrijkste indicatoren die bij de meeste van de vraag- en probleemstellingen naar voor komen. We staan dan ook iets meer uitvoerig stil bij de non-respons op deze variabelen. Gezien de verschillende wijze waarop de vragen die betrekking hebben op de binnen- en buitenlandse diploma's werden gecorrigeerd en om de voorstelling van onze bevindingen nietodeloos te compliceren behandelen we hieronder beide afzonderlijk.

Belgische diploma's

In tabel 4 geven we de verschillende bronnen van de ontbrekende of mogelijk foutieve gegevens aan op de variabelen van onderwijs in België. Daarbij maken we ten eerste een onderscheid tussen een al dan niet gekende leeftijd van het laatste onderwijs. Ten tweede welke code voor het al dan niet behalen van een diploma in België ingevuld is en ten derde of we effectief een geldige code van een diploma terugvinden. We vinden in tabel 4 drie types van combinaties terug. Een eerste type is de combinatie waarbij de informatie correct is. Een tweede type bestaat uit inconsistenties in de informatie (schuin) en een derde type zijn dan die gevallen waarvan de gegevens geheel of gedeeltelijk ontbreken (vet).

Voor 88% vinden we alle informatie terug. De leeftijd van het laatste onderwijs is gekend (kan nul zijn voor diegenen die geen diploma behaalden) en we weten of ze al dan niet een diploma in België behaalden en zo ja welk diploma dit dan is. Het probleem dat zich bij deze gevallen stelt is de correspondentie tussen de leeftijd van het laatste onderwijs en het behaalde diploma. De correctie die met betrekking tot de leeftijd van het onderwijs

werd doorgevoerd bestond erin dat deze werd opgetrokken indien het diploma inconsistent (te laag) was volgens de leeftijd van het laatste onderwijs. Anderzijds werd er geen correctie doorgevoerd naar een maximum leeftijdsgrens. Hierdoor vinden we nogal wat personen terug met een leeftijd van onderwijs die hoger is dan 14 zonder dat ze een diploma behaalden (40% van diegenen zonder diploma). Uiteraard is het niet mogelijk zomaar uit de eindleeftijd van het onderwijs het diploma vast te stellen en werd deze correctie terecht niet doorgevoerd. Door het hoge aandeel van een relatief hoge leeftijd van het laatste onderwijs hebben we echter de indruk dat heel wat mensen een diploma van lager onderwijs niet als een diploma opgeven. We kunnen hierdoor aannemen dat een gedeelte van deze personen toch minimaal het lager onderwijs afronden.

In beperkte mate vinden we ook enkele inconsistenties terug in de combinatie van de drie onderwijsvariabelen in tabel 4. Deze komen er meestal op neer dat personen gecodeerd zijn als 'heeft geen diploma in België behaald' of 'onbekend of diploma in België werd behaald' maar waarvoor we toch een geldige diploma-code terugvinden. In principe gaat het hier niet over ontbrekende gegevens of non-respons. We zullen deze inconsistenties dan ook niet opnemen in de analyse van de achtergrondkenmerken van de non-respons (zie verder). Het probleem bij deze inconsistenties is echter wel dat de gebruiker de geldige codes slechts kan recupereren indien men weet welke variabelen bruikbaar zijn. Zo vinden we voor diegenen waarvan de leeftijd van het laatste onderwijs gekend is 5,89% inconsistenties terug waarbij een persoon enerzijds gecodeerd staat als 'heeft geen Belgisch diploma' maar waar er wel een code voor een Belgisch diploma voorkomt. De oorzaak hiervan vonden we grotendeels terug in het instructieboek. Indien het opgegeven diploma enkel een diploma van lager onderwijs was, werd de variabele die aanduidt of er een diploma in België werd behaald, niet gecorrigeerd. Op deze manier bleef voor meer dan 30% van de personen die een lager school diploma opgaven een foutieve code 'nee' bestaan op het bezitten van een Belgisch diploma. In dit geval geeft de variabele die betrekking heeft op het soort van diploma de correcte informatie en niet de filtervraag van het al dan niet hebben van een Belgisch diploma.

Tenslotte zijn er nog die combinaties waarvan de leeftijd van het laatste onderwijs onbekend is en/of onbekend is of deze personen een diploma in België behaalden. Tezamen maken ze ongeveer 5,84% van de ganse populatie uit die we hier afbakenden. Voor deze gevallen voerden we ook een logistische regressie-analyse uit om na te gaan door welke achtergrondkenmerken deze ontbrekende gegevens kunnen worden verklaard (zie bijlage).

De analyse van de non-respons voor de onderwijsvariabelen van België (zie bijlage) toont duidelijk de invloed van de leeftijd aan. Voor jongeren vinden we minder item non-respons terug. Verder is er een beperkt effect van de gezinsgrootte. Hieruit blijkt weer dat een hoger aantal personen in het gezin de kans tot een correcte invulling van de gegevens verhoogt. Naar woonplaats stellen we nogmaals vast dat de provincie Brabant het slechtst scoort en er voor Limburg meer betrouwbare informatie bestaat. Ook het effect van de nationaliteit is

overduidelijk aanwezig. Hierbij is het echter zo dat niet-Europese vreemdelingen meer ontbrekende gegevens hebben op de onderwijsvariabelen dan de Europese vreemdelingen. Indien de volkstellingsenquête door deze laatsten wordt ingevuld, wat in mindere mate het geval is (zie boven), wordt dit voor de onderwijsvariabelen op een betere manier gedaan dan door andere vreemdelingen.

Buitenlandse diploma's

In de volkstellingsenquête werden afzonderlijke vragen voorzien die peilen naar het onderwijs dat iemand in het buitenland genoot. De non-respons op de vraag of iemand al dan niet een diploma in het buitenland behaalde is voor Belgen niet na te gaan. Voor vreemdelingen waarvan het jaar van vestiging in België gekend is kunnen we opmaken of hun schoolloopbaan mogelijk in het buitenland plaatsvond. Dit indien we ervan uitgaan dat de laatste vestiging in België ook hun eerste was. We hebben voor de vreemdelingen die zich op een leeftijd van ouder dan 18 in België vestigden onderzocht wat hun mogelijke non-respons is. Voor meer dan 18% van hen is de leeftijd van het laatste onderwijs ongekend⁸. Voor diegenen waar de leeftijd van het laatste onderwijs wel gekend is en waarvan geen diploma in België is opgegeven, duiden slechts 50% aan dat ze een diploma in het buitenland behaalden. Van de anderen hebben nochtans meer dan de helft een eindleeftijd van onderwijs die hoger ligt dan 14 jaar. We kunnen hierbij veronderstellen dat het gekende onderwijspeil van de vreemdelingen slechts gedeeltelijk het werkelijke onderwijspeil weergeeft.

Voor diegenen, zowel Belgen als vreemdelingen, waarvan gekend is dat ze in het buitenland een diploma behaalden is (bijna) steeds de graad van het onderwijs gekend. In het instructieboek staat immers vermeld dat wanneer er geen geldige code is voor de graad van het buitenlandse diploma dit onverenigbaar is met het status van 'houder van een buitenlands diploma'. Indien iemand dus opgaf dat hij een buitenlands diploma behaalde maar hierbij niet de aard van dit diploma specificeerde, werd hij/zij niet langer beschouwd als iemand met een buitenlands diploma. Wel was het voor de corrector mogelijk om toch nog een graad van het buitenlands diploma toe te wijzen aan de hand van de leeftijd van het laatste onderwijs, indien deze gekend was. Dit staat dan weer in contrast met de correcties voor variabelen met betrekking tot de diploma's in België, waar een dergelijke operatie op basis van de eindleeftijd van het onderwijs niet kon worden uitgevoerd. De werkelijke non-respons voor de graad van het buitenlands diploma is gezien deze correcties dan ook niet gekend.

Conclusies

Over het algemeen kunnen we voor de variabelen van het onderwijs besluiten dat de foutenmarge die ontstaat door de non-respons en de inconsistenties in de gegevens en correcties zeker voor bepaalde groepen

⁸ Dit is 13% meer dan bij de belgen.

(vreemdelingen en ouderen) vrij hoog kan oplopen. Waarschijnlijk situeert de non-respons zich bij personen die geen of slechts een laag onderwijsniveau behaalden.

2.2.2.4. Beroepspositie.

De analyse van de non-respons met betrekking tot de beroepsvariabelen beperken we tot de non-respons op de vraag naar de aard van dit beroep. De non-respons hierop bedraagt over het algemeen 5,5% van al diegenen die opgeven een beroepsactiviteit uit te oefenen. Gezien de vraag over het beroep een open vraag is, is het mogelijk dat de code 'beroep ongekend' ook afkomstig is van personen die een slechte of zeer vage omschrijving gaven van hun beroepsactiviteiten.

De hoogste kans op een non-respons op de beroepsvariabelen ligt bij ouderen, alleenstaanden, vrouwen en niet-Europese vreemdelingen. Ook voor gezinnen met meer dan 4 personen vinden we een relatief grotere kans op non-respons. De kans dat in die gezinnen de gegevens van personen door iemand anders in het gezin werden ingevuld is hoger. Dit kan dan resulteren in een slechte omschrijving van de beroepsactiviteit. Gecontroleerd voor deze effecten vinden we echter nog een zeer sterk effect terug van de woonplaats op de non-respons van het beroep. In de provincie Henegouwen is deze kans voor bijna alle arrondissementen 2 maal hoger dan men algemeen gemiddeld in België waarneemt. Ook in de provincie Namen, waar we in het arrondissement Namen een tot 2 maal hogere kans op non-respons vaststellen, is dit regionale effect sterk aanwezig. In Limburg daarentegen stellen we tot 2 maal minder onbekende beroepen vast dan algemeen gemiddeld in België voorkomt. Voor de provincie Brabant waar de non-respons voor andere variabelen over het algemeen vrij hoog ligt is de non-respons van de beroepspositie eerder beperkt (1,1 maal meer voor de provincie Brabant en 1,5 maal meer in Brussel-Hoofdstad).

2.2.2.5. Besluit

De item non-respons op de vier thema's die we behandelen geven een iets ander beeld dan de record non-respons. We zien dat de item non-respons, uitgezuiverd van de record non-respons, in bepaalde mate samenhangt met de karakteristieken van de vraag. Zo stellen we vast dat er voor Belgen een hogere non-respons is dan voor vreemdelingen voor het vestigingsjaar in België indien ze in het buitenland geboren zijn. Deze Belgen, zo kunnen we veronderstellen, zijn in meerdere mate op zeer jeugdige leeftijd in België gevestigd en herinneren zich minder goed hun exacte vestigingsjaar. Zo vinden we ook meer non-respons op het eerste huwelijksjaar voor gescheiden vrouwen waarbij de referentiepersoon waarmee ze nu samenleven, indien hij de informatie voor de vrouw invult, meer kans heeft om het juiste antwoord schuldig te blijven.

In tegenstelling tot de record non-respons is de item non-respons voor jongeren minder hoog dan voor ouderen. Indien men de jongeren op hun domicilie-adres terugvindt is hun gegeven informatie over het algemeen

kwalitatief beter dan bij ouderen. Voor de alleenstaanden was de record non-respons vrij hoog en ook kwaliteit van de gegevens voor de afzonderlijke vragen is bij hen systematisch laag. Wanneer het aantal personen in een huishouden groter is neemt blijkbaar de kans toe dat minimaal één iemand het invullen van de volkstellingsenquête voor zichzelf en de andere leden van het gezin grondig ter harte neemt. Voor de open beroepsvraag neemt waarschijnlijk hierdoor wel de item non-respons toe. Door de hoge graad van zowel record non-respons als item non-respons bij alleenstaanden dringt zich de vraag op of deze doelgroep bij de census niet beter dient opgevolgd te worden. Hetzelfde geldt ook voor vreemdelingen.

Verder wordt het ook de noodzaak duidelijk van een expliciete en consistente codering van de 'ontbrekende' en 'ongekende' informatie. Men heeft als gebruiker van de gegevens immers geen houvast indien dezelfde code tegelijkertijd 'onbekend' of 'neen' kan betekenen al naargelang bepaalde correctieprocedures werden toegepast.

3. BESLUIT

Door het spoorwerk naar de oorsprong van ontbrekende of afwijkende codes en de bereidwillige medewerking van het N.I.S. hierin, en door de terugkoppeling van informatie via de gebruikers van toegeleverde databestanden zij we inmiddels in staat voor zowat alle variabelen vrij goed te documenteren welke factoren aanleiding zijn geweest voor het voorkomen van ontbrekende of onduidelijke codes, zodat -ook al is de precieze omvang van het probleem vaak niet bekend- tenminste de richting en een idee van de omvang van mogelijke vertekeningen kan ingeschat worden.

We gingen in dit artikel uitvoerig in op het voorkomen en de aard van de non-respons in de Volkstelling van 1991. Hierbij bleek het ook noodzakelijk om stil te staan bij de correcties en aanvullingen die door het N.I.S. in het databestand van de optisch ingelezen formulieren werden verricht. Aan de hand van onze bevindingen willen we nu even kort stilstaan bij de mogelijke gevolgen van deze non-respons en correcties.

Voor de record non-respons bleek dat het vaak ging om personen die niet op hun formele domicilie werden teruggevonden. Dit betekent dat er toch wel een vrij selectieve ondervertegenwoordiging is van bepaalde categorieën van personen; met name alleenstaanden, Europese vreemdelingen en jongeren. We kennen in België voor de volkstellingen niet het systeem, zoals in Turkije, waarbij iedereen op de dag van de Volkstellingen op de plaats moet blijven waar hij of zij zich bevindt. Dit betekent dat de meer mobiele personen, of althans diegenen die aan hun interne migratie geen formele domiciliewijziging verbinden, vaker in de gehele non-respons terecht komen. Zeker voor het hoge aantal jongeren die een record non-respons vertonen kunnen we vermoeden dat er sprake is van een selectieve uitval. Het gaat daarbij dan om jongeren die in losse en mobiele samenlevingsverbanden leven of om die zich in een overgangsfase bevinden (bv.: van samenwonen

naar huwelijk) en waarschijnlijk bevinden zich bij de record non-respons ook een aantal 'kot'-studenten. Voor de vreemdelingen bleek voornamelijk de hoge non-respons van Europese vreemdelingen problematisch te zijn. Het gaat hierbij waarschijnlijk in meerdere mate om Brusselse EG-ambtenaren of kaders van grote bedrijven of instellingen. We kunnen ons hierbij de vraag stellen in welke mate hun verblijf in België als een reële en vaste residentie gezien moet worden.

Voor de item non-respons die we onderzochten kunnen we stellen dat deze ten dele vraagspecifiek is. Verder zijn het voornamelijk de ouderen die een slechtere kwaliteit van antwoorden geven op de volkstellingen. De item non-respons is over het algemeen iets hoger bij alleenstaanden. Toch zien we dat de item respons ook voor gezinnen met meerdere personen niet altijd goed is. Dit is gedeeltelijk te wijten aan het feit dat het bij de volkstellings-enquête vaak voorkomt dat de informatie door medegezinsleden van de betrokkene wordt ingevuld. Daar waar de Europese vreemdelingen vooral slecht scoorden met betrekking tot de record non-respons is de kwaliteit van hun respons op afzonderlijke vragen beter dan voor andere vreemdelingen. Dit kunnen we toeschrijven aan de taalproblemen die voor deze laatste ontstaan bij het invullen van de formulieren. Het invullen van de volkstellings-enquête gebeurt waarschijnlijk zorgvuldiger door personen met een beter scholingsniveau en met kennis van één van onze landstalen.

Ons onderzoek heeft ook markante regionale verschillen in de non-respons aan het licht gebracht, die in enkele gevallen zelfs los staan van de plaatselijke bevolkingscompositie. Vooral de zwakke respons, beduidend slechter nog in het woningtelformulier dan in de individuele formulieren, in enkele Naamse en Henegouwse gemeenten trok onze aandacht. De selectieve uitval ten gevolge van non-respons heeft ontegensprekelijk een autonome regionale component.

Voor elke enquête geldt dat personen die moeilijker te vatten zijn in het vooropgestelde gestandaardiseerde beeld sneller uit de boot vallen. Dit is zeker ook bij de volkstellingen het geval. De correctieprocedures waarvan we er enkele illustreerden wijzen erop dat er meestal voor een vrij stereotyp antwoord werd gekozen wanneer er sprake was van non-respons of mogelijke foutieve codes. Dit is op zich perfect verdedigbaar en wordt algemeen toegepast. Toch dient de gebruiker van de gegevens zich van de aard van deze correcties bewust te zijn. Op deze manier worden immers bepaalde personen (te snel?) naar het meest plausibele antwoord gehercodeerd en blijven anderzijds selectief atypische personen en gezinnen waarvoor geen plausibel antwoord valt op te maken met een non-respons in hun gegevens achter.

Vooraf voor de wetenschappelijke gebruiker van de census-informatie blijft het feit dat moet uitgegaan worden van een reeds verregaand opgekuiste datafile een belangrijk nadeel. Hoewel de gehanteerde correctieregels over het algemeen zeer plausibel waren, blijft vooral het (onomkeerbare) dirigeren van ontbrekende antwoorden naar

een geldige maar nogal stereotype antwoordcategorie, vaak van de strekking "geen ... (vb. buitenlands diploma,...)", hoe plausibel ook, moeilijk verteerbaar.

De volkstellingen bevatten een schat aan informatie van (bijna) alle in België verblijvende personen. Onze conclusie is echter dat de gebruiker van deze gegevens niet blind mag zijn voor het feit dat de volkstelling in de eerste plaats een enquête is met al de voor- en nadelen die daaraan verbonden zijn. Eén van de nadelen is een beperkte selectieve uitval van bepaalde categorieën van personen. Een meer expliciete codering en identificeerbaarheid van de non-respons zou ons in de toekomst al zeker een stuk vooruit helpen om de draagwijdte hiervan correct in te schatten.

BIJLAGE: Logistische regressie-analyse naar de item non-respons van het vestigingsjaar, de vragen over fertilitet, de onderwijsvragen voor België, en de aard van het beroep.

| ITEM NON-Respons | Vestigingsjaar non-repons: 10,7% | | | fertilitet non-repons: 2,6% | | | Onderwijs in B. non-repons: 5,6% | | | Beroep non-repons: 5,5% | | |
|-----------------------------|-------------------------------------|--------|--------------|--------------------------------|--------|--------------|-------------------------------------|--------|--------------|----------------------------|--------|--------------|
| | BRUTO NETTO | | sign 0,05 | BRUTO NETTO | | sign 0,05 | BRUTO NETTO | | sign 0,05 | BRUTO NETTO | | sign 0,05 |
| Exp(B) | Exp(B) | Exp(B) | | Exp(B) | Exp(B) | | Exp(B) | Exp(B) | | Exp(B) | Exp(B) | |
| PROVINCIE | | | | | | | | | | | | |
| Antwerpen | 1,4573 | 1,4584 | * | 1,1226 | 1,1097 | | 0,8526 | 0,904 | * | 1,1295 | 1,1506 | * |
| Brabant | 1,2689 | 1,281 | * | 1,3992 | 1,2391 | * | 1,6302 | 1,411 | * | 1,2045 | 1,1042 | * |
| W-VL | 0,7861 | 0,7302 | | 0,9541 | 1,0157 | | 1,1975 | 1,2437 | * | 0,6654 | 0,6851 | * |
| O-VL | 0,6038 | 0,6097 | * | 0,9627 | 0,9831 | | 1,2452 | 1,2734 | * | 0,6649 | 0,6851 | * |
| Henegouwen | 1,0729 | 1,1164 | | 1,1053 | 1,0847 | | 0,926 | 0,84 | * | 2,1927 | 2,1776 | * |
| Luik | 0,8766 | 0,8927 | | 0,7252 | 0,7066 | * | 0,8528 | 0,795 | * | 0,6352 | 0,6241 | * |
| Limburg | 0,7578 | 0,8044 | | 0,8623 | 0,8457 | | 1,1398 | 1,1318 | * | 0,9649 | 0,9956 | |
| Luxemburg | 1,2469 | 1,1958 | | 0,6665 | 0,7064 | | 0,7114 | 0,7989 | * | 0,5732 | 0,5686 | * |
| Namen | 1,2821 | 1,2542 | | 1,5046 | 1,5906 | | 0,7535 | 0,8198 | * | 2,1567 | 2,1797 | * |
| NATIONALITEIT | | | | | | | | | | | | |
| Belg | 1,2173 | 1,2660 | * | 0,4788 | 0,4457 | * | 0,3198 | 0,2996 | * | 0,4697 | 0,4833 | * |
| Eur | 0,8796 | 0,9116 | | 0,6337 | 0,6202 | * | 0,8447 | 0,9210 | | 0,7267 | 0,6829 | * |
| Turken en Marokkanen | 0,9173 | 0,9390 | | 2,3049 | 2,5638 | * | 2,4273 | 2,3380 | * | 1,4502 | 1,4869 | * |
| Rest | 1,0181 | 0,9228 | | 1,4299 | 1,4110 | * | 1,5251 | 1,5501 | * | 2,0202 | 2,0377 | * |
| LEEFTIJD | | | | | | | | | | | | |
| 18-24 | 1,0482 | 1,0477 | | 0,3411 | 0,325 | * | 0,5734 | 0,531 | * | 1,0364 | 0,9893 | |
| 25-29 | 1,1666 | 1,1530 | | 0,8264 | 0,8014 | * | 0,8157 | 0,7839 | * | 0,9026 | 0,9018 | * |
| 30-34 | 1,0309 | 1,0108 | | 1,0349 | 1,0485 | | 0,9498 | 0,9640 | | 0,937 | 0,9006 | * |
| 35-39 | 1,0246 | 1,026 | | 1,1774 | 1,2037 | * | 1,0083 | 1,0169 | | 0,9084 | 0,8578 | * |
| 40-44 | 0,8177 | 0,8491 | | 1,1350 | 1,1918 | * | 1,0656 | 1,1048 | * | 0,8742 | 0,8375 | * |
| 45-49 | 0,972 | 0,9743 | | 1,2669 | 1,3136 | * | 1,1419 | 1,1901 | * | 0,9656 | 0,9784 | * |
| 50-54 | 0,9968 | 1,0121 | | 1,4105 | 1,3991 | * | 1,3214 | 1,3116 | * | 1,0834 | 1,1511 | * |
| 55-59 | 0,9772 | 0,9533 | | 1,4356 | 1,3889 | * | 1,3885 | 1,4211 | * | 1,3733 | 1,5383 | * |
| GESLACHT | | | | | | | | | | | | |
| Man | 0,9657 | 0,9916 | | nvt | nvt | | 0,9572 | 0,9514 | * | 0,8421 | 0,8091 | * |
| Vrouw | 1,0355 | 1,0085 | * | nvt | nvt | | 1,0447 | 1,0511 | * | 1,1875 | 1,2359 | * |
| AANTAL PERSONEN | | | | | | | | | | | | |
| 1 | 1,3527 | 1,3194 | * | 1,0621 | 1,0289 | | 1,1547 | 1,0958 | * | 1,2660 | 1,2010 | * |
| 2 | 1,0298 | 1,0212 | | 1,2629 | 1,2724 | * | 0,9739 | 0,9492 | | 0,8774 | 0,8445 | * |
| 3 | 1,0777 | 1,0678 | | 0,8428 | 0,8956 | | 0,8793 | 0,9533 | | 0,8606 | 0,8745 | * |
| 4 | 0,7605 | 0,7693 | * | 0,8457 | 0,9171 | | 0,8311 | 0,9312 | * | 0,8898 | 0,9590 | |
| 5+ | 0,8759 | 0,9034 | * | 1,0460 | 0,9300 | | 1,2168 | 1,0830 | | 1,1756 | 1,1757 | |