

Steunpunt Welzijn, Volksgezondheid en Gezin

**Kwetsbare gezinnen in Vlaanderen.
Evidentie en aanbevelingen op basis van de gezinsenquête 2016.**

Methodologisch rapport

Laure-lise Robben
Prof. dr. Wim Van Lancker



Steunpunt Welzijn, Volksgezondheid en Gezin
 Minderbroedersstraat 8 – B-3000 Leuven
 Tel 0032 16 37 34 32
 E-mail: swvg@kuleuven.be
 Website: <http://www.steunpuntwvg.be>



Publicatie nr. 2020/06

SWVG-Rapport 37/1

Steunpunt Welzijn, Volksgezondheid en Gezin

Titel rapport: Kwetsbare gezinnen in Vlaanderen. Evidentie en aanbevelingen op basis van de gezinsenquête 2016. Een methodologische nota

Promotor: Prof. dr. Wim Van Lancker¹

Onderzoekers: Laure-lise Robben¹

¹ Centrum voor Sociologisch Onderzoek, KU Leuven

Dit rapport kwam tot stand met de steun van de Vlaamse Overheid. In deze tekst komen onderzoeksresultaten van de auteur(s) naar voor en niet die van de Vlaamse Overheid. De Vlaamse Overheid kan niet aansprakelijk gesteld worden voor het gebruik dat kan worden gemaakt van de meegedeelde gegevens.

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt zonder uitdrukkelijk te verwijzen naar de bron.

No material may be made public without an explicit reference to the source.

Het consortium SWVG

<i>Deelnemende instellingen - Dagelijks bestuur</i>	
KU Leuven	
· LUCAS, Centrum voor zorgonderzoek en consultancy	Prof. dr. Chantal Van Audenhove
· Instituut voor Sociaal Recht	Prof. dr. Anja Declercq
· HIVA - Onderzoeksinstituut voor Arbeid en Samenleving	Prof. dr. Johan Put
· Leuvens Instituut voor Gezondheidsbeleid	Prof. dr. Tine Van Regenmortel
· Faculteit Economie en Bedrijfswetenschappen	Prof. dr. Erik Schokkaert
UGent	
· Vakgroep Innovatie, Ondernemerschap en Dienstenmanagement	Prof. dr. Paul Gemmel
· Vakgroep Publieke Governance, Management & Financiën	Prof. dr. Joris Voets
· Vakgroep Volksgezondheid en Eerstelijnszorg	Prof. dr. An De Sutter
Vrije Universiteit Brussel	
· Faculteit Psychologie en Educatiewetenschappen	Prof. dr. Dominique Verté
Universiteit Antwerpen	
· CRESC, UAntwerpen	Prof. dr. Peter Raeymaeckers
<i>Partnerinstellingen</i>	
Universiteit Hasselt	
· Centrum voor Statistiek	
Thomas More Hogeschool	
· Opleiding Toegepaste Psychologie	Dr. Peter De Graef
Arteveldehogeschool	
· Dienst onderzoek en dienstverlening	
Hogeschool Gent	
· Faculteit Mens en Welzijn	
UC Leuven-Limburg	
· Groep Gezondheid en Welzijn	
<i>Coördinatie en secretariaat</i>	
Coördinator SWVG	Dr. Kathleen De Cuyper
Administratieve ondersteuning	Manuela Schröder Lut Van Hoof

Rapport 37/1

Juni 2020

**Kwetsbare gezinnen in Vlaanderen. Evidentie en aanbevelingen op basis van de gezinsenquête 2016.
Een methodologische nota**

Onderzoekers: Laure-lise Robben¹
Promotor: Prof. dr. Wim Van Lancker¹

¹ Centrum voor Sociologisch Onderzoek, KU Leuven

Samenvatting

De gezinsenquête streeft ernaar om de vinger aan de pols te houden bij gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen en (Nederlandstaligen in) het Brussels Hoofdstedelijk Gewest. De voorliggende studie werd opgezet vanuit de bezorgdheid dat de eerste golf van de gezinsenquête in 2016 onvoldoende representatief is om uitspraken te kunnen doen over kwetsbare gezinnen. In dit rapport presenteren we een literatuurstudie en een statistische analyse naar de representativiteit van de gezinsenquête op basis van een koppeling tussen de gegevens van de gezinsenquête en administratieve gegevens van het Datawarehouse Arbeidsmarkt en Sociale Bescherming. We vinden dat de gezinsenquête ondanks een lage responsgraad representatief is voor de doelpopulatie. Dat is zeker het geval voor de zogenaamde 'basis' steekproef. In het kader van de gezinsenquête werd echter ook een 'extra' steekproef van personen met een niet-EU-herkomst gerealiseerd. Deze extra steekproef zorgt voor een voldoende vertegenwoordiging van een aantal specifieke groepen van kwetsbare gezinnen (gezinnen met een leefloon, laaggeschoolden) maar zorgt terzelfdertijd voor een oververtegenwoordiging van gezinnen met een niet-EU-herkomst. Om een voldoende, representatief beeld van Vlaamse gezinnen met kinderen te verkrijgen, zeker met betrekking tot herkomst als criteria, was deze extra steekproef niet nodig geweest. Volgende edities van de gezinsenquête vragen dan ook om voldoende reflectie: wat is het onderzoeksopzet en wie is de doelpopulatie? Wat wil men bereiken met een extra steekproef (inhoudelijke verdieping en/of methodologisch vergroten van het bereik)? We suggereren om gebruik te maken van reeds bestaande datasets, zoals de kansarmoedegegevens verzameld door Kind en Gezin om een extra steekproef van kwetsbare gezinnen te trekken. We bespreken tenslotte nog de mogelijke rol van administratieve gegevens in het organiseren van een nieuwe golf van de gezinsenquête en in het vermijden van item non respons, en we stellen een aantal strategieën voor om de respons te verhogen.

Inhoud

Inleiding	9
Methodologie	13
Hoofdstuk 1 De doelpopulatie	17
1 Definiëren van 'het gezin'	17
2 Populatiegegevens	20
3 Nederlandstalige inwoners in Brussel	20
Hoofdstuk 2 Het steekproefkader en de steekproeftrekking	23
1 Het Rijksregister als steekproefkader	23
2 De steekproeftrekking	26
3 Longitudinaal onderzoek	31
4 Scenario's voor de toekomst	33
Hoofdstuk 3 De dataverzameling	39
1 De postenquête	39
2 Alternatieve modi	45
3 Mixed-mode survey	48
4 Merging of knowledge en de extra steekproef	50
Hoofdstuk 4 Respons en representativiteit	53
1 Responsgraad in de gezinsenquête	53
2 Representativiteit van de gerealiseerde steekproef	55
2.1 Representativiteit ten aanzien van de populatie	57
2.2 Representativiteit ten aanzien van de geselecteerde steekproef	60
3 Multivariate analyse	75
4 Item non-respons	78
5 Strategieën om de respons te verhogen	84
6 Kritische kanttekening bij administratieve gegevens	90
Hoofdstuk 5 Populatiegewichten	93
1 Construeren van gewichten	93
2 Weegcoëfficiënten op basis van de eerste kruistabel	96
3 Weegcoëfficiënten op basis van de tweede kruistabel	98
4 Toepassen van de gewichten	99

Hoofdstuk 6	Conclusie	101
	Referenties	105
	Bijlagen	111
1	Overzicht gevraagde gegevens per instelling uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB)	111
2	Afbakening herkomst	119
3	Overzicht construeren van gewichten (Kruistabel 1)	122
4	Overzicht construeren van gewichten (Kruistabel 2)	131

Lijst Tabellen & Figuren

Tabel 1	Aandeel personen in de doelpopulatie in Vlaanderen en Brussel (Bron: Eigen berekeningen op basis van DWH AM&SB)	21
Tabel 2	Definiëring groepen zoals gebruikt in de stratificatiemethode van de Measuring Equivalent Incomes (of MEqIN)- survey (Maniquet et al., 2018, p. 3)	27
Tabel 3	Provinciale verdeling van de gerealiseerde steekproef in de gezinsenquête naar herkomst (Bron: Eigen berekeningen via Datawarehouse Arbeidsmarkt en Sociale Bescherming)	30
Tabel 4	Overzicht scenario's design vervolgenquêtes gezinsenquête (Geïnspireerd op Frans, Mortelmans, Sels, & Forrier, 2009)	33
Tabel 5	Potentiële thema's voor jaarlijkse bevraging of modules: een voorbeeld van Bircan et al. (2017, p. 7)	43
Tabel 6	Het schillenmodel van het Permanent Onderzoek Leefsituatie (Sociaal en Cultureel Planbureau, 2007)	44
Tabel 7	Overzicht van de verkregen gegevens van respondenten en niet-respondenten voor de basis en extra steekproef (Bron: DWH AM&SB)	55
Tabel 8	Overzicht gerealiseerde en geselecteerde (basis en extra) steekproef ten aanzien van populatiegegevens (Bron: Ongewogen cijfers vanuit het DWH AM&SB)	59
Tabel 9	Overzicht gerealiseerde en geselecteerde (basis en extra) steekproef tegenover de totale gerealiseerde steekproef voor enkele sociodemografische kenmerken (Bron: Ongewogen cijfers uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming)	61
Tabel 10	Overzicht opleidingsniveau gerealiseerde steekproeven (Basis - Extra - Totaal) op basis van enquêtegegevens (Bron: ongewogen cijfers uit de gezinsenquête 2016)	64
Tabel 11	Overzicht gerealiseerde en geselecteerde (basis en extra) steekproef tegenover de totale gerealiseerde steekproef voor enkele socio-economische kenmerken (Bron: Ongewogen cijfers uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming).	65
Tabel 12	Werkintensiteit van basisrespondenten versus basis niet-respondenten (Bron: DWH AM&SB)	74
Tabel 13	Overzicht resultaten logistische regressie van de deelname van de gerealiseerde basis, extra en totale steekproef (Bron: DWH AM&SB)	76
Tabel 14	Item non respons voor een selectie aan items (Bron: ongewogen cijfers uit de gezinsenquête 2016)	78
Tabel 15	Profielen van item non respondenten voor het opleidingsniveau (Bron: ongewogen cijfers via DWH AM&SB en de gezinsenquête 2016)	79

Tabel 16	Profielen van item non respondenten voor materiële deprivatie (1: het kunnen rondkomen met het beschikbare gezinsinkomen)(Bron: ongewogen cijfers via DWH AM&SB).	80
Tabel 17	Profielen van item non respondenten voor materiële deprivatie (2: het kunnen veroorloven van een jaarlijkse vakantie)(Bron: ongewogen cijfers via DWH AM&SB en gezinsenquête 2016).	81
Tabel 18	Profiel van item niet-respondenten voor de inkomensvraag (Bron: ongewogen cijfers via DWH AM&SB en gezinsenquête 2016)	83
Tabel 19	Kruistabel 1: Aantal gezinnen met kinderen jonger dan 25 jaar naar huishoudtype, aantal kinderen per leeftijdsklasse en herkomst in Vlaanderen.	94
Tabel 20	Kruistabel 2: Aantal gezinnen met kinderen jonger dan 25 jaar naar huishoudtype en beroepsstatuut in Vlaanderen	94
Tabel 21	Voorbeeld combinatie bij het construeren van gewichten.	95
Tabel 22	Originele combinaties gezinsenquête (volledig overzicht in Bijlage 3).	97
Tabel 23	Voorbeeld toepassing vuistregel (1)	98
Tabel 24	Voorbeeld toepassing nieuwe gewichten (Bron: DWH AM&SB en gezinsenquête 2016).	100
Tabel 25	Overzicht aanvraag bij het Datawarehouse Arbeidsmarkt en Sociale Bescherming met de gegevens per instelling	112
Tabel 26	Overzicht afbakening herkomstlanden (Bron: DWH AM&SB)	119

Lijst figuren

Figuur 1	Types of survey error (Gebaseerd op Weisberg, 2009, p. 3)	13
Figuur 2	Schematisch overzicht respons	54

Inleiding

De gezinsenquête streeft ernaar om ‘het gezin’ in Vlaanderen en Nederlandstalig Brussel in kaart te brengen. De enquête peilt naar het gezinsleven, de opvoeding, de balans tussen gezin, zorg en werk, de relatie met de partner, en het gebruik van diensten specifiek voor gezinnen¹. Op basis hiervan wil het Departement Welzijn, Volksgezondheid en Gezin (WVG) de effectiviteit van verschillende beleidsinstrumenten analyseren en het gezinsbeleid op deze resultaten afstemmen. Het wetenschappelijke onderzoek ter zake is immers vaak niet gedetailleerd genoeg om de situatie van gezinnen in Vlaanderen en het Vlaamse beleid in kaart te brengen en te evalueren, in het bijzonder omwille van databeperkingen. In de Vlaamse Regionale Indicatoren (VRIND) werd in 2013 al aangegeven dat er onvoldoende indicatoren beschikbaar zijn om een *evidence based* beleid mee te ondersteunen. Om die redenen organiseerde het Departement WVG in het voorjaar van 2016 een bevraging bij gezinnen met kinderen in Vlaanderen, met aansluitend een bijkomende bevraging specifiek gericht op gezinnen met een niet-Europese herkomst. In deze bijdrage onderzoeken we in welke mate alle gezinnen met kinderen in Vlaanderen werden bereikt, en meer specifiek of de kwetsbare gezinnen met kinderen voldoende vertegenwoordigd worden in de gezinsenquête.

In deze studie staan volgende vragen centraal: slaagde men erin om met de gezinsenquête kwetsbare gezinnen te bereiken? In welke mate laat de gezinsenquête toe om de levensomstandigheden van kwetsbare gezinnen te analyseren? En op welke manier kan de maatschappelijke positie van deze kwetsbare gezinnen in vervolgenquêtes op betrouwbare wijze in kaart worden gebracht? De vaststellingen worden onderverdeeld in een inhoudelijk en in een methodologisch rapport.

In dit methodologisch rapport ontleden we stap voor stap de methodologie die men in de gezinsenquête heeft toegepast. In het eerste hoofdstuk bekijken we de doelpopulatie en in welke mate het mogelijk is om deze te bereiken. Het tweede hoofdstuk omschrijft het steekproefkader van waaruit men de steekproef heeft getrokken. We bespreken de voordelen, maar ook enkele aandachtspunten bij het hanteren van dit specifieke kader. Hierbij aansluitend bekijken we de steekproeftrekking en in welke mate alternatieven kunnen ingezet worden in het doelgericht bereiken van kwetsbare gezinnen. Het derde hoofdstuk bespreekt de dataverzamelmethode: de voor- en nadelen van het werken met een postenquête en potentiële alternatieve methodieken om de doelpopulatie in kaart te brengen. Hoofdstuk vier analyseert de respons en de representativiteit van de gezinsenquête: zowel de item non respons als de unit non respons wordt hier onder de loep genomen. We plaatsen karakteristieken van de netto steekproef (i.e. respondenten) tegenover de bruto steekproef (i.e. alle aangeschreven personen) en enkele populatiegegevens. Daaropvolgend bespreken we de meerwaarde van de extra steekproef voor deze editie van de gezinsenquête en mogelijke strategieën om de respons te verhogen. In het vijfde hoofdstuk gaan we dieper in op de weegcoëfficiënten die we construeerden aan de hand van populatiegegevens. In elk hoofdstuk eindigen we met een korte samenvatting met daarin de essentiële informatie die het desbetreffende hoofdstuk bevat. Tot slot sluiten we in Hoofdstuk 6 af met een algemene conclusie en de kern van de aanbevelingen die eerder per hoofdstuk werden beschreven.

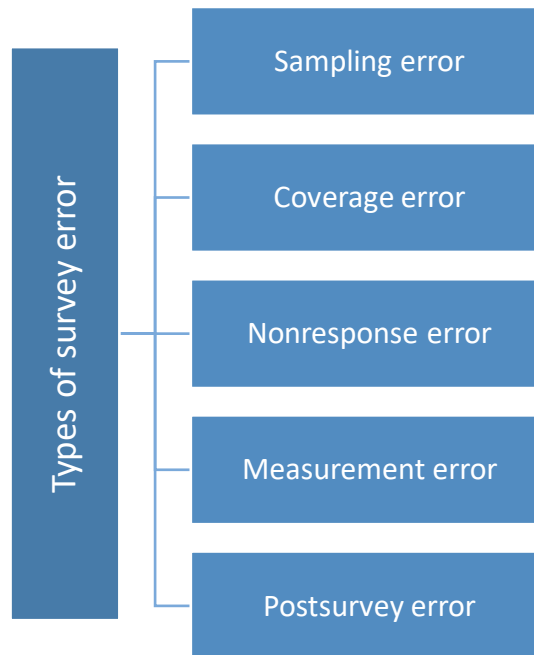
¹ Meer informatie over de onderzochte thema's via www.gezinsenquête.be

Terminologische noot

In dit rapport hanteren we enkele begrippen die vaak terugkeren maar enige uitleg behoeven. We spreken over de *geselecteerde basissteekproef* wanneer we het hebben over de initiële ‘basis’ selectie van 9.000 personen. De *gerealiseerde basissteekproef* zijn daarbij alle respondenten die effectief hebben deelgenomen aan de gezinsenquête. We spreken over de *geselecteerde extra steekproef* als de uitbreidende steekproef van personen met een niet-EU-herkomst, die in samenwerking met het Agentschap Binnenlands Bestuur werd getrokken. Daaropvolgend bestaat de *gerealiseerde extra steekproef* uit deze ‘extra’ respondenten die effectief deelnemen aan de enquête. In diezelfde context spreken we respectievelijk eveneens over de basisselectie, de basisrespondenten, de extra selectie of de extra respondenten.

Methodologie

In dit rapport baseren we ons op het Total Survey Error Model, die de representativiteit van een enquête bestudeert aan de hand van vijf aspecten (zie Figuur 1): 1) de steekproeffout (*sampling error*), 2) de dekkingfout (*coverage error*), 3) de non responsfout (*non respons error at unit and item level*), 4) de meetfout (*measurement error*) en 5) de postsurvey error.



Figuur 1 Types of survey error (Gebaseerd op Weisberg, 2009, p. 3)

De eerste drie fouten hebben te maken met het selecteren van respondenten. De eerste fout is de *steekproeffout*. Tenzij men over volledige populatiedata beschikt, zal een steekproef altijd enigszins afwijken van de werkelijke populatie. De steekproeffout is de mate waarin de steekproef afwijkt van de werkelijke populatie. De gezinsenquête is een aselechte steekproef waarin ieder lid van de populatie een gelijke kans heeft om getrokken te worden op basis van de informatie van het Rijksregister (het steekproefkader). De steekproeftrekking en de mate waarin de steekproef nauwkeurig is, namelijk hoe dicht de gezinsenquête de werkelijke populatiewaarden benadert, bespreken we in Hoofdstuk 2.3.

Ten tweede spreken we van de *dekkingfout* als de fout die optreedt wanneer de doelpopulatie geen perfecte match is met het steekproefkader (Weisberg, 2009). Concreet voor de gezinsenquête betekent dit dat men er niet in zou slagen om specifieke doelgroepen te bereiken omdat die niet opgenomen zijn in het Rijksregister (denk daarbij aan dak- en thuislozen, rondtrekkende bevolkingsgroepen, maar ook aan personen in het wachtregister). Aan deze dekkingfout schenken we aandacht in Hoofdstuk 1 en Hoofdstuk 2.1. In Hoofdstuk 2 bespreken we immers de implicaties van het afbakenen van de verkozen doelpopulatie in functie van het kunnen bestuderen van het gezinsleven in Vlaanderen en (Nederlandstalig) Brussel. In Hoofdstuk 3 analyseren we verder het gekozen steekproefkader en in welke mate deze ervoor zorgt dat bepaalde bevolkingsgroepen niet geselecteerd werden.

Ten derde verdiepen we ons in dit rapport eveneens in de potentiële non responsfout van de gezinsenquête. Deze fout treedt op wanneer de achtergrondkenmerken van de respondenten die

hebben deelgenomen aan de enquête afwijken van deze van de niet-respondenten, namelijk de gezinnen die werden aangeschreven maar niet hebben deelgenomen. Het gevolg zou kunnen zijn dat de enquête vertekend is en geen representatief beeld schetst van de populatie. De representativiteit en de respons van de basis- en extra steekproef worden nader bekeken in Hoofdstuk 4.

De vierde fout is de *meetfout*, de fout die optreedt wanneer wat men *wil* meten afwijkt van wat men effectief *heeft* gemeten. Wanneer men bijvoorbeeld de inkomens wil opvragen, maar de respondent vult deze –om welke reden dan ook- foutief in, maar je gebruikt deze alsnog in je analyses, dan is er sprake van een fout in de meting. Om deze fout op te sporen, maakt men gebruik van administratieve gegevens om enquêtegegevens te vergelijken en/of waar nodig te imputeren.

Ten slotte omschrijft de *postsurvey fout* de onregelmatigheden die optreden tijdens het verwerken of analyseren van de enquêtegegevens. Hierbij horen ook de *mode effects* (bijvoorbeeld de effecten van het kiezen van een postenquête) (Weisberg, 2009). In Hoofdstuk 3 bespreken we de postenquête als dataverzamelingsmodus, en in welke mate mogelijke alternatieve modi soelaas kunnen bieden in het bereiken van kwetsbare gezinnen.

Om deze fouten te bestuderen, maken we gebruik van administratieve gegevens van diverse overheidsinstanties. De Kruispuntbank Sociale Zekerheid (KSZ) creëerde het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB)², een instelling die administratieve gegevens van verscheidene instellingen verzamelt. Naast het aanbieden van enkele statistieken online voorziet de KSZ eveneens de mogelijkheid om een gegevensaanvraag in te dienen. Daarbij kunnen gegevens van individuen of gezinnen worden gekoppeld uit verschillende gegevensbronnen, bijvoorbeeld door gebruik te maken van gegevens uit het Rijksregister, de Rijksdienst voor Sociale Zekerheid, de Rijksdienst voor Arbeidsvoorziening, en andere instellingen van de sociale zekerheid³. Daarnaast kan men een cross-sectionele of longitudinale steekproef opvragen. Voor deze studie werd een koppeling gevraagd tussen de gegevens van de gezinsenquête met gegevens uit het Datawarehouse. Het gaat om gegevens van alle aangeschreven steekproefpersonen en hun gezinsleden, dus niet alleen van de respondenten die deelnamen aan de gezinsenquête, maar ook van de niet-respondenten. Hiervoor werd machtiging nr. 19/006 van 15/01/2019 goedgekeurd bij het Informatieveilighedscomité (IVC)⁴.

Zoals verder besproken zal worden in Hoofdstuk 2.3 heeft het werken met administratieve gegevens voordelen, maar ook enkele nadelen. Een eerste voordeel is dat het gaat om (quasi) populatiegegevens. Deze relatief complete dataset maakt het mogelijk om cijfers op te splitsen naar gezinstype, leeftijdsgroepen, en dergelijke. Maar deze bron kent ook enkele beperkingen. De gegevens, en voornamelijk de inkomensgegevens, zijn steeds wat ouder. De meest recent beschikbare inkomensgegevens die wij verkregen zijn deze van 2015. Specifiek wat betreft de inkomens beschikt men enkel over bruto belastbare inkomens, die alle looncomponenten omvat waarop gewone sociale zekerheidsbijdragen verschuldigd zijn. Waar men niet over beschikt is gegevens over het *beschikbare* inkomen, noch over inkomens uit alternatieve databronnen (denk aan inkomsten uit buitenlandse arbeid, zwart werk, enzovoort). Het gaat om beroepsinkomsten en uitkeringen, maar men houdt geen rekening met huurinkomsten, onroerend vermogen, dividenden, en dergelijke. Bovendien schiet deze informatie ook

² Meer informatie over het Datawarehouse Arbeidsmarkt en Sociale Bescherming via de website: www.ksz-bcss.fgov.be/nl/dwh/homepage/index.html

³ Meer informatie over de kenmerken (per instelling) in de aanvraag die wij opstelden voor het DWH AM&SB is terug te vinden in Bijlage A.

⁴ Online beschikbaar via www.ksz-bcss.fgov.be/nl/beraadslagingen-ivc/beraadslagingen-over-de-verwerking-van-persoonsgegevens/19006.

tekort wanneer we willen nagaan of dit inkomen voldoende hoog is voor het individu of het gezin om rond te kunnen komen. Verder verkrijgen we gegevens in inkomensklassen om de privacy van gezinnen te waarborgen, wat ervoor zorgt dat we geen exacte berekeningen kunnen maken wat betreft de totale inkomens van de steekproefpersonen.

Hoofdstuk 1

De doelpopulatie

De doelpopulatie is de populatie die we vanwege het onderzoeksopzet willen bestuderen. Voor de gezinsenquête wordt deze gedefinieerd als de gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen en (de Nederlandstaligen in) het Brussels Hoofdstedelijk Gewest. De referentiepersoon is het gezinshoofd die minstens 18 jaar is op 1 januari 2016. Deze persoon moet verblijven in een Vlaamse gemeente of in het Brussels Hoofdstedelijk Gewest (op voorwaarde dat de identiteitskaart in het Nederlands werd uitgereikt). Het kind is jonger dan 25 jaar en staat gedomicilieerd op hetzelfde adres dan van de referentiepersoon. In dit hoofdstuk staan we stil bij de gekozen afbakening van deze doelpopulatie, en beschrijven we elk component die deze definiëring naar voren schuift.

1 Definiëren van ‘het gezin’

In het midden van de 20^e eeuw overheerste de traditionele opvatting dat er sprake is van ‘een gezin’ wanneer een getrouwd koppel (meestal man en vrouw) één of meerdere kinderen heeft. Samen met de opkomst van de postindustriële samenleving wijzigden deze structuur en samenstelling. Hoewel de aanwezigheid van kinderen een belangrijk criterium blijft in het definiëren van een gezin, is het huwelijk als institutie daarbij niet meer essentieel (Hantrais, 2004). Bij de vorming van nieuw samengestelde gezinnen, na echtscheiding of partnerbreuk, kan één persoon tot meerdere gezinnen behoren. Kortom, het is niet duidelijk hoe men het gezin vandaag adequaat kan omschrijven (Luyten, 2017). In de wetenschappelijke literatuur wordt het gezin gedefinieerd op basis van een brede waaier aan relaties, leefsituaties, emotionele en financiële regelingen, terwijl tegelijkertijd rekening wordt gehouden met de subjectieve afbakening die gezinsleden zelf hanteren om ‘het gezin’ te bepalen (Harris, 2008). Veeleer gaat het vandaag om de verwantschapsrelatie tussen de gezinsleden, en de zorg die men opneemt. In dit kader vervaagt vaak de grens tussen het begrip ‘kerngezin’ en de ‘familie’ (Yorburg, 1975). Deze meer moderne notie om het gezin af te bakenen schuift niettemin de methodologische vraag naar voren hoe we het ‘gezin’ kunnen operationaliseren in een enquête die net dit gezin centraal zet.

Algemeen vinden we in de literatuur twee specifieke invalshoeken om het lidmaatschap van het gezin af te bakenen. Ten eerste is er het gezinssociologisch perspectief die benadrukt dat individuen vaak geen één lidmaatschap in één enkel gezin hebben. Intuïtief denkt men bij ‘het gezin’ aan de term ‘familie’ en heeft men het over (1) een zekere band, dan wel een bloedband, een huwelijk, een partnerschap; (2) de mate waarin men contact met elkaar heeft; en (3) een zekere rechtstreekse of onrechtstreekse ‘afstand’ (Haskey, 2010). Voor Van Crombrugge (2013) gaat het om een samenleven van personen waarbij één of meerdere personen de verantwoordelijkheid opnemen voor de zorg en opvoeding van de aanwezige kinderen. De gezinssociologische definiëring hanteert, met andere woorden, een brede afbakening, zeker in vergelijking met het economische perspectief. In deze economische invalshoek, ten tweede, spreekt men over een ‘huishouden’ als personen die daadwerkelijk een woning (of een dagelijkse maaltijd) delen. In plaats van aandacht te schenken aan de wederzijdse relatie en steun die men elkaar geeft, bekijkt men dit als een ‘economische eenheid’ die men vormt door gemeenschappelijke uitgaven. Een economische definiëring hanteert dus een meer strikte afbakening.

Administratieve databronnen vertrekken vandaag veeleer vanuit dit economische perspectief. Het Datawarehouse Arbeidsmarkt & Sociale Bescherming (DWH AM&SB), gecreëerd door de Kruispuntbank Sociale Zekerheid (KSZ) om administratieve gegevens vanuit verschillende instellingen te koppelen⁵, bakent huishoudens af op basis van het domicilie-adres. Een huishouden zijn alle personen die gewoonlijk eenzelfde woning betrekken er samenleven. Men kijkt dus naar de wettelijke hoofdverblijfplaats. Het hebben van kinderen is bijgevolg geen voorwaarde, noch het ervaren van een zekere verwantschapsrelatie met een (niet-)inwonende persoon (Datawarehouse Arbeidsmarkt en Sociale Bescherming, z.j.-b; Lodewijckx, 2001). Fiscale statistieken baseren zich dan weer op de belastingaangifte. Een ‘fiscaal huishouden’ bestaat uit alle personen die in een belastingaangifte worden vermeld, als aangever of als persoon ten laste. Deze fiscale aangiften kunnen enerzijds individueel aangewend worden of gemeenschappelijk (voor gehuwde of wettelijke samenwonende koppels). Volgens fiscale statistieken is een koppel dat feitelijk (maar niet wettelijk) samenwoont, dus niet één maar twee te onderscheiden huishoudens (BISA, 2016).

Toch moet men voorzichtig omspringen met het definiëren van het gezin op basis van administratieve gegevens. Terwijl we een verschuiving vaststellen van het klassieke gezin naar meerdere typen van gezinsstructuren, verschuift het administratief registreren van deze nieuwe samenlevingsvormen niet automatisch mee. Hoe een huishouden officieel geregistreerd is, weerspiegelt niet noodzakelijk de werkelijke leefsituatie. De klassieke administratieve gegevens bevatten dan ook geen adequate informatie over de verschillende gezinstypes. Bovendien kunnen er verscheidene transities plaats vinden in één jaar tijd, die niet altijd gecaptureerd staan in desbetreffende data. Algemeen worden administratieve gegevens beschouwd als datasets die de moderne notie van het gezin niet (kunnen) omschrijven, wat enquêtes het geschikte instrument maken om een zicht te krijgen op werkelijke leefsituaties van de gezinnen die men wil onderzoeken (Haskey, 2010; Lodewijckx & Deboosere, 2011).

Let wel, ook enquêtes passen vaak een eigen definitie van ‘het gezin’ of ‘het huishouden’ toe. De European Union Statistics on Income and Living Conditions (of de EU-SILC) definieert het huishouden als een persoon die alleen leeft, of een groep van personen die samenleven, in één gemeenschappelijke woning. Naast deze woning delen ze ook de uitgaven en voorzien ze samen in essentiële materiële behoeften (Eurostat, n.d.). De Generations and Gender Survey (of de GGS) hanteert een meer concrete definitie. Volgens hen bestaat een huishouden uit de personen die *“gedurende een normale week minstens vier dagen onder hetzelfde dak wonen* (Lodewijckx & Deboosere, 2011, p. 6)”. Wie volgens deze studie ook behoort tot het huishouden, zijn (1) kinderen die elders gedomicilieerd zijn, (2) personen die tijdelijk in de gevangenis of het ziekenhuis verblijven en (3) personen die er minder dagen zijn doordat ze ver van huis studeren of werken. Het GGS- onderzoek maakt een duidelijk onderscheid met het ‘kerngezin’, dus de biologische- of adoptieouder(s) met het kind of de kinderen. Andere personen kunnen in dit gezin samenwonen, maar niet noodzakelijk behoren tot dit kerngezin (Lodewijckx & Deboosere, 2011). Er zijn ook verschillende enquêtes die vertrekken vanuit het gezin als de eenheid die potentieel uitkeringsgerechtigd is voor specifieke gezinsuitkeringen (Haskey, 2010). In het geval van België betekent dit het recht hebben op kraamgeld, een adoptiepremie, kinderbijslag of toeslagen en/of een schoolpremie. Ook dit is een alternatieve manier om het gezin af te bakenen.

Het verschil tussen deze afbakeningen van de verschillende databronnen maakt duidelijk dat er geen wetenschappelijke consensus bestaat over de omschrijving van ‘het gezin’, noch vanuit welke invalshoek (zijnde economisch of sociologisch) we dit gezin moeten benaderen. Omdat het steekproefkader veeleer

⁵ Meer informatie over het Datawarehouse Arbeidsmarkt en Sociale Bescherming op Pagina 14.

het huishouden op economische wijze benadert, vertrekken enquêtes noodgedwongen vanuit een algemene definitie van 'het huishouden', om verder de mogelijkheid open te laten voor datagebruikers om zelf 'het gezin' te kunnen construeren. Aansluitend vinden we in de literatuur tevens bewijs dat ook 'het kind' geen eenduidige definitie kent. Geadopteerde kinderen worden, bijvoorbeeld, vaak (ook wettelijk) beschouwd als gezinsleden, terwijl pleegkinderen die op tijdelijke basis worden opgenomen vaak niet worden gezien als leden van het gezin. Het Rijksregister stelt dat een huishouden kinderen heeft wanneer het om een eigen kind of stiefkind gaat. Daarnaast kunnen het ook kinderen zijn 'zonder familieverband'. Of deze niet-verwante gezinsleden adoptiekinderen zijn, schoon- of pleegkinderen wordt niet gespecificeerd. Hoewel het Rijksregister wel verwantschapsrelaties omvat, is informatie omtrent zorgrelaties niet beschikbaar. Denk daarbij aan kinderen die mantelzorg verlenen aan hun ouders (Lodewijckx & Deboosere, 2011). De doelpopulatie van de gezinsenquête wordt gedefinieerd als de gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen en (de Nederlandstaligen in) het Brussels Hoofdstedelijk Gewest. De referentiepersoon is het gezinshoofd die minstens 18 jaar is op 1 januari 2016, en verblijft in een Vlaamse gemeente of in het Brussels Hoofdstedelijk Gewest (op voorwaarde dat de identiteitskaart werd uitgereikt in het Nederlands). Het kind is jonger dan 25 jaar en staat gedomicilieerd op hetzelfde adres dan van de referentiepersoon. Deze leeftijdskeuze wordt eveneens gehanteerd bij financiële tegemoetkomingen zoals de kinderbijslag (tegenwoordig het *Groei pakket*). Andere databronnen zoals fiscale statistieken praten over 'kinderen ten laste' die ouder kunnen zijn dan 25 jaar. De definitie van het kind is, met andere woorden, net zoals de definitie van het gezin afhankelijk van enerzijds het steekproefkader en anderzijds van de keuzes die men maakt op basis van het onderzoeksopzet.

Voor de gezinsenquête maakte het onderzoeksteam de beslissing om de leeftijdsgrens van het kind af te stemmen op de leeftijdsafbakening van het (toen nog Belgische) kinderbijslagstelsel. Om het gezin zo ruim mogelijk te benaderen, besliste het onderzoeksteam om de verwantschapsrelaties zo breed mogelijk op te nemen. 'Het kind' betekent voor de gezinsenquête dus het eigen kind, het stiefkind, het niet-verwant kind of het kleinkind (indien de grootouders instaan voor de zorg en opvoeding). Daarnaast maakte het onderzoeksteam van de gezinsenquête ook de weldoordachte keuze om enkel gezinnen met kinderen te onderzoeken. Bijgevolg is het niet mogelijk om, op basis van deze enquête, een vergelijking op te stellen van gezinnen met minstens één kind jonger dan 25 jaar en gezinnen zonder kinderen of gezinnen met oudere kinderen.

Samengevat toont de literatuurstudie aan dat enquêtegegevens er veeleer in slagen om een moderne notie van het gezinsleven te capteren, terwijl administratieve gegevens blijvend botsen op beperkingen om het gezin af te bakenen. De enquêtes zijn echter wel vaak beperkt net doordat ze vertrekken vanuit administratieve gegevens. Met het oog op volgende edities blijft het belangrijk om te reflecteren over het onderzoeksopzet en de keuzes die men in dit licht kan maken. Wil men het gezin nog breder benaderen, door bijvoorbeeld aanstaande ouders of ouders met oudere kinderen eveneens op te nemen, dan zijn andere methodologische keuzes nodig. Dit kan interessant zijn wanneer men meer informatie wil verzamelen over, bijvoorbeeld, mantelzorg, of de ondersteuning waarop aanstaande moeders beroep kunnen doen. Maar dit heeft ook methodologische implicaties, omdat het in dat geval noodzakelijk is om de vragenlijst anders te formuleren voor specifieke groepen. Hoe dan ook moet de gekozen afbakening helder omschreven zijn enerzijds en vergelijking met andere databronnen mogelijk maken anderzijds.

2 Populatiegegevens

In wat volgt bespreken we de gekende cijfers van huishoudens en gezinnen in Vlaanderen en (Nederlandstalig) Brussel. Anno 2016 wonen er in Vlaanderen 2.769.259 private huishoudens (i.e. alle personen die gewoonlijk eenzelfde woning betrekken en er samen leven). De gemiddelde huishoudgrootte is in datzelfde jaar 2,3 (Vlaamse Overheid, 2019). Het gros van de private huishoudens heeft echter geen kinderen (ten laste) en maakt niet deel uit van onderzoek naar gezinnen (Dekkers, Tarantchenko, & Van den Bosch, 2019).

Daarnaast weten we dat, van alle gezinnen met minstens één kind jonger dan 25 jaar, er 61% een gehuwd paar zijn, 21% een ongehuwd paar en 19% alleenstaand (Audenaert & Stuyck, 2018a). Op basis van de door ons gekregen geaggregeerde gegevens van het DWH AM&SB over de doelpopulatie weten we het volgende. Op 31 december 2015 zijn er in Vlaanderen 905.954 gezinnen die minstens één kind hebben dat jonger is dan 25 jaar⁶. Hiervan zijn 18,8% alleenstaande ouders, 81,2% zijn gehuwde of ongehuwde ouders. In het Brussels Hoofdstedelijk Gewest tellen we 174.382 gezinnen met een kind jonger dan 25 jaar, en een totaal van 339.073 kinderen. Brussel kent een andere verhouding met betrekking tot het gezinstype: hier zien we dat 30,7% alleenstaande ouders zijn, 69,3% zijn gehuwde of ongehuwde ouders. Deze cijfers bevestigen dat gezinnen met kinderen in Vlaanderen op socio-demografische kenmerken sterk verschillen van de gezinnen met kinderen in Brussel.

3 Nederlandstalige inwoners in Brussel

Om, met het oog op het beleid, inzicht te verkrijgen in het gezinsleven, werd de gezinsenquête uitgevoerd in Vlaanderen en het Brussels Hoofdstedelijk Gewest. Wat betreft de steekproeftrekking werd de keuze gemaakt om personen in het Brussels Hoofdstedelijk Gewest te selecteren waarvan hun identiteitskaart in het Nederlands werd uitgereikt. Maar deze selectie heeft enkele methodologische implicaties. Ten eerste is het zo dat het Rijksregister individuen of huishoudens kan selecteren op basis van deze taalcriteria, maar andere instellingen kunnen dit niet. Zo werden voor dit rapport de populatiegegevens van alle gezinnen (met minstens één kind jonger dan 25 jaar) in Vlaanderen en het gehele Brusselse Hoofdstedelijke Gewest ter beschikking gesteld. Ten tweede de taal waarin de identiteitskaart wordt uitgeraakt eigenlijk een ongeschikte manier om Nederlandstalige gezinnen met kinderen te selecteren. Het is niet noodzakelijk zo dat het deze taal is die de gezinsleden het beste beheersen noch de taal waarmee deze personen zich identificeren (Janssens, 2001; Skutnabb-Kangas & Phillipson, 1989). Ten derde streeft de gezinsenquête ernaar om inzicht te verwerven in het gezinsleven in functie van het gezinsbeleid. Maar in Brussel wijkt de uitvoering van dit gezinsbeleid af van het beleid in Vlaanderen. Bijvoorbeeld: het Huis van het Kind en Kind en Gezin zijn ook gevestigd in Brussel maar worden hier op andere wijze georganiseerd. De kinderopvanginitiatieven van Kind en Gezin zijn, bijvoorbeeld, enkel genoodzaakt om het Nederlands te verwerken in het administratieve aspect (in de statuten, arbeidsovereenkomsten, huishoudelijk reglement, enzovoort), maar het is niet noodzakelijk de taal waarin men de kinderen aanspreekt in de opvang zelf (Kind en Gezin, z.d.). Bovendien is het

⁶ Dit cijfers wijkt af van de gegevens die door de Studiedienst van de Vlaamse Regering (SVR) ter beschikking werden gesteld aan de initiatiefnemers van de Gezinsenquête. Op basis van de data aangeleverd door de KSZ zijn er 905.954 gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen; de SVR rapporteert 890.696 gezinnen, een verschil van 14.258 gezinnen. Hoewel de oorspronkelijke bron telkens het rijksregister is, worden er bij elke transfer tussen instellingen correcties doorgevoerd. Het is op het moment van schrijven onduidelijk waar de verschillen vandaan komen.

Brusselse bereik van de gezinsenquête, zeker voor de geselecteerde basissteekproef, zeer klein. Dat wordt zichtbaar in Tabel 1, waar we het aandeel van Vlaamse en Brusselse steekproefpersonen naast elkaar plaatsen.

Tabel 1 Aandeel personen in de doelpopulatie in Vlaanderen en Brussel (Bron: Eigen berekeningen op basis van DWH AM&SB)

	Vlaanderen		Brussel		Totaal	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Doelpopulatie	904.954	83,8	174.382	16,2	1.079.336	100,0
Geselecteerde steekproef	10.794	98,6	156	1,4	10.950	100,0
Gerealiseerde steekproef	2.629	98,6	37	1,4	2.666	100,0

De selectie gaat steeds om private huishoudens met minstens één kind jonger dan 25 jaar. Wat betreft de Brusselse doelpopulatie tonen we hier alle gezinnen waarvan de identiteitskaart van de referentiepersoon in het Nederlands werd uitgereikt.

Tabel 1 geeft een overzicht van het aantal bereikte personen in de drie verschillende populaties, zijnde de doelpopulatie, de geselecteerde steekproef en de gerealiseerde steekproef (het onderscheid tussen deze twee laatste groepen bespreken we verder in detail, zie Hoofdstuk 4). Zoals verwacht zien we dat de Vlaamse gezinnen de meerderheid van de doelpopulatie uitmaken: 904.954 gezinnen in Vlaanderen hebben minstens één kind jonger dan 25 jaar. In Brussel zijn dat er 174.382 (Nederlandstalige) gezinnen. Uit deze doelpopulatie selecteerde men 11.000 personen om deel te nemen aan de gezinsenquête: 98,6% woont daarvan in Vlaanderen, 1,4% in Brussel (over 50 personen heeft het DWH AM&SB geen informatie). Dit betekent dat de focus voornamelijk lag op het selecteren en bereiken van Vlaamse gezinnen. Uiteindelijk zien we dat de gerealiseerde steekproef (i.e. alle deelnemende respondenten) bestaan uit 98,6% Vlaamse gezinnen en 1,4% Brusselse gezinnen. Concreet zien we dat, van de 156 geselecteerde Nederlandstalige Brusselse gezinnen, slechts 37 respondenten uiteindelijk de gezinsenquête invulden. Het gaat om een laag absoluut aantal van gezinnen waarop men zich moet baseren om ook de Brusselse context te integreren in deze dataverzameling. Opnieuw blijft hier de vraag belangrijk welk onderzoeksopzet men naar voren schuift: wanneer de gezinsenquête een instrument wil zijn en blijven om het gezinsbeleid mee te evalueren, kan het, bijvoorbeeld, een meerwaarde bieden om ofwel zich enkel te concentreren op Vlaanderen, of om een grotere selectie en bereik na te streven in Brussel. In dat laatste geval kan het interessant zijn om een extra steekproeftrekking hierop te voorzien (zie ook Hoofdstuk 2.2 voor meer informatie omtrent alternatieve steekproeftrekkingen) en in functie daarvan een meer *aangepaste* vragenlijst te formuleren die specifiek gericht is op het gezinsbeleid in Brussel.

Samenvattend

De doelpopulatie van de gezinsenquête zijn de gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen en (de personen waarvan de identiteitskaart in het Nederlands werd uitgereikt in) het Brussels Hoofdstedelijk Gewest. Dit impliceert dat het kind gedomicilieerd staat op hetzelfde adres als dat van de referentiepersoon. In dit hoofdstuk stonden we stil bij de methodologische implicaties van deze doelpopulatie. Er ging een heel traject aan vooraf waarin het onderzoeksteam weloverwogen keuzes maakte op basis van methodologische en praktische haalbaarheid. De praktische haalbaarheid betekent voornamelijk dat men beperkt is omwille van het gekozen steekproefkader. De methodologische haalbaarheid heeft betrekking op het ontwikkelen van een uniform, heldere en generieke vragenlijst die door elk geselecteerd gezin kan worden ingevuld. Met het oog op volgende edities blijft het echter essentieel om te reflecteren over de afbakening van 'het gezin': wat is 'het gezin' waarover we meer willen weten? Welke personen moeten als gezinslid behandeld worden: alle personen die gedomicilieerd staan op één adres? Wat met, bijvoorbeeld, alleenstaande vaders waarvan het kind niet gedomicilieerd staat op zijn woning? Wat met aanstaande ouders? Het uitbreiden van deze afbakening betekent dat de doelpopulatie herzien moet worden, en dat er moet nagedacht worden over de vragenlijst: andere vragen kunnen relevanter zijn voor aanstaande ouders of mensen met een kindervens dan vragen die relevant zijn voor ouders met oudere kinderen.

Daarnaast stellen we ook vast dat men, in functie van volgende edities, eveneens moet blijven reflecteren over de rol van Brussel in dit onderzoek. Wanneer de gezinsenquête een instrument wil zijn en blijven om het gezinsbeleid mee te evalueren, kan het, bijvoorbeeld, een meerwaarde bieden om ofwel zich enkel te concentreren op Vlaanderen, of om een grotere selectie en bereik na te streven in Brussel.

Hoofdstuk 2

Het steekproefkader en de steekproeftrekking

1 Het Rijksregister als steekproefkader

Om een steekproef te trekken, vertrekt men vanuit een steekproefkader. Dit steekproefkader wordt gedefinieerd als een lijst waarin alle eenheden van de doelpopulatie opgesomd staan (De Leeuw, Hox, & Dillman, 2008). De gezinsenquête richt zich tot de gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen en (de Nederlandstaligen in) het Brussels Hoofdstedelijk Gewest. Het meest volledige steekproefkader om deze doelgroep te bereiken, zijn de bevolkingsregisters van het Rijksregister. Het Rijksregister heeft namelijk een centrale databank waarin geregistreerde gegevens vanuit de gemeenten en de Dienst Vreemdelingenzaken worden geregistreerd. Hierin vinden we (1) alle Belgen die in België verblijven; (2) Alle Belgen die in het buitenland verblijven en die zich inschreven in de bevolkingsregisters; (3) alle vreemdelingen die verblijven in België en ook de toelating hebben om in dit land te verblijven; (4) alle vreemdelingen (en familieleden) die verklaren (of vragen voor erkenning) ‘vluchteling’ te zijn (FOD Binnenlandse Zaken, 2019).

Het Rijksregister omvat volgende registers (Schockaert, Morissens, Cincinnato, & Nicaise, 2012, pp. 7-8): de bevolkingsregisters van alle gemeenten; het wachtregister (personen in asielpcedure); het vreemdelingenregister van een gemeente (personen die gerechtigd zijn tot een verblijf van meer dan 3 maanden) en de diplomatieke en consulaire registers. Voor de gezinsenquête selecteerde men enkel personen uit de bevolkingsregisters die *of* woonachtig zijn in het Vlaams Gewest *of* woonachtig zijn in het Brussels Hoofdstedelijk Gewest en hier ook een in het Nederlands uitgereikte identiteitskaart hebben verkregen. Daarnaast verkreeg men de toestemming om informatie te onttrekken uit het vreemdelingenregister over de taal van de uitgifte van de identiteitskaart om Nederlandstalige inwoners in Brussel te kunnen weerhouden. Het wachtregister werd niet opgenomen in de steekproeftrekking van de gezinsenquête. Dit is een lijst met personen die officieel een asielaanvraag indienden en wachten op goedkeuring, maar ook personen met een afgewezen asielaanvraag en een bevel om het grondgebied te verlaten (Audenaert & Stuyck, 2018a). Toch zijn er bepaalde groepen in dit wachtregister die rechten kunnen opnemen op (vormen van) gezinsbeleid, maar uitgesloten worden om deel te nemen aan de bevraging. Zo zijn er asielzoekers die pas na latere erkenning als vluchteling retroactief de gezinsbijslag toegekend krijgen tot op het moment van de asielaanvraag⁷. Ook EU- burgers die een verklaring tot inschrijving doen of gezinshereniging aanvragen met een persoon in België hebben het recht om retroactief de gezinsbijslag te ontvangen eens hun inschrijving goedgekeurd is. Tot slot hebben niet-begeleide minderjarige vreemdelingen⁸ meteen recht op gezinsbijslag, zelfs als ze in een opvangcentrum verblijven.

Het Rijksregister, en meer specifiek het bevolkingsregister, is één van de meest complete databronnen om een steekproef te trekken. Desalniettemin gaan er enkele **beperkingen** gepaard met het gebruik van dit steekproefkader. Ten eerste spraken we al eerder van het Rijksregister als de bron die enkel een

⁷ Op het moment van schrijven zou deze retroactieve toekenning van de kinderbijslag afgeschaft worden; op het moment van de steekproeftrekking was dit beleid nog in voege.

⁸ Specifiek de minderjarigen die aan de verblijfsvoorwaarde voldoen met een attest van Immatriculatie. Meer informatie hierover: <http://www.agii.be>.

beeld verschaft van de *wettelijke* situatie, maar niet van de *werkelijke* woonsituatie van talrijke gezinnen. Het Rijksregister bouwt namelijk aan de hand van het domicilie-adres haar huishoudens op uit individuele gegevens. Denk bijvoorbeeld aan kinderen die afwisselend deeltijds wonen bij de vader of moeder met slechts één officieel adres (Studiedienst van de Vlaamse Regering, 2016), studenten die op kot zitten maar nog steeds thuis gedomicilieerd staan, enzovoort. Een voorbeeld van dit knelpunt is dat gescheiden ouders, die geen (stief)kind op haar domicilie gevestigd hebben, niet in de steekproef worden opgenomen, ook al spreekt dit de werkelijke leefsituatie tegen en verblijft het kind occasioneel bij deze ouder. Na scheiding of relatiebreuk is het vaak het adres van de moeder waar het kind op gedomicilieerd staat, waardoor gescheiden vaders ondervertegenwoordigd zijn in de bevolkingsregisters en dus ook in de steekproef van de gezinsenquête (Audenaert & Stuyck, 2018a). Een ander voorbeeld zijn stiefouders die een biologisch kind hebben met een domicilie op een ander adres. Klaarblijkelijk kunnen we niet vertrekken vanuit de veronderstelling dat alle kinderen van elke ouder op eenzelfde adres verblijven (Haskey, 2010). Kortom, het Rijksregister schetst een beeld van de wettelijke woonplaats van een persoon, wat (on)rechtsreeks een uitsluiting van verscheidene bevolkingsgroepen teweegbrengt. Omdat het Rijksregister aan de hand van domicilie huishoudens opbouwt uit individuele gegevens, is vooralsnog geen pasklare oplossing voorhanden (Audenaert & Stuyck, 2018a). Voor een enquête omtrent het gezinsleven en –beleid is het belangrijk om dit in het achterhoofd te houden.

Een tweede beperking, hierbij aansluitend, is dat men geen continue controle toepast op de officieel geregistreerde gezinssamenstelling. De feitelijke gezinssamenstelling kan steeds onderhevig zijn aan wijzigingen, en kan dus afwijken van de wettelijk genoteerde situatie (Audenaert & Stuyck, 2018a). Gezinstransities kunnen, met andere woorden, niet officieel geregistreerd worden maar wel opgevangen worden via surveygegevens. Het gaat dan over het feitelijk uit elkaar gaan van koppels zonder dat ze officieel de burgerlijke staat wijzigen, het starten en beëindigen van ongehuwde samenwoonrelaties, of de verwantschap tussen de gezinsleden (Pasteels, Lodewijckx, & Mortelmans, 2013). Het constant opvolgen van de correctheid in gegevens kan het mogelijk maken om de administratieve data up-to-date te houden en kan de potentiële kloof tussen de werkelijke en feitelijke woonsituatie dichten. De gezinsenquête schenkt aandacht aan de perceptie van de respondent en haar of zijn feitelijke woonsituatie, en levert op die manier een meerwaarde in het sociaalwetenschappelijk onderzoek wanneer we de leefsituatie van gezinnen willen analyseren. Let wel dat het tijdstip van de steekproeftrekking, hoewel zo kort mogelijk gehouden, verschilt met het tijdstip dat men de vragenboekjes verzendt, ontvangt en verwerkt. Bepaalde transitie zoals echtscheiding, relatiebreuk of verhuizen kunnen in die tijd hebben plaatsgevonden en hebben vaak een nadrukkelijke impact op het huishouden. De gezinsenquête tracht dit deels te ondervangen door de respondent te bevragen naar de huidige woonsituatie, via de vraag ‘U woont samen sinds dd/mm/jjjj’ of ‘U woont niet meer samen sinds dd/mm/jjjj’ met aan deze laatste vraag gekoppeld de reden waarom dit niet meer het geval is (bijvoorbeeld financiële of gezondheidsredenen, gezinsomstandigheden, de partners willen apart wonen, enzovoort). Een enquête die administratieve gegevens gebruikt om enquêtegegevens te verrijken, is de European Union Statistics on Income and Living Conditions (EU-SILC), een Europese survey die vandaag de meest gebruikte bron is inzake inkomens, armoede, sociale uitsluiting en levensomstandigheden in Europa. Om de EU-SILC flexibel te kunnen gebruiken, maakt Eurostat (die de survey coördineert) het mogelijk voor landen om enquête- en administratieve gegevens te combineren. De EU-SILC vraagt echter op regelmatige basis een update van het Rijksregister om uitval als gevolg van tijdsverloop te beperken (Jäntti, Törmälehto, & Marlier, 2013).

Een derde beperking van het trekken van een steekproef uit het Rijksregister is dat men bepaalde groepen van huishoudens en personen onvermijdelijk uitsluit (Schockaert et al., 2012). Dit gaat over personen die het OCMW als referentieadres moeten opgeven, zoals personen die illegaal in België terecht kwamen en hier onwettig verblijven, of dak- en thuislozen. Ook personen met een toeristenvisum dat niet meer geldig is maar beslist te blijven, en personen die in het buitenland wonen en/of werken en deze adreswijziging niet hebben aangegeven, zijn personen waarvan de werkelijke woonplaats zich elders bevindt dan aangegeven waardoor je hen moeilijk of niet kan bereiken. Denk ook aan dak- en thuislozen, campingbewoners, rondtrekkende bevolkingsgroepen, et cetera. Het zijn deze bevolkingsgroepen die vaak niet meegenomen worden in steekproeven omwille van hun onbereikbaarheid op basis van wettelijke gegevens. Het uitsluiten van bepaalde groepen op basis van de keuze in steekproefkader is een dekkingfout (coverage error). Nog specifiekere spreekt men over frame coverage error wanneer er een mismatch is tussen het steekproefkader en de doelpopulatie (De Leeuw et al., 2008; Schockaert et al., 2012).

Ten vierde selecteerde men omwille van pragmatische redenen de private huishoudens en sluit men de collectieve huishoudens uit. Dit zijn personen wonende in rust- en verzorgingstehuizen, een militaire instelling, een ziekenhuis, een zorginstelling, een religieuze gemeenschap of een gevangenis. Volgens Statbel telde het Vlaamse Gewest in 2018 80.332 personen die in collectieve huishoudens wonen (Statbel, 2019b).

Afsluitend benadrukken we in deze sectie opnieuw de noodzaak tot het blijvend reflecteren met betrekking tot de keuzes die men maakt. Is dit kader het beste instrument om de doelpopulatie in zijn totaliteit in kaart te brengen? Is dit het enige kader dat men hiervoor kan gebruiken? Het naast elkaar leggen van potentiële steekproefkaders kan hierbij een hulpmiddel zijn om de voor- en nadelen van het werken met het Rijksregister te bekijken. De onderzoeker moet namelijk kunnen beargumenteren waarom hij of zij beslist om, bijvoorbeeld, enkel de bevolkingsregisters op te nemen. Een voorbeeldvraag die men in de discussie in functie van vervolgenresearch kan voorleggen, is in welke mate het Rijksregister een adequaat kader is om kwetsbare gezinnen te bereiken (zie verder).

2 De steekproeftrekking

Voor de gezinsenquête werd gebruik gemaakt van het Rijksregister als het steekproefkader om van daaruit een basis- en extra steekproef te trekken. Het verzamelen van de steekproefgegevens gebeurde via een systematische en gestratificeerde steekproeftrekking. Ten eerste verliep het systematisch, door via het Rijksregister een extractie door te voeren van alle personen die (1) deel uitmaken van alle registers, behalve het wachtregister; (2) minstens 18 jaar zijn op 1 januari 2016; (3) gezinshoofd zijn (enkel private huishoudens); (4) woonachtig zijn in een Vlaamse gemeente of in het Brussels Hoofdstedelijk Gewest en (5) in geval deze in het Brussels Gewest verblijft, enkel te selecteren als zijn of haar identiteitskaart in het Nederlands werd uitgereikt. In een volgende stap reduceerde men deze selectie verder, door te kijken naar de verwantschapscodes gekoppeld aan de geselecteerde referentiepersonen. Deze persoon moet minstens één kind hebben, jonger dan 25 jaar met als verwantschapscode 3 (dochter/zoon), 13 (stiefdochter/zoon), 5 (kleinkind) of 12 (niet-verwant). Vanuit dit steekproefkader trok men vervolgens twee aselechte steekproeven (Audenaert & Stuyck, 2018b). De eerste noemen we in dit rapport de ‘basissteekproef’, de tweede steekproef kwam tot stand door een samenwerking met het Agentschap Binnenlands Bestuur (ABB) en noemen we de ‘extra steekproef’. Dit is de uitbreidende steekproef bij gezinnen met een herkomst van buiten de Europese Unie. Na het selecteren van de gezinnen, werden de referentiepersonen gesorteerd naar dalende leeftijd. In dit design start de onderzoeker vanuit een willekeurig gekozen item in het beschikbare steekproefkader, om van daaruit een *n*-de subject te selecteren. Om het toeval hierbij een rol te laten spelen, kiest men het startpunt lukraak (Audenaert & Stuyck, 2018a). Deze selectie kan enkel *random* gebeuren indien men toegang heeft tot gemakkelijk te identificeren populatiegegevens. Voor de gezinsenquête 2016 werd een algoritme uitgewerkt dat het gezinstype en de positie van gezinsleden ten aanzien van de referentiepersoon in rekening brengt. Per geselecteerd gezin liet men een iteratie lopen met het algoritme dat de steekproefpersoon selecteert (Audenaert & Stuyck, 2018a).

Ten tweede spreken we van een gestratificeerde steekproef omdat de basissteekproef van de gezinsenquête werd uitgebreid met een extra steekproef van personen met niet-EU-herkomst. In de eerste steekproef die men trok, of de basissteekproef, kregen 9.000 referentiepersonen en hun partners een vragenboekje. In de uitbreiding, of extra steekproef, werden 2.000 referentiepersonen (maar niet de partners) aangeschreven. Deze basis en extra steekproef worden uitgebreid besproken in Hoofdstuk 4. Hier staan we stil bij de steekproeftrekking. We hebben hier namelijk te maken met een disproportioneel gestratificeerde random sample: de extra steekproef zorgt ervoor dat men een groter aandeel niet-EU- steekproefpersonen selecteert zodat deze groep meer kans heeft om geselecteerd te worden en waardoor de steekproefuitkomsten moeten worden herwogen (Audenaert & Stuyck, 2018a). Deze uitbreidende steekproef kwam tot stand vanwege een bezorgdheid van het Agentschap Binnenlands Bestuur dat enquêtes er vaak amper of niet in slagen om voldoende personen met een herkomst van buiten de Europese Unie te bereiken. Met andere woorden, dit criterium werd toegepast omwille van een methodologische keuze, eerder dan dat men deze groep inhoudelijk verder wil verdiepen.

Disproportionele stratificatie vinden we eveneens terug in andere enquêtes. De MEQIN⁹- survey (of ‘Measuring Equivalent Incomes’) maakt, bijvoorbeeld, gebruik van zulk een gestratificeerde

⁹ De UCL vormt, samen met de UA, UB en KUL, het consortium MEQIN. Dit is een afkorting voor ‘Measuring Equivalent Incomes’, waarbij men via face-to-face interviews bij 4.000 steekproefpersonen onderzoek doet naar de levensstandaard in België, waarbij men het begrip ‘equivalent inkomen’ concreet probeerde in te vullen.

steekproeftrekking om de doelgezinnen te bereiken. Eveneens vertrekkend vanuit het Rijksregister, selecteerde men veertien demografische groepen. Dit deed men via drie criteria: (1) de leeftijd, (2) de gezinsgrootte en (3) of men een eenoudergezin is of niet. Aan de Kruispuntbank Sociale Zekerheid (KSZ) vroeg men de trekking uit het steekproefkader te stratificeren volgens volgende schema (Tabel 2) (Maniquet, Schokkaert, Cantillon, & De Rock, 2018, p. 3).

Tabel 2 Definiëring groepen zoals gebruikt in de stratificatiemethode van de Measuring Equivalent Incomes (of MEqIN)- survey (Maniquet et al., 2018, p. 3)

Gezinsgrootte	Jong (18-60 jaar)		Oud (60+)
	Alleenstaande ouder	Geen alleenstaande ouder	
1	-	Groep 1	Groep 2
2	Groep 3	Groep 4	Groep 5
3	Groep 6	Groep 7	Groep 8
4	Groep 9	Groep 10	Groep 11
Meer dan 4	Groep 12	Groep 13	Groep 14

Door zich te baseren op deze drie criteria konden de onderzoekers een voldoende hoge respons garanderen voor alle demografische groepen. Deze survey illustreert hoe men, met het onderzoeksopzet in het achterhoofd, de doelpopulatie ontleedt om een voldoende hoge respons te krijgen van elke groep waarin men geïnteresseerd is.

Een andere enquête die deze disproportionele stratificatie hanteert, is de survey Gemeente- en Stadsmonitor 2017. Voor deze enquête ordende men het bevolkingsregister volgens relevante kenmerken (namelijk geslacht, leeftijd en nationaliteit) waaruit systematisch eenheden werden geselecteerd. Vervolgens sorteerde men alle Belgische mannen van jong naar oud op de lijst, dan alle Belgische vrouwen naar leeftijd. Verder zetten de onderzoekers ‘de sprong’ (gelijk aan de steekproef-fractie) vast op 100, bijvoorbeeld door in een stad van 50.000 inwoners 500 personen te selecteren. Men kiest dan een toevalsgetal (t) tussen 1 en 100, waardoor de steekproef volgende elementen bevat: t, 100+t, 200+t, enzovoort. Opnieuw gebeurt dit, ook in deze survey, in functie van hun onderzoeks-opzet: de Gemeente- en Stadsmonitor streeft ernaar om in elke centrum een representatief zicht te krijgen op de door hen geselecteerde relevante kenmerken.

Is het noodzakelijk voor de gezinsenquête om specifieke relevante kenmerken naar voren te schuiven? Is het voldoende om één willekeurige basissteekproef te trekken om uitspraken te kunnen doen over de hele doelpopulatie? Is het enkel een kwestie van het voldoende aandacht schenken aan de herkomst opdat men een representatief staal van de gezinnen in Vlaanderen en (Nederlandstalig) Brussel bekomt? Doorheen dit rapport beargumenteren we dat er, in functie van vervolgenquêtes, een zekere reflectie genoodzaakt is voor deze keuze in steekproeftrekking. We doen hiervoor eerst inspiratie op bij andere studies met een gelijkaardig opzet.

In 2004 organiseerde het Centrum voor Sociaal Beleid Herman Deleeck (CSB) het project “Zorgen voor kinderen in Vlaanderen”. Het betrof onder meer het opzetten van een enquête naar het hoe en waarom van zorg en opvang bij ongeveer 3.000 gezinnen met kinderen jonger dan 16 jaar. De vragen die ze

stelden, komen gedeeltelijk overeen met deze van de gezinsenquête. Men vroeg de gezinnen naar de arbeidsmarktsituatie, arbeidscontract en flexibiliteit van de job, de (in)formele opvang, de zorg-werk balans, en dergelijke. Naast een aselechte steekproef van gezinnen in Vlaanderen, trok men ook twee extra steekproeven: een steekproef bij gezinnen met een kind met een handicap en een steekproef bij kansarme gezinnen. Op die manier wilde de studie nadrukkelijk de focus leggen op kwetsbare gezinnen. Voor de eerste extra steekproef selecteerde men aselekt een groep uit de databank van het Vlaams Agentschap voor Personen met een Handicap (VAPH). De laatste extra steekproef bij kansarme gezinnen werd getrokken uit de databank van Kind en Gezin. Kind en Gezin verzamelt reeds jarenlang cijfergegevens over de leefsituatie van gezinnen met kinderen. Wat Kind en Gezin daarnaast ook doet is de kansarmoede in kaart brengen, via een geconstrueerde kansarmoede-index, die aangeeft hoeveel procent van kinderen (tussen 0 en 3 jaar) van bij de geboorte opgroeit in een situatie van kansarmoede. Concreet bekijkt Kind en Gezin voor zes verschillende levensdomeinen: (1) het inkomen van het gezin, (2) de arbeidssituatie van de ouders, (3) de opleiding van ouders, (4) de kwaliteit van de huisvesting, (5) het stimulatie-niveau van de kinderen en (6) de gezondheid. Wanneer een gezin onvoldoende scoort op drie of meer van de zes criteria, spreekt Kind en Gezin over een kansarm gezin (Kind en Gezin, 2018). Voor de CSB- studie kwamen de steekproefpersonen voor de extra steekproef bij kwetsbare gezinnen slechts in aanmerking wanneer het gezin slecht scoorde op minstens drie van de zes eerder genoemde criteria. De onderzoeksresultaten toonden aan dat kansarme gezinnen en eenoudergezinnen de meest kwetsbare groepen bleken die meer beleidsaandacht verdienen. Ook scoren kinderen van laag-opgeleide, alleenstaande, werkloze moeders beduidend zwakker op vlak van welbevinden in vergelijking met kinderen die opgroeien in een tweeoudergezin met een werkende, hoger opgeleide moeder (Ghysels & Debacker, 2007a). *“Er is met andere woorden duidelijk nood aan een zorgbeleid dat oog heeft voor de moeilijke situatie waarin deze kwetsbare groepen zich bevinden en dat werkt aan oplossingen op maat waarin tewerkstelling en zorg voor kinderen niet los van elkaar, maar binnen de hele gezinscontext worden bekeken (Ghysels & Debacker, 2007b, p. 145)”. Deze studie maakt duidelijk dat, indien men de gehele gezinscontext wil kunnen bestuderen, het noodzakelijk is om aandacht te schenken aan specifiek kwetsbare groepen. Dit kan door het trekken van een ‘brede steekproef’ van kansarme gezinnen, maar ook door een meer ‘enge’ steekproef van kinderen met een handicap.*

Ten tweede verkennen we de uitbreiding die het JOnG!- onderzoek¹⁰ (2016) doorvoerde vanuit het onderzoeksprogramma ‘Jeugd en Gezin’, uitgevoerd door het Steunpunt Welzijn, Volksgezondheid en Gezin (SWVG). Dit JOnG!- onderzoek is een longitudinale studie naar opvoeding, gezondheid, gedrag en ontwikkeling bij 0-, 6-, en 12- jarige kinderen en hun gezin. Deze unieke gegevens laten toe om de invloed van verscheidene factoren op de ontwikkeling van kinderen in Vlaanderen te bestuderen. Specifiek omdat er een gebrek aan longitudinaal onderzoek is naar oorzaken en gevolgen van kansarmoede in Vlaanderen, streefde een extra onderzoek van het Vlaams Armoedesteunpunt (VLAS) ernaar om deze JOnG! –data uit te breiden met een extra steekproef van kansarme gezinnen. De focus van dit VLAS-onderzoek ligt bij armoede in de vroege levensjaren (i.e. 0-3 jaar) en het belang van *Early Childhood Education and Care (ECEC)* waardoor enkel de cohorte van de 0-jarigen werd uitgebreid met een steekproef van kansarme gezinnen. In deze studie erkende men expliciet dat de kansarme gezinnen in de JOnG!- steekproef disproportioneel laag vertegenwoordigd werden (Lemmens & Groenez, 2015a). Uit deze studie bleken voornamelijk kinderen op basis van socio-economisch profiel, gezinskenmerken en sociale context sterk af te wijken van de JOnG!-cohorte en de Vlaamse gemiddelden. De meest kwetsbare kinderen waren diegene waarvan de ouders laagopgeleid zijn, een beperkt inkomen hebben,

¹⁰ Meer informatie: <http://www.steunpuntwvg.be/jong>.

een precare arbeidsmarktsituatie en opgroeien in een groot gezin (drie kinderen of meer) (Juchtmans & Groenez, 2016; Kuppens et al., 2016). Het gaat echter over een kleine steekproef met niet-toevallige uitval waardoor de resultaten met enige voorzichtigheid moeten worden geïnterpreteerd. Hoewel de resultaten niet te veralgemenen zijn naar alle gezinnen in armoede, zo zeggen de onderzoekers, zijn de bevindingen interessant als indicatie voor toekomstig onderzoek (Kuppens et al., 2016). De steekproeftrekkingen van zowel de CSB- als de VLAS-studie tonen aan dat de keuzes gemaakt in de steekproeftrekking steeds in de lijn liggen van het onderzoeksopzet¹¹.

Een alternatieve keuze, naast de focus op, bijvoorbeeld kwetsbare gezinnen, kan ook zijn om aandacht te schenken aan de geografische representativiteit van de gezinsenquête. In geval men er zeker van wil zijn dat alle provincies en/of gemeenten correct vertegenwoordigd zijn, is een geografische stratificatie aangewezen. Een geografische stratificatie, specifiek voor de gezinsenquête, zou kunnen inhouden dat in de zes betrokken *strata* (vijf Vlaamse provincies en het Brussels Hoofdstedelijk Gewest) personen worden geselecteerd omdat men voldoende huishoudens bereikt in alle strata. De MEQIN-survey schonk hier aandacht aan door gebruik te maken van wat ze ‘pseudo-clustering’ noemen. In elke provincie werden bepaalde steden selectief gekozen, terwijl andere kleinere steden of dorpen random werden geselecteerd. Elke entiteit (dus zowel de grote als de kleine steden of dorpen) wordt hier gedefinieerd als ‘pseudo-cluster’ (Maniquet et al., 2018). Door deze stratificatie te integreren in het design kan men garanderen om in elke provincie voldoende personen te bereiken. Wanneer de gezinsenquête bijvoorbeeld Brussel wil blijven betrekken in haar analyse, kan het daarbij de mogelijkheid zijn om geografisch te stratificeren om voldoende Brusselse gezinnen in de steekproef te garanderen.

Zoals we verder zullen zien in Hoofdstuk 4.3 slaagde de basis enquête erin om een representatieve provinciale verdeling in de gerealiseerde steekproef te bekomen. Met andere woorden, de geselecteerde basissteekproef selecteerde een zeker percentage van inwoners in een provincie, en kon een voldoende hoog aandeel personen van deze provincie effectief bevragen (zie Tabel 3). Dankzij de verkregen administratieve gegevens vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming construeerden we het overzicht zoals weergeven in Tabel 3.

¹¹ Een zekere nuancering is hierbij belangrijk. De longitudinale VLAS-studie beschikte slechts over een beperkte steekproefgrootte en een niet-toevallige uitval over tijd (Kuppens et al., 2016).

Tabel 3 Provinciale verdeling van de gerealiseerde steekproef in de gezinsenquête naar herkomst (Bron: Eigen berekeningen via Datawarehouse Arbeidsmarkt en Sociale Bescherming)

	BE	EU	Niet-EU	<i>n</i>
Provincie				
Antwerpen	63,4	11,8	24,8	750
Brussel	20,0	24,4	55,6	29
Limburg	62,5	15,1	22,3	322
Oost-Vlaanderen	74,6	4,0	21,4	590
Vlaams-Brabant	64,3	10,2	25,5	437
West-Vlaanderen	83,2	5,8	11,9	417
				<i>Totaal: 2.545</i>

De weergegeven percentages zijn rijpercentages.

Uit deze tabel blijkt, zoals we eerder zagen in Hoofdstuk 1, dat Brussel slechts een klein deel uitmaakt van de gerealiseerde steekproef. Wat we verder ook zullen zien in Hoofdstuk 4 is dat het voornamelijk de extra steekproef was die ervoor zorgt dat we een hoger aandeel van Brusselse respondenten konden bereiken. Wat we in dit hoofdstuk voornamelijk willen bespreken, is het verschil in herkomstgroepen per provincie die men heeft kunnen bevragen. Van alle bevroegde personen, wonen, bijvoorbeeld, 16,4% in West-Vlaanderen. Maar deze West-Vlamingen bestaan voornamelijk uit Belgen (83,2%), in mindere mate uit personen met een herkomst binnen de Europese Unie (5,8%) en erbuiten (11,9%). Aansluitend op Hoofdstuk 1 zien we hier dat voornamelijk Brussel erg afwijkt van de vijf Vlaamse provincies op vlak van herkomst. Hoewel het bereik in Brussel klein was (slechts 29 personen bereikt), bestaat deze groep voornamelijk uit personen met een vreemde herkomst: 55,6% van deze respondenten heeft een herkomst buiten de Europese Unie, 24,4% heeft een herkomst binnen de Europese Unie (maar niet België). Hieruit vloeit de discussie voort of men, in functie van vervolgenquêtes, Brussel moet blijven betrekken in de gezinsenquête. Het blijvend integreren van deze stad zorgt ervoor dat er voldoende personen met een diverse herkomst worden betrokken, maar anderzijds kan je via deze respondenten ook peilen naar het gezinsleven in Brussel. Wanneer de gezinsenquête de Brusselse context echter meer diepgaand bereiken, is het noodzakelijk om meer personen in dit gewest te bereiken. Net zoals de Gemeente-en stadsmonitorenquête kan men in dat geval doen aan 'multi-stage sampling', een steekproeftrekking die vertrekt vanuit de premisse dat socio-economische status van verschillende huishoudens geclusterd zijn binnen steden, en dat men via de postcode in principe een vernauwing kan maken van de geselecteerde steekproefeenheden (Agentschap Binnenlands Bestuur, 2018a). Maar de gezinsenquête kan ook opteren voor een extra steekproef van gezinnen in deze Brusselse context, om meer of andere soorten vragen te stellen over de organisatie van het gezinsleven in Brussel.

3 Longitudinaal onderzoek

Het gezin is een dynamisch gegeven. Luyten (2017) pleit voor meer beleidsaandacht voor de gezinsdynamiek en gezinstransities. Deze transities zijn moeilijk te capteren via cross-sectionele bevragingen omdat het hier gaat om een momentopname van de leefsituatie van de respondent. Temporele veranderingen bij individuen of gezinnen zijn dan ook bij voorkeur te observeren door middel van herhaaldelijke bevragingen van dezelfde personen via een longitudinaal design. In deze sectie omschrijven we de meerwaarde van het longitudinaal onderzoeken van gezinnen, en specifiek kwetsbare gezinnen.

In de voorbereidingsfase van de gezinsenquête bracht de stuurgroep de mogelijkheid tot het opnemen van een longitudinaal onderzoeksdesign meermaals ter sprake. Uit academische hoek bleek er vraag naar een enquête die via een longitudinaal design gezinnen opvolgt, maar uit pragmatische redenen besliste de stuurgroep uiteindelijk dit niet te doen. Zo bleek dat men, ten eerste, dit design van bij de start moest integreren en verantwoorden bij de aanvraag voor de Commissie ter Bescherming van de Persoonlijke Levenssfeer (CBPL) om een steekproef te verkrijgen uit het Rijksregister. Een koppeling met de KSZ-gegevens op longitudinale basis is namelijk geen vanzelfsprekend gegeven. De individuele gegevens van de steekproefpersonen moeten in dat geval voor een langere periode bewaard blijven, wat de aanvraag complexer maakt. Ten tweede is het niet evident om gezinnen op te volgen met minstens één kind jonger dan 25 jaar. In geval deze in een tweede wave al ouder zijn dan deze leeftijdsgrens, vallen ze al uit de omschreven doelpopulatie die men wil bestuderen. Bovendien kende het onderzoeksteam de knelpunten van een longitudinaal design, namelijk uitval tussen de verschillende golven (of attritie) als gevolg van ontbinding van een huishouden, verhuizen, overlijden, enquête-moeheid, waardoor men extra moet inzetten op het zo groot mogelijk houden van de steekproefgrootte. In geval van een longitudinale studie is het in dat opzicht interessant om een bredere steekproef te trekken, zodat 'gezinnen in wording' ook opgenomen kunnen worden. Omwille van deze overwegingen opteerde het onderzoeksteam van de gezinsenquête voor een cross-sectioneel design (Audenaert & Stuyck, 2018a).

Het longitudinaal opvolgen van gezinnen, en met name kwetsbare gezinnen, zou echter voordelen opleveren om goed de vinger aan de pols te houden omtrent het gezinsleven in Vlaanderen. Het cross-sectioneel bevragen van gezinnen is interessant om een momentopname te schetsen van de gezinnen in Vlaanderen. Een dwarsdoorsnede van de doelpopulatie leert ons vaak welke personen kwetsbaar zijn op één moment, maar geven niet noodzakelijk aan welke personen (een groot) risico lopen om kwetsbaar te zijn in de (nabije) toekomst.

In Vlaanderen zijn er amper gegevens beschikbaar over de evolutie en trends van gezinnen op lange termijn. Een longitudinaal design vinden we terug bij de Panel Studie van Belgische Huishoudens (PSBH), die werd opgenomen in het geharmoniseerde European Community Household Panel (ECHP), en bij de European Union Statistics on Income and Living Conditions (EU-SILC). De EU-SILC-survey beschikt over zowel een cross-sectioneel als een longitudinale component waarbij huishoudens gedurende een periode van vier jaar gevolgd worden. Door dezelfde huishoudens over een langere periode te volgen is het mogelijk om na te gaan welke veranderingen die huishoudens ervaren over de jaren heen op allerlei dimensies bevraagd in de studie. De keuze voor een longitudinaal onderzoeksdesign moet gemaakt worden in functie van het onderzoeksopzet en de praktische haalbaarheid: Welke vragen wil men beantwoorden? En hoe kunnen we dit het beste bestuderen en analyseren? In geval men steeds een dwarsdoorsnede van de verschillende huishoudens wil bekijken, is het voldoende om een cross-

sectioneel design te hanteren. In geval men graag dieper wil ingaan op de kwetsbaarheid van specifieke gezinnen over de jaren heen, is het wel een meerwaarde om een longitudinaal design te overwegen. De literatuur bespreekt namelijk verschillende voordelen van het hanteren van een longitudinaal design in het bestuderen van kwetsbare bevolkingsgroepen. Volgens Bircan, Nicaise, Gábos, and Tóth (2017) laat het roterende panel in de EU-SILC toe om (1) inkomensmobiliteit te analyseren in een levensloop-perspectief; (2) een vergelijking maken van korte en lange termijn impact van beleidsmaatregelen of sociale investeringen op bepaalde levensloopgebeurtenissen; en (3) een zicht krijgen op armoede voor een korte periode of persistente armoede.

Bovendien moet men hiervoor het warme water niet uitvinden, maar er kan ook worden geleerd uit de lessen van andere longitudinale bevestigingen zoals bij de EU-SILC. Een studie van Bircan et al. (2017) vat enkele van deze lessen samen, namelijk het vergroten van de steekproef, het gebruik van een geharmoniseerde *identifier* zodat men de cross-sectionele en de longitudinale dataset beter kan combineren, en het reduceren van attritie. Bovendien raden de auteurs aan om het tijds kader breder te installeren, dat huishoudens niet vier maar vijf jaar worden gevolgd om potentiële transitie in of uit armoede nog beter in beeld te brengen. De onderzoekers suggereren ook af te stappen van een roterend panel. De longitudinale steekproef van de EU-SILC bestaat namelijk uit vier roterende groepen van huishoudens die deelnemen aan de bevestiging voor vier opeenvolgende jaren, waarvan elk jaar één van deze vier groepen wordt ingewisseld met een nieuwe roterende groep. Volgens de studie kan men beter één grote steekproef trekken van personen om deze voor de volledige periode jaarlijks te bevestiging.

Een ander, specifiek Vlaams, longitudinaal onderzoek is, zoals eerder besproken, de uitbreiding die plaatsvond bij het JOnG!-onderzoek¹². De JOnG!-studie is een longitudinaal onderzoek naar opvoeding, gezondheid, gedrag en ontwikkeling bij 0-, 6-, en 12- jarige kinderen en hun gezin. Deze unieke gegevens laten toe om de invloed van verscheidene factoren op de ontwikkeling van kinderen in Vlaanderen te bestuderen. Specifiek omdat er een gebrek aan longitudinaal onderzoek is naar oorzaken en gevolgen van kansarmoede in Vlaanderen, vond in het kader van het Vlaams Armoedesteunpunt (VLAS) een onderzoek plaats om deze JOnG! –data uit te breiden met een sample van kansarme gezinnen. De focus van dit VLAS- onderzoek ligt bij armoede in de vroege levensjaren (i.e. 0- 3 jaar) en het belang van voorschoolse voorzieningen zoals kinderopvang waardoor enkel de cohorte van de 0- jarigen werd uitgebreid met een steekproef van kansarme gezinnen. “*Het doel van dit VLAS – onderzoek is bijgevolg om de 0- jarige JOnG! –cohorte aan te vullen met een bijkomende sample zodat longitudinaal onderzoek naar de effecten van armoede op de vroege ontwikkeling ook in Vlaanderen mogelijk wordt* (Lemmens & Groenez, 2015a, p. 4)”. Uit dit onderzoek bleek echter dat het moeilijk was om de doelgroep van kansarme kinderen goed af te bakenen, met als gevolg dat het moeilijk was om een voldoende grote bijkomende sample te rekruteren (Lemmens & Groenez, 2015b).

¹² JOnG! is een longitudinaal onderzoek vanuit het onderzoeksprogramma ‘Jeugd en Gezin’, uitgevoerd door het Steunpunt Welzijn, Volksgezondheid en Gezin (SWVG). Meer informatie: <http://www.steunpuntwvg.be/jong>.

4 Scenario's voor de toekomst

Wat de voorgaande studies duidelijk maken is dat het enerzijds inhoudelijk interessant is om zich te verdiepen in een longitudinaal design om het gezinsleven vast te leggen en dat er anderzijds ook blijvend vraag is naar data over evoluties en trends binnenin dit gezinsleven. Of dit praktisch en methodologisch haalbaar is, bespreken we door enkele scenario's te presenteren voor potentiële designs in functie van vervolgenquêtes. Een overzicht van deze scenario's presenteren we in Tabel 4.

Tabel 4 Overzicht scenario's design vervolgenquêtes gezinsenquête (Geïnspireerd op Frans, Mortelmans, Sels, & Forrier, 2009)

Survey design	
Scenario 1	Opeenvolgende cross-secties gezinsenquête
Scenario 2	Longitudinale survey gezinsenquête
Scenario 3	Administratieve data Datawarehouse
Scenario 4	Administratieve data Datawarehouse + cross-secties gezinsenquête
Scenario 5	Administratieve data Datawarehouse + longitudinale gezinsenquête

Het eerste scenario zou verder gaan op het huidige design van de gezinsenquête, waarbij op cross-sectionele wijze data verzameld wordt met een steeds wijzigende representatieve steekproef. Herhaalde cross-secties vermijden een aantal nadelen van een longitudinaal design, zoals bijvoorbeeld de uitval van respondenten (attritie). Het bestuderen van de doelpopulatie via deze momentopnames maakt het alsook mogelijk om trends op te volgen. Wanneer de gezinsenquête in vervolgenquêtes gebruik maakt van eenzelfde gestandaardiseerde vragenlijst, maakt dit het mogelijk om de bevindingen over tijd te vergelijken. Op die manier kan een algemeen inzicht over het gezinsleven verworven worden (Frans, Mortelmans, Sels, & Forrier, 2009; Ruspini, 2002). Wat ook interessant kan zijn, is dat men steeds bewust andere groepen gaat opmeten, naargelang de interesse of het thema waar men zich in wil verdiepen (Frans et al., 2009). Werken via wisselende thematische modules als aanvulling op een standaard vragenlijst is ook een mogelijkheid om zich te verdiepen in een specifiek thema. Een zekere flexibiliteit in het design van de enquête laat toe om beleidsspecifieke vragen te beantwoorden.

Desalniettemin kent dit cross-sectionele design nadelen, en deze zwaktes worden vaak beschreven als de sterktes van het longitudinale design. Het is enkel mogelijk om via een panelstudie individuele of huishoudelijke veranderingen of transities te onderzoeken. De veranderingen die in een cross-sectioneel design worden geregistreerd, zijn enkel geldig op geaggregeerd niveau omdat men geen uitspraken kan doen over individuele trajecten (Cohen & Manion, 1980; Frans et al., 2009). Bovendien kan er met het huidige design in de breedte gewerkt worden, maar niet in de diepte. En dit is meteen het belangrijkste nadeel in het bestuderen van het gezinsleven: dergelijk onderzoek laat enkel toe om een statisch beeld te verschaffen van de doelpopulatie, wat het bestuderen van dynamische transities en de causale factoren die deze sturen onmogelijk maakt. Causale relaties tussen specifieke factoren kunnen de facto pas na een bepaalde tijdsperiode zichtbaar worden. Bestuderen van veranderingen en verklaringen op cross-sectionele wijze werkt vertekening in de hand, omdat de onderzoekers in dat geval moeten vertrouwen op het geheugen of de herinneringen van de respondent (Frans et al., 2009).

In scenario twee gaan we uit van een paneldesign waarbij eenzelfde groep personen of gezinnen doorheen de tijd gevolgd wordt. De dataverzameling kan jaarlijks of om de x aantal jaren gebeuren. In de EU-SILC hanteert men een roterende panel waarbij de gekozen steekproefpersonen om de vier jaar worden bevraagd. De keuze van dit tijdsinterval heeft veeleer te maken met praktische haalbaarheid van het uitvoeren van de enquête. Het jaarlijks organiseren van een bevraging vraagt voldoende tijd, budget, personeel, en dergelijke (Frans et al., 2009).

Een belangrijke en vaak vergeten meerwaarde van een longitudinaal design is het betrekken van respondenten als partners om de dataverzameling tot een goed einde te brengen. Hiermee bedoelen we dat onderzoekers de respondenten het gevoel kunnen geven dat hun bijdrage een cruciale rol speelt in het verzamelen van kennis om het gezinsbeleid mee te ondersteunen. Dankzij het engagement van deze deelnemende gezinnen kan de studie een vinger aan de pols houden bij het gezinsleven, zodat de gezinnen ook aangemoedigd worden om blijvend betrouwbare informatie te delen (Couper, 2011).

We vertrekken vanuit het concept van 'kwetsbare gezinnen' waar dit rapport zich op focust. De kwetsbare gezinnen anno 2016 waren niet noodzakelijk dezelfde in 2000, en zullen niet noodzakelijk dezelfde zijn in 2026. Het is een belangrijke beleidskwestie welke gezinnen vanuit een kwetsbare positie niet langer kwetsbaar zijn, en welke beleidsingrepen daarmee samenhangen. De dynamieken en transitie in en uit kwetsbaarheid c.q. armoede kunnen alleen in beeld gebracht worden door middel van een paneldesign. Of een gezin vandaag kwetsbaar is, is niet noodzakelijk de beste indicator om te voorspellen of iemand kwetsbaar is in de toekomst (Chaudhuri, Jalan, & Suryahadi, 2002). Zeker wanneer men beleid wil betrekken in deze vraag, moet men verder gaan dan het louter registreren van de kwetsbare gezinnen van vandaag, om zicht te krijgen op de kwetsbaarheid over de tijd heen.

Een alternatieve mogelijkheid, scenario drie, is om volledig gebruik te maken van administratieve gegevens uit de Kruispuntbank Sociale Zekerheid (KSZ). De KSZ creëerde via het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB) een mogelijkheid om administratieve gegevens vanuit verscheidene sociale zekerheids- en overheidsinstellingen te koppelen. Het Datawarehouse voorziet in (quasi) populatiegegevens omtrent individuele- en gezinskenmerken. Van hieruit kan men gegevens van individuen koppelen aan gegevens van de gezinsleden via het Rijksregister. Concreet gaat het om het opvolgen van een vaste procedure van gegevensaanvragen bij de KSZ, door ofwel telkens een representatieve steekproef te trekken ofwel een representatieve steekproef te trekken die individuen over de tijd heen opvolgt. Men geeft de beschikbare gegevens in dat geval een panelkarakter waardoor individuele- en gezin levenslopen opgevolgd kunnen worden (Frans et al., 2009)¹³.

Het werken met administratieve data heeft in principe een aantal voordelen. Het gaat om een databestand van de populatie met uitgebreide en gedetailleerde gegevens over verschillende relevante kenmerken van personen, zoals nationaliteit, gezinssamenstelling, arbeidsregime, loon, en dergelijke. Het is ook mogelijk om gezinskenmerken te koppelen aan individuele kenmerken (zoals de arbeidsmarktgegevens) zodat de gezinscontext geïntegreerd wordt in de analyse (De Lathouwer, Bogaerts, & Dillen, 2006; Frans et al., 2009). Bovendien moeten er geen gevoelige vragen aan respondenten worden gesteld, bijvoorbeeld over hun inkomsten. Administratieve gegevens, bijvoorbeeld over deze inkomsten, zijn vaak meer accuraat dan enquêtegegevens (Frans et al., 2009).

Bovendien kunnen uit administratieve gegevens grotere steekproeven worden getrokken. Daarnaast valt het te verwachten dat de potentiële attritie of uitval van respondenten in dit geval minder selectief

¹³ Meer informatie over deze specifieke databron in 4.1

en beperkter zal zijn. Het verkrijgen van een lage respons door het bevragen van kleine subgroepen in de doelpopulatie, valt weg (Ruspini, 2002). Een longitudinale studie met administratieve gegevens maken het dus mogelijk om specifieke gebeurtenissen of transities bij kwetsbare gezinnen, die men vaak bij een gewone enquête niet kan bereiken, te bestuderen. Tot slot is dit design een goedkoper alternatief omdat (1) de dataverzameling reeds gebeurde, (2) alle analyses meteen uitgevoerd kunnen worden en (3) de data in gestandaardiseerde formats geleverd worden (Frans et al., 2009; Ruspini, 2002).

Maar het enkel en alleen werken met administratieve data kent ook belangrijke nadelen. De variabelen werden vanuit bepaalde administratieve doelstellingen gedefinieerd en geoperationaliseerd waardoor deze inhoudelijk beperkt kunnen zijn en kunnen afwijken van de vooropgestelde onderzoeksdoelstellingen. Het Datawarehouse beschikt over bruto belastbare inkomens (dus inkomens voor belastingen, maar na de aftrek van sociale bijdrage) die worden opgebouwd uit de inkomsten gekend door de sociale zekerheid. Andere inkomens, zoals inkomsten uit buitenlands arbeid of zwartwerk, worden hier niet bij betrokken. Verschillende dimensies van kwetsbaarheid, zoals het besteedbare inkomen of materiële deprivatie kunnen hier niet uit worden afgeleid. Een zeer belangrijke beperking is tevens het ontbreken van gegevens over de percepties, motivaties en intenties van de respondent. De subjectieve inschatting of iemand rond kan komen met het beschikbare inkomen kan niet worden meegenomen (Frans et al., 2009). We stellen ook vast dat een administratieve databron beperkt kan zijn voor specifieke variabelen, zoals de inkomens. Omwille van privacy redenen worden niet de precieze inkomens meegedeeld, maar werkt men met inkomensklassen. Daarnaast houdt het inkomensbegrip van de KSZ eenzijdig rekening met beroepsinkomsten en uitkeringen, maar niet met huurinkomsten, onroerend vermogen, dividenden, en dergelijke. Tot slot merkte het Departement WVG reeds snel op dat het geen gemakkelijke oefening is om toegang te verkrijgen tot deze gegevens. Als gevolg van privacy- en vertrouwensredenen moet elke data-aanvraag voorbij het informatieveiligheidscomité, waardoor het effectief verkrijgen van de gegevens vertraging kan oplopen. Daarnaast zit er een zekere vertraging op de beschikbaarheid van de gegevens. Dit maakt het moeilijk voor beleidsmakers om de huidige stand van zaken steeds nauwgezet op te volgen (Bakker, Arts, & Toor, 2005; Frans et al., 2009). Op het moment van schrijven konden we beroep doen op administratieve gegevens van 2015. Voor het beschrijven en evalueren van de gezinssituaties in Vlaanderen lijkt een puur administratieve route dus minder geschikt.

In scenario vier worden de administratieve data van het Datawarehouse gekoppeld aan cross-sectionele enquêtegegevens van de gezinsenquête. Door scenario's 1 en 3 te combineren kunnen de beperkingen van beiden gereduceerd worden. Er zijn twee mogelijkheden. Een eerste mogelijkheid is om op gezette tijdstippen een nieuwe golf van de gezinsenquête te organiseren (zoals in scenario 1), maar bepaalde gegevens laten invullen op basis van de administratieve data. Een tweede mogelijkheid is om een steekproef te trekken op basis van administratieve data (zoals in scenario 3) en die mensen via een postenquête (of via andere modi, zie Hoofdstuk 3) bijkomende vragen te stellen. Dit wordt in de literatuur ook wel *data linkage* genoemd, een methode waarin men enquête- met administratieve gegevens koppelt om het potentieel van de beschikbare data te vergroten. In geval men meer diepgaand het Vlaamse gezinsleven wil leren kennen, kan men afstappen van de vragen die in feite af te leiden zijn uit de administratieve data. Indien respondenten deze vragen niet moeten beantwoorden, kan er meer aandacht besteed worden aan andere vragen die niet zomaar in te vullen zijn op basis van administratieve gegevens.

Het linken van enquête- en administratieve gegevens kent een aantal voordelen: (1) Terwijl de enquêtegegevens vaak slechts een momentopname zijn van de respondent, worden administratieve data in

principe consequent en nauwkeurig geregistreerd, wat resulteert in een hogere betrouwbaarheid van de gegevens. In plaats van data te hebben voor één specifiek jaar, kan men vaak beroep doen op verscheidene jaren en een chronologisch beeld van het gezin of het gezinslid schetsen: veranderingen in gezinsstructuren, transities in het gezin worden hier op objectieve wijze vastgelegd; (2) Het *matchen* van de data kan ervoor zorgen dat men de vragenlijst kan inkorten, wat ook budgettair gunstige gevolgen kan hebben. Het antwoord op sommige items in de vragenlijst kan namelijk rechtstreeks gehaald worden uit administratieve data, zodat de vragenlijst voor de respondenten niet te langdradig wordt. Maar bovenal kan men hierdoor gevoelige vragen, zoals het inkomen, weglaten. Deze gevoelige vragen zorgen er namelijk voor dat er item non respons ontstaat, of doen respondenten algeheel afhaken (Künn, 2015; O'Hara, Shattuck, & Goerge, 2017). Wanneer men bijvoorbeeld expliciet wil nagaan in welke mate de huishoudens gebruik maken van voor hen beschikbare voorzieningen, kan men deze praktische implicaties van het beleid nagaan door gebruik te maken van geregistreerde data bij de betrokken voorzieningen (O'Hara et al., 2017); (3) Daarnaast kan men complementair te werk gaan: variabelen die ontbreken in administratieve data kunnen bevraagd worden via een enquête, bijvoorbeeld opinies, subjectieve inschatting van de eigen leefsituatie, tevredenheid, werkomstandigheden, enzovoort (De Lathouwer et al., 2006; Frans et al., 2009). Samenvattend betekent het betrekken van een externe databron zoals administratieve gegevens dat men de tekortkomingen van beide databronnen kan beperken.

Maar er bestaan ook belangrijke nadelen wanneer men beide datasets wil matchen. Ten eerste is het linken van verschillende databronnen een duur, arbeidsintensief en een tijdrovend proces. Ten tweede meldde men correct in de *informed consent* van de vragenlijst van de gezinsenquête dat de data na verzameling gekoppeld werd met administratieve data. Indien personen afhaken omwille van deze boodschap, wordt consent bias geïntroduceerd (verschil tussen mensen die toestemming geven en mensen die geen toestemming geven), een andere vorm van selectie *bias* (Künn, 2015). Ten derde, en zoals eerder vermeld, bevatten de administratieve databronnen variabelen die op basis van administratieve definitie tot stand zijn gekomen. Het is maar zeer de vraag of deze variabelen bruikbaar zijn voor sociaalwetenschappelijk onderzoek naar het Vlaamse gezinsleven. De variabelen over inkomens of werk, bijvoorbeeld, matchen niet noodzakelijk met het inkomen of de werksituatie op het moment van bevraging (zie verder). Hier is ook de definitie van het gezin cruciaal: een gezin zoals gedefinieerd in het Rijksregister is geen sociologische definitie van het gezin (zoals eerder aangegeven in Hoofdstuk 1). De keuze in onderzoeksvragen zijn relevant om de bruikbaarheid van administratieve variabelen te evalueren.

In het laatste en vijfde scenario kan men nog een stap verder gaan door beschikbare administratieve data te koppelen aan een longitudinale studie. Het gaat dus om een combinatie van scenario's 2 en 3. Ook hier zijn twee mogelijkheden: een longitudinale versie van de gezinsenquête te organiseren (zoals in scenario 2), maar bepaalde gegevens laten invullen op basis van de administratieve data. Een tweede mogelijkheid is om een steekproef te trekken op basis van administratieve data (zoals in scenario 3) en die mensen op basis van administratieve data verder opvolgen, en dan via een postenquête (of via andere modi, zie Hoofdstuk 3) bijkomende vragen te stellen. Een bijkomstig voordeel waarin dit scenario zich onderscheidt van het vierde scenario, is dat het de ruimste mogelijkheid is om gegevens verklarend te analyseren, zowel voor objectieve als subjectieve individuele- of gezinskenmerken. Dat is voor beleidsmakers het meest bruikbare instrument om op robuuste wijze de effecten van het beleid op een aantal specifieke gezinstransities en -beslissingen op te volgen. Afhankelijk van de keuzes rond de doelpopulatie en de definitie van het gezin (zie Hoofdstuk 1) zou dit ook toelaten om aanstaande ouders

te betrekken bij het onderzoek, die dan opgevolgd kunnen worden tijdens de jaren van geboorte, opvoeding, zorg, en dergelijke. Een nadeel is dan weer dat men de respondent moet overtuigen om enerzijds meermaals deel te nemen aan dit onderzoek en anderzijds dat zijn of haar administratieve gegevens gekoppeld zullen worden. Dit kan voor respondenten als indringend overkomen en kan tot een hoge non responsgraad leiden. Daarnaast houdt het ook een lange termijnfinanciering in, men moet op voorhand zeker zijn van het budget om deze studie te kunnen blijven uitvoeren (Frans et al., 2009).

Kortom, er zijn verschillende keuzes mogelijk om het gezinsleven in kaart te brengen. Het is van essentieel belang om grondig te reflecteren over het zo diepgaand en breed mogelijk bereiken van gezinnen in Vlaanderen voor de steekproeftrekking plaatsvindt. Het koppelen van een externe dataset met gezinsenquêtegegevens kan een meerwaarde betekenen ten opzichte van de huidige werkwijze. Desalniettemin zal men, in geval men kiest voor een longitudinaal design, een knoop moeten doorhakken betreffende de onderzoekseenheid. De eenheid van analyse is in longitudinaal onderzoek doorgaans het individu eerder dan het huishouden, omdat het niet evident is om de eenheden van eenzelfde huishouden over een bepaalde tijd te volgen. De gezinsstructuur kan namelijk veranderen over de tijd heen, nieuwe gezinnen worden gevormd of bestaande gezinnen wijzigen qua structuur. Het individu als analyse eenheid is stabiel om te bestuderen in een longitudinale context, van waaruit men huishoudelijke kenmerken kan afleiden.

In hoofdstukken 4 en 5 van dit rapport maken we gebruik van een koppeling tussen de Datawarehouse en de gezinsenquête om de representativiteit van en de mate van non respons in de gezinsenquête uitgebreid te analyseren. Dit levert ons ook het testmateriaal om de proef op de som te nemen of de scenario's waarin administratieve data gekoppeld met surveygegevens effectief een meerwaarde kunnen betekenen voor nieuwe golven van de gezinsenquête. Het leert ons ook of, en in hoeverre, administratieve data bruikbaar zijn om gegevens vooraf in te vullen of om diverse vormen van non-respons te vermijden.

Samenvattend

In dit hoofdstuk bekeken we of het Rijksregister het meest adequate steekproefkader is om de gehele doelpopulatie in kaart te brengen. Dit Rijksregister kent immers enkele voordelen: het blijft de meest complete databron voor handen en dus het meest passende steekproefkader om van daaruit een steekproef te trekken. De beperkingen van dit kader zijn echter zaken waar men voldoende bij stil moet staan. Het Rijksregister 1) geeft een overzicht van de wettelijke woonsituatie, maar slaagt er niet in om de werkelijke situatie te capteren; 2) noch om gezinstransities vast te leggen en 3) sluit onvermijdelijk bepaalde groepen uit.

Via de steekproeftrekking wil men een representatief staal van de doelpopulatie kunnen bereiken om uitspraken te kunnen doen over álle Vlaamse en Nederlandstalige Brusselse gezinnen met minstens één kind jonger dan 25 jaar. Om dit te kunnen bewerkstelligen, selecteerde men voor de gezinsenquête een willekeurige steekproef van *alle* Vlaamse en (Nederlandstalige) Brusselse gezinnen met minstens één kind jonger dan 25 jaar enerzijds (basissteekproef) en een extra steekproef van personen met een herkomst buiten de Europese Unie (extra steekproef) anderzijds. Deze disproportionele stratificatie vinden we eveneens terug in andere bevestigingen, waar men stratificeert op basis van de meest relevante kenmerken die in hun onderzoeksopzet naar voren treden. De gezinsenquête verschilt met deze enquêtes doordat de extra steekproef in feite tot stand kwam door een samenwerkingsverband met het Agentschap Binnenlands Bestuur uit bezorgdheid dat personen met een niet-EU-herkomst vaak onvoldoende worden vertegenwoordigd in enquêtes. In functie van vervolgenquêtes kan men verkiezen om herkomst als criteria verder te zetten opdat men zeker voldoende diversiteit in herkomst blijft bereiken. Maar men kan ook andere criteria vooropstellen. De centrale vraag die hier naar voren wordt geschoven, is of én het steekproefkader én de steekproeftrekking die men kiest, het beste alternatief zijn om álle gezinnen te kunnen bereiken om een representatief beeld van hét gezinsleven te kunnen verkrijgen.

Daaropvolgend zien we, door een vergelijking met andere enquêtes, dat het Rijksregister niet noodzakelijk het enige steekproefkader hoeft te zijn, net zoals de *basissteekproef* niet de enige steekproeftrekking hoeft te zijn. De databank van Kind en Gezin zou bijvoorbeeld ook een gepast steekproefkader kunnen vormen om kwetsbare gezinnen met jonge kinderen te selecteren in een bijkomende steekproef.

Tot slot concludeerden we dat het cross-sectioneel bevragen van gezinnen minder geschikt is om een beeld te krijgen van de dynamieken en transitieën waar gezinnen zich in bevinden of in kunnen terecht komen. Indien men de langetermijneffecten van het gezinsbeleid wil bestuderen, of transitieën in het gezinsleven diepgaand wil opvolgen, moet men gezinnen op een meer dynamische wijze benaderen via een longitudinaal design. Maar ook dit design kent haar beperkingen. We bespraken een aantal mogelijke scenario's met al hun voor- en nadelen. In de volgende hoofdstukken worden een aantal van die scenario's getoetst aan de realiteit.

Hoofdstuk 3

De dataverzameling

Dataverzameling bij de gekozen doelpopulatie kan worden volbracht via verschillende onderzoeksmethoden, zoals een online enquête, een persoonlijk of face-to-face interview, een telefonische bevraging, een postenquête of een combinatie van deze methodieken. In deze sectie omschrijven we de *mode* of de methode die men hanteerde in de gezinsenquête, en de potentiële alternatieven die men kan gebruiken in het verzamelen van data over kwetsbare gezinnen.

1 De postenquête

Postenquêtes zijn in feite nog steeds bekend als één van de meest effectieve instrumenten wat betreft dataverzameling (Dillman, Smyth, & Christian, 2009). Het gaat om gestandaardiseerde vragenlijsten (i.e. elke respondent krijgt hetzelfde vragenboekje) en wordt aangeleverd via de post. Het is een *self-administered questionnaire* (SAQ) waarin men data verzamelt door de respondent volledig zelf in eigen beheer de vragenlijst te laten invullen en terugsturen. De respondent kan geen beroep doen op persoonlijke begeleiding bij het beantwoorden van de vragen. In dit opzicht zijn ze te vergelijken met online enquêtes, waarin men de verantwoordelijkheid legt bij de respondent om de vragenlijst in zijn volledigheid in te vullen en terug te zenden (Harrison, 2010). In deze sectie beschrijven we de voordelen en beperkingen van deze methode. Vervolgens gaan we in op de constructie van de vragenlijst en de mogelijke implicaties van de vragen en lay-out hiervan.

Een eerste, en vaak meest doorslaggevende **voordeel** van het hanteren van een postenquête is het kostenplaatje. Deze papieren bevraging kost minder dan een telefonisch of persoonlijk interview. Men drukt namelijk de kost van het aanwerven van interviewers die de bevestigingen afnemen. De grootste kost is het opstellen, afprinten, opsturen en verwerken van de vragenboekjes. De enige methode die meer kostenefficiënt is, is een online enquête (*zie verder*) (Dillman et al., 2009; Harrison, 2010).

Een tweede argument ten voordele van deze papieren bevraging is dat het vaak om een simpel, conventioneel en gemakkelijk verstaanbaar formaat gaat. De respondent kan de vragenlijst invullen waar en wanneer het hem of haar uitkomt (Harrison, 2010). Technisch kan een postenquête ook visuele stimuli aanreiken in de vorm van grafieken of tabellen, op voorwaarde dat deze door het grote publiek begrepen worden. Je hebt een vaste structuur in de vragen en het formaat van de vragenlijst. Merk wel dat men extra aandacht, tijd maar ook kosten moet schenken aan het aantrekkelijk maken van de lay-out. Op een visuele en inhoudelijke manier de respondent stimuleren om de volledige vragenlijst in te vullen, is een opdracht die vaak onderschat wordt (Schockaert et al., 2012).

Een derde voordeel, gekoppeld aan de specifieke doelpopulatie van deze enquête, is dat we een uitgebreid steekproefkader ter beschikking hebben. Men kan een postenquête namelijk enkel effectief inzetten in geval de adressen van de steekproefeenheden ter beschikking worden gesteld. Dankzij het bevolkingsregister kunnen we beroep doen op een lijst met de adressen van de getrokken steekproef. Zulk een lijst is vaak eerder beschikbaar dan wanneer men telefoonnummers of e-mailadressen moet verzamelen (Harrison, 2010).

Tenslotte is de postenquête een methode die de nadelen van de alternatieve *modes* tegengaat (zie verder). Zo kan er bij een face-to-face bevraging fouten optreden zoals een interviewereffect, i.e. de attitude of het gedrag van de interviewer beïnvloedt de respondent in het beantwoorden van de vragen. Een ander voorbeeld is het sociaal wenselijk antwoorden van respondenten wanneer men tegenover een interviewer zit, terwijl men in een postenquête sneller gevoelige of persoonlijke informatie wenst te beantwoorden. Specifieke bevolkingsgroepen, zoals ouderen, hebben bijvoorbeeld ook moeite met het gebruiken van het internet om een bevraging in te vullen. Anderen kunnen ook wantrouwig zijn over het delen van persoonlijke informatie online (Stern, Bilgen, & Dillman, 2014) en voelen hun anonimiteit eerder gegarandeerd via een postenquête (Harrison, 2010).

Toch heeft deze manier van bevraging ook enkele **beperkingen**, wat het niet noodzakelijk de meest effectieve of efficiënte manier maakt om data te verzamelen over de gekozen doelpopulatie. Ten eerste gaat een postenquête ervan uit dat de aangeschreven persoon kan lezen en schrijven. In deze bevraging sluit men laaggeletterden uit, alsook personen die de taal van de vragenlijst niet (voldoende) machtig zijn.

Ten tweede gaat het om een eenzijdige interactie waarbij de onderzoeker de vragen en de manier van vraagstelling bepaalt, terwijl deze van de respondent verwacht te antwoorden binnen de mogelijke, en vaak schaarse antwoordcategorieën (Blasius & Thiessen, 2012). Ten derde is het formaat van een postenquête geen flexibel instrument. Dat is de schaduwzijde van een eerder vermeld voordeel van deze mode. Het stellen van bijvragen over specifieke thema's is zeer moeilijk. Denk hierbij aan de vraag 'indien ja, ga dan naar vraag 2'. Wat men niet in persoonlijke of telefonische bevragingen moet doen, maar wat in een postenquête absoluut noodzakelijk is, is de extra aandacht die uitgaat naar het design en de lay-out van het vragenboekje opdat de lezer begrijpt wat de volgende vraag is. In de andere soorten bevragingen moet de respondent zelf niet achterhalen wat de volgende vraag is, maar krijgt hij of zij deze gewoon voorgeschoteld. Denk maar aan online enquêtes die de respondent automatisch doorverwijzen naar de volgende relevante vraag (Stern et al., 2014). Bovendien moet alle informatie beschikbaar zijn in de vragenlijst zelf. De respondent moet zich zelf inlezen over het doel van de studie, de uitnodiging tot deelname, de informatie omtrent de privacybescherming en potentiële koppeling aan administratieve gegevens, specifieke instructies omtrent het beantwoorden van de vragen, en dergelijke. Dit vraagt in se een extra inspanning van de respondent wat kan resulteren in item- of unit non respons (Zie Hoofdstuk 4) (Schockaert et al., 2012).

Ten vierde beargumenteert Dillman (2012) dat item non respons in een postenquête hoger ligt dan bij andere bevragingen. De respondent geeft vaker een niet-volledig ingevulde vragenlijst aan de onderzoeker. Er is geen interviewer om erop toe te zien dat de respondent zo volledig mogelijk te werk gaat en geen vragen overslaat.

Ten vijfde is er ook geen interviewer om toe te zien dat het daadwerkelijk de aangeschreven persoon is die de vragenlijst beantwoordt. De uitnodiging tot deelname moet specifiek op naam gericht zijn aan desbetreffende persoon, en de onderzoeker moet naderhand kunnen achterhalen of het wel degelijk de aangeschreven persoon is die de vragenlijst invulde.

Een zesde knelpunt is dat men met een vragenboekje via de post steevast enkele personen niet bereikt. Denk maar aan personen die zich juist op dit moment in een transitie begeven, als gevolg van een echtscheiding of een verhuis, e.d. (Stern et al., 2014).

Bovendien vinden we veel tegenstrijdige literatuur over de responsgraad gekoppeld aan deze *mode*. Krosnick (1991) omschrijft de respons kwaliteit als het resultaat van de moeilijkheid van de vragenlijst en de taak *an sich*, de motivatie om deel te nemen aan de enquête en de cognitieve vaardigheden van de

respondent. Telefonische of persoonlijke interviews maken dat de interviewer zelf actief betrokken is in het verkrijgen van unit- en item respons. De interviewer kan extra informatie verschaffen bij moeilijke vragen, kan de respondent motiveren om de vragenlijst af te ronden en algeheel overtuigen om deel te nemen (Schockaert et al., 2012).

Verder staan we stil bij het design van de **vragenlijst** zelf. Gegeven de lengte van de vragenlijst en in dit geval dus de dikte van het vragenboekje, is het cruciaal om stil te staan bij de vraag of men zo kort en bondig mogelijk te werk is gegaan. Zoals eerder vermeld vraagt deze *mode* (i.e. de postenquôte) een extra inspanning van haar respondenten om zich in te lezen in de instructies en de vragenlijst, waardoor het noodzakelijk is om het vragenboekje zo kort mogelijk op te stellen opdat het zo min mogelijk tijd van de respondent inneemt. Het gaat om een *trade off* tussen theoretische overwegingen (zoveel mogelijk thema's verwerken in de vragenlijst zodat we de doelpopulatie breed kunnen onderzoeken) en praktische overwegingen (i.e. budget, maar ook de inspanning zo klein en timing zo kort mogelijk houden voor de respondent). Aan de vragenboekjes van de gezinsenquête ging een lange weg vooraf: verschillende experts werden al vanaf het startpunt betrokken om de vragenlijst te vormen, maar ook om de meerwaarde en opportuniteiten van deze enquête te bespreken. Gebaseerd op een initieel opgestelde projectfiche waarin de onderzoeksvragen geformuleerd werden, startte het onderzoeksteam met het opzetten van een lijst met concepten die nadien werden voorgelegd aan de begeleidende stuurgroep. Daaropvolgend construeerde men een eerste draft van de vragenlijst die werd voorgelegd aan 40 vrijwillige proefpersonen (Audenaert & Stuyck, 2018a). In deze testronde konden de testrespondenten aangeven in welke mate de vragen onduidelijk of moeilijk gesteld waren, of de vragenlijst te lang was, of het interessante vragen waren, of de volgorde van vervolgvragen correct was, en dergelijke. Deze pilotstudie wordt sterk aangemoedigd in de *survey* literatuur: in elke *mode* is het namelijk essentieel om zowel de vragen als de lay-out voor te leggen bij een testpubliek vooraleer je start met de grootschalige uitvoering van de bevraging. Het is dus financieel, maar ook zeker inhoudelijk voordelig om deze test door te voeren vooraleer deze maar ook toekomstige enquêtes (Harrison, 2010). Een aanbeveling die men maakt in de uitvoering van een pilotstudie is om steekproefpersonen te selecteren die grosso modo representatief staan voor de doelpopulatie. Het zijn net deze personen die je wil bereiken, waardoor hun evaluatie van de vragenlijst belangrijk is om te implementeren in de finale versie. Interessant kan dan ook zijn om in deze pilotstudie ook de kwetsbare gezinnen expliciet mee te nemen. Peilen naar hun ervaring met het invullen van de vragenboekjes kan bijzonder interessant zijn om de respons op relatief goedkope wijze in deze doelgroep te verhogen.

Naast deze pilotstudie toetste de organisator van de gezinsenquête af bij vzw *Wablieft* of de vragenlijst was opgebouwd uit helder en toegankelijk taalgebruik. Opnieuw zou men ook hier andere organisaties kunnen betrekken die specifiek deskundig zijn in het bereiken van verschillende kwetsbare bevolkingssegmenten, zoals het Netwerk tegen Armoede.

In de wetenschappelijke literatuur werd in toenemende mate aandacht geschonken aan het visuele aspect van de enquête. Tijdens de jaren '90 vonden surveymethodologen dat kleine, en soms bijna onzichtbare manipulaties in het design een sterke invloed kunnen hebben op de antwoorden van de respondenten. Hun onderzoek toonde aan dat de lay-out of het visuele aspect van een enquête minstens even invloedrijk kan zijn dan de interviewers- of verwoordingseffecten (Jenkins & Dillman, 1997; Stern et al., 2014). Concreet gaat het om de symbolen die men gebruikt, de kleuren, lettertype en lettergrootte, ruimte die men laat tussen de vragen, en dergelijke meer. Zeker wat betreft het stellen van vervolgvragen, is een heldere lay-out essentieel om een respondent in een postenquôte tot de volgende vraag te leiden. In geval men correct gebruik maakt van visuele elementen in een postenquôte kunnen ze de respondent

begeleiden van de ene naar de ander vraag. Indien er onvoldoende aandacht naar uitgaat, kunnen ze onbedoelde effecten hebben en verwarring creëren zodat de respondent afhaakt of de vraag anders interpreteert en dus foutief antwoordt (Harrison, 2010). Ook dit is een element dat kan opgenomen worden in de pilootstudie. In het huidige opgestelde evaluatieformulier van de eerst doorgevoerde pilootstudie van de gezinsenquête bevrage men voornamelijk of de documenten duidelijk zijn, maar gaat men niet in op het vormelijke aspect en in welke mate men het vragenboekje de respondent ook aansprak. Dit vormelijke aspect kan het onderzoeksteam voor vervolgenquêtes aftoetsen bij een nieuwe testgroep die bestaat uit zeer heterogene profielen: niet enkel kwetsbare gezinnen zijn hierin belangrijk, maar ook een diverse groep wat betreft leeftijd en herkomst.

Het opnemen van modules

Gegeven de lengte van een vragenlijst en de tijd nodig om de vragen in te vullen, is het belangrijk om de vragen kort en bondig op te stellen. Kiezen is verliezen: de keuze voor één thema maakt dat een andere topic mogelijk niet (uitgebreid) besproken wordt. Een brede en gevarieerde set aan thema's over gezinnen wordt zo ingeperkt, waardoor potentieel interessante thema's om kwetsbare doelgroepen te identificeren niet of amper worden vermeld. Gegeven de eerder vernoemde afruil tussen theoretische overwegingen (vragenlijst verlengen met nieuwe topics) versus praktische overwegingen (tijd, budget) kan het interessant zijn om het aandeel van reeds veelgevraagde thema's te laten afnemen in functie van nieuwe, ongekende topics via zogenaamde ad hoc modules.

Een bekend voorbeeld van een enquête die gebruik maakt van modules is de European Union Statistics on Income and Living Conditions (EU-SILC) survey. De bevraging bestaat uit een kerndeel waar achtergrondvariabelen jaarlijks bevrage worden en een ad-hoc module die jaarlijks wijzigt. Enkele voorbeelden hiervan zijn materiële deprivatie (in 2009 en 2014), intra- huishoudelijk delen van middelen (2010), welzijn (2013), gezondheid (2017), of drempels bij het gebruik van dienstverlening (2016). Bircan et al. (2017) schreven een uitvoerige analyse over de nood aan kwaliteitsvolle data inzake armoede en levensomstandigheden en beschreven hoe de EU-SILC als enquête nog optimaler kan worden ingezet om desbetreffende data te verzamelen. Via een bevraging bij ervaringsdeskundigen verkregen de onderzoekers een overzicht over thema's die jaarlijks bevrage moeten worden, en thema's die geïntegreerd kunnen worden in ad hoc modules. Om armoede en sociale uitsluiting te kunnen identificeren, suggereren de onderzoekers om de focus te verbreden van een eenzijdige financiële invalshoek naar een multidimensionele benadering waarin men de *mogelijkheden* die een persoon krijgt, bestudeert. Daarbij aansluitend raden Bircan en collega's aan om meer aandacht te schenken aan sociaalpsychologische en culturele dimensies gekoppeld aan sociale cohesie en ongelijkheden, maar ook aan het gebruik van en ervaring met bepaalde diensten en beleid. Bircan et al. (2017) bestudeerden specifiek welke topics kunnen worden toegevoegd aan de EU-SILC vragenlijst opdat men het multidimensionale concept dat 'armoede' is, operationeel kan maken (zie Tabel 5).

Tabel 5 Potentiële thema's voor jaarlijkse bevraging of modules: een voorbeeld van Bircan et al. (2017, p. 7)

(Voorbeelden van) onderwerpen	In de jaarlijkse survey	In een module
Financiële informatie (vermogen, sparen, schulden, uitgaven, en dergelijke)		✓
Tijdsbestedingen (bijvoorbeeld zorg verantwoordelijkheden)		✓
Multidimensionele benadering op armoede: competenties, materiële welvaart, materiële en sociale deprivaties, gezondheid, en dergelijke.	✓	✓
Gebruik van (de voordelen van) sociaal beleid, bijvoorbeeld sociale diensten, sociale bijstand, uitkeringen, sociale huisvesting, activeringsmaatregelen, enzovoort	✓	

Geïnspireerd op het overzicht van Bircan en collega's kan men voor vervolgenquêtes van de gezinsenquête eveneens reflecteren over thema's die men wil bevragen in elke editie en thema's die men ad hoc wil bevragen. Per editie kan men dan concentreren op één specifiek thema. Ook in het geval van niet-jaarlijkse enquêtes zoals bij de gezinsenquête het geval is kan dit nuttig zijn, al moet het aantal roterende modules dan uiteraard beperkt zijn.

Een andere studie die frequent gebruik maakt van inhoudsgerichte modules, is het Nederlandse Permanent Onderzoek Leefsituatie (verder: POLS). Via één identieke basisvragenlijst voor alle respondenten tracht men inzicht te verwerven in de leefsituatie van de Nederlandse bevolking. De vragenlijst bestaat uit vragen omtrent de socio-economische (opleiding, beroep, bedrijfsklasse) en demografische (geslacht, leeftijd, samenstelling huishouden, nationaliteit, burgerlijke staat, etc.) eigenschappen, maar ook kernindicatoren die belangrijke aspecten van de leefsituatie naar voren schuiven (Sociaal en Cultureel Planbureau, 2007). Het onderzoeksopzet bestaat uit een schillenmodel, waarbij de vragen verdeeld worden over vier 'schillen' (Tabel 6).

Tabel 6 Het schillenmodel van het Permanent Onderzoek Leefsituatie (Sociaal en Cultureel Planbureau, 2007)

Schillen	Omschrijving
Schil 1	<ul style="list-style-type: none"> - Vaste vragenlijst: Demografische kenmerken (geslacht, burgerlijke staat, leeftijd, samenstelling huishouden, nationaliteit, woongemeente) en socio-economische kenmerken (opleiding, socio-economische positie, bedrijfsklasse, beroep) - Gericht op de gehele steekproef - Methode: CAPI
Schil 2	<ul style="list-style-type: none"> - Vaste vragenlijst: kernindicatoren leefsituatie (wonen, gezondheid, leefstijlen, tijdsbesteding, milieu, arbeid, politieke betrokkenheid, (on)veiligheid, maatschappelijke participatie) ➔ Gericht op de gehele steekproef - Variabele vragenlijst: lichamelijke- en psychische belemmeringen (voor personen die worden geselecteerd met een specifieke eigenschap) - Methode: CAPI
Schil 3	<ul style="list-style-type: none"> - Thematische vragenlijstmodules (bv. Gezondheid) - Methode: CAPI en PAPI
Schil 4	<ul style="list-style-type: none"> - Thematische vragenlijstmodules

In de eerste schil bevaart de enquête de gehele steekproef naar demografische en socio-economische kenmerken. Ook bevaart men de steekproef naar de inkomens(componenten). Aanpassingen in deze schil gebeuren enkel wanneer definities van de betrokken variabelen wijzigen. De variabelen die men in schil 1 bevaart worden geïnterpreteerd als achtergrondvariabelen om enerzijds de samenhang met andere variabelen te kunnen bepalen en anderzijds meteen een onderscheid te kunnen maken tussen verschillende relevante bevolkingsgroepen. In de tweede schil bevaart men de gehele steekproef naar kernindicatoren van verschillende leefsituatieonderwerpen. Dit zijn in feite variabelen waar men met een zekere regelmaat informatie over wil verkrijgen. In diezelfde schil is er ruimte voorzien voor variabele vragenblokken met 'screeningsvragen'. De personen die hierin worden geselecteerd krijgen in een later stadium (i.e. schil vier) bovenop de standaard vragenlijst additionele vragen voorgelegd. Zowel in schil 1 als in schil 2 maakt men gebruik van *computer-assisted personal interviewing* (of CAPI) waarin men voor 15 minuten de respondenten bevaart. In de derde en vierde schil stellen de onderzoekers meer diverse onderwerpgerichte vragen. In de modules voorziet men een interviewduur van 30 minuten. Op verzoek van derden kan de steekproefomvang voor specifieke modules worden vergroot. Het aantal onderwerpgerichte modules varieert jaar op jaar, en afhankelijk van de fluctuerende betrokkenheid van externe partners wordt dit design herzien. Via deze schillen willen de onderzoekers tegemoetkomen aan de vraag naar een combinatie van gegevens over verscheidene onderwerpen bij subgroepen in de populatie (Sociaal en Cultureel Planbureau, 2007). Hoewel deze studie in 2009 werd beëindigd, kan men via dit schillenmodel reflecteren over enerzijds de standaard thema's die men blijvend bevaart bij de respondent, en anderzijds de aanvullende thema's die men kan bevragen door middel van samenwerking met derden.

Ook de European Social Survey (verder: ESS) maakt gebruik van roterende modules in haar vragenlijst. De vragenlijst bestaat uit een kernsectie en een roterende sectie. De kernsectie bestaat, zoals bij de bovengenoemde POLS-bevraging, uit achtergrondvariabelen. Om de roterende sectie op te stellen, selecteert men in elke wave onderzoekers uit verschillende deelnemende landen om twee roterende modules in de vragenlijst te construeren. Enkele thema's die tot nu toe behandeld werden sociale ongelijkheden in gezondheid; tewerkstelling, gezin en welzijn; vergrijzing; persoonlijk en sociaal welzijn; enzovoort. De geselecteerde onderzoekers dienen een aanvraag in om een terugkerende of nieuwe module te installeren in de eerstvolgende ESS- wave.

Tot slot vinden we, sinds 2000, eveneens een ad hoc module in de vragenlijst van de Enquête naar Arbeidskrachten (EAK). Het gaat hier over een beperkte set van vragen die zich specifiek richten op een arbeidsmarkt gerelateerd onderwerp, dewelke elk jaar wijzigt. In 2017 ging het, bijvoorbeeld, over vragen over het werken als zelfstandige om zich inhoudelijk te verdiepen in de arbeidsmarktsituatie van dit specifieke beroepsstatuut. In 2011 werd een module 'mobiliteit' toegevoegd, specifiek in het kader van de census (Termote & Depickere, 2018).

2 Alternatieve modi

Is de postenquête het beste alternatief om alle gezinnen te bereiken, kwetsbare gezinnen inclusief? In wat volgt bespreken we wat de literatuur stelt over alternatieven. Algemeen kunnen we een onderscheid maken tussen twee types van *modes*: één waarbij een interviewer wordt ingezet om data te verzamelen, en één waar men niet beroep kan doen op een interviewer, ook wel *self-administered surveys* genoemd (De Leeuw et al., 2008).

Eén van de manieren waarop een interviewer wordt ingezet, is via **telefonische** bevestigingen. Dit staat bekend als een relatief kostefficiënte manier van dataverzameling, maar wordt enkel ingezet voor bepaalde doelpopulaties. Onderzoekers maken gebruik van telefonische bevestigingen in geval men een aanvulling wil op een andere *mode*, of in geval de doelpopulatie in zijn geheel telefonisch te bereiken is. Eén van de manieren om een telefonische bevestiging uit te voeren, is via de *random digital dial* (RDD) methode waarbij men random telefoonnummers selecteert om op te bellen. Deze methode gebruikt men veeleer in electoraal onderzoek of het selecteren van een experimentele controlegroep. Een andere methode is CATI, of *computer-assisted telephone interviewing*, waarbij men aan de hand van instructies op de computer een telefonisch interview afneemt (De Leeuw et al., 2008).

Sinds de jaren '2000 zien we een afname van de responsgraad bij deze specifieke interviews (De Leeuw et al., 2008) mede door het wijdverspreide gebruik van voicemail en de barrière van de *caller identification* (in geval het een onbekend nummer is, heeft men de neiging niet op te nemen)(Groves & Couper, 1998; Stern et al., 2014). Bovendien neemt het aantal Belgische huishoudens met een vaste telefoonlijn almaar af, waardoor de telefonische enquêtes aan populariteit inboeten. Een extra moeilijkheid in deze *mode* is het feit dat verscheidene steekproeftrekkingen uitgaan van telefoonnummers op basis van vaste telefoonlijnen, eerder dan van mobiele telefoonnummers. Onderzoek waarschuwt meermaals dat deze methode resulteert in een zeer selectief bereik: jongeren, laaggeschoolden, huishoudens in een transitie (bv. Door een echtscheiding) en alleenstaanden zijn groepen die worden uitgesloten (Blumberg et al., 2011; Pickery, 2010b).

Een tweede alternatief op postenquêtes zijn de **face-to-face of persoonlijke bevestigingen**. Algemeen is het een belangrijke en efficiënte manier om data te verzamelen. Essentieel in deze *mode* is het directe,

persoonlijke contact tussen een interviewer en de respondent, het gebruik van een gestandaardiseerde vragenlijst, en de specifieke verdeling van de taken waarin de interviewer de vragen stelt en de antwoorden registreert. Deze *mode* verschilt dus van de post- en telefonische bevraging doordat er sprake is van direct contact met de interviewer. Dit contact zorgt voor mogelijkheden, zoals het kunnen stellen van vragen in geval een instructie of vraag niet duidelijk is, maar draagt ook risico's met zich mee, zoals het interviewer effect (i.e. de aanwezigheid van de interviewer kan een effect of invloed hebben op de antwoorden van de respondent), het niet of foutief bereiken van de respondent, het niet of fout stimuleren van respondenten om vragen te interpreteren en beantwoorden, en het niet of fout registreren van desbetreffende antwoorden (Beullens & Loosveldt, 2016; Boyd & Westfall, 1985). Bovendien kunnen deze persoonlijke interviews impliciet sociaalwenselijke antwoorden aanmoedigen, eerder dan de werkelijke attitudes van een persoon doordat de respondent niet graag iets wil toegeven aan de persoon die recht tegenover hem zit. Men neigt ook eerder naar het akkoord gaan met enkele stellingen (bv. zodat het gesprek des te sneller verloopt), eerder dan in discussie te treden met de interviewer (Schuman & Presser, 1981; Stern et al., 2014). Daarnaast is deze mode over het algemeen duur (De Leeuw et al., 2008). Eén vorm van deze bevraging is CAPI, of het *computer-assisted personal interview*, waarin de respondent of de interviewer gebruik maakt van een elektronisch hulpmiddel (bv. iPad, computer, laptop, en dergelijke) om de antwoorden te registreren. Indien er geen interviewer aanwezig is, is er sprake van CASI of het *computer-assisted self interviewing* (De Leeuw et al., 2008). In de MEqIN- survey van Belspo¹⁴ werkte de onderzoekers met een professionele survey agentschap (KANTAR) om CAPI- bevestigingen af te nemen. In deze enquête wil men namelijk breed en diep ingaan op verschillende levensdomeinen van de respondent, waarvoor het persoonlijk bevragen van de respondenten de meest ideale methode bleek (Maniquet et al., 2018). Ook in de Enquête naar Arbeidskrachten (EAK) ondervroeg men (voor 2017) de respondenten face-to-face via CAPI. Enkel voor huishoudens die uitsluitend uit personen van 65 jaar en ouder bestonden, liet men een telefonische bevraging toe (via de CATI- methode)(Termote & Depickere, 2018).

Naast deze knelpunten van het persoonlijk bevragen van de respondent, blijft het een onmiskenbare methode om specifieke bevolkingssegmenten te bereiken. Het is een manier om aandacht te schenken aan de asymmetrische relatie tussen de onderzoeker en de te onderzoeken kwetsbare populatie en aan de deconstructie van deze aanwezige machtsrelatie. In een onderzoek van het Vlaams Armoedesteunpunt organiseerde de onderzoekers overlegmomenten met armoedeverenigingen en ervaringsdeskundigen waaruit bleek dat deze partners, in het onderzoek naar kwetsbare populaties, een voorkeur geven voor kwalitatief onderzoek. Ook eerder onderzoek pleitte voor een kwalitatieve benadering in het armoede-onderzoeksveld (Driessens, 2003; Lemmens & Groenez, 2015b). Interviewers in face-to-face- bevestigingen zouden bovendien het hoogst scoren als het aankomt op responscijfers (De Leeuw et al., 2008). Een nieuw fenomeen om kwetsbare doelgroepen te benaderen, is door het rekruteren van *peer* interviewers. Dit zijn personen met eenzelfde etnische-, culturele- of specifieke taalachtergrond als de personen die je wenst te bereiken (Schepers, Juchtmans, & Nicaise, 2017). Werken met deze personen maakt dat je het interview zelf laagdrempelig en toegankelijk maakt voor de doelgroep. Men tracht een zekere vertrouwensrelatie op te bouwen waarin de respondent aangemoedigd wordt deel te nemen en vragen kan stellen indien iets onduidelijk blijkt. Het uitgangspunt van deze methode is dat geïnterviewde personen zich gemakkelijker openstellen en ervaringen uitwisselen tegen personen die bijvoorbeeld ook het Frans machtig zijn, of een persoon die van hetzelfde geslacht is, en dergelijke. De literatuur heeft het over de meerwaarde van een geheel gedeelde (culturele) identiteit (Lemmens & Groenez, 2015b). De rekrutering in de eerder besproken uitbreiding van het JOnG!-onderzoek vond bijvoorbeeld plaats via lokale verenigingen waar

¹⁴ Voor meer gedetailleerde informatie over deze studie verwijzen we naar het inhoudelijk rapport.

armen het woord nemen. Kwetsbare gezinnen maken deel uit van een sociaal netwerk, en het is dit netwerk waar men beroep op doet om de steekproefpersonen te kunnen bereiken. Respondenten hadden bovendien tijdens de bevraging zelf de mogelijkheid om een vertrouwenspersoon uit te nodigen om bij dit gesprek aanwezig te zijn. Het ging om gestandaardiseerde vragenlijsten die mondeling werden afgenomen, maar er werd – specifiek voor deze kwetsbare steekproef - expliciet en meerdere malen vermeld dat men enkel de vragen diende te beantwoorden die men wilde beantwoorden (Lemmens & Groenez, 2015b). Schockaert et al. (2012) reflecteerden dan weer over het effect van de persoonskenmerken van de betrokken interviewer bij het bereiken van de respondent. Volgens deze studie zouden genderaspecten voornamelijk van belang zijn in samenhang met de etniciteit van de steekproefpersoon. Men raadt dan ook aan om de interviewer en respondent qua gender op elkaar af te stemmen. Voor vervolgenquêtes van de Gezinsenquête zou men bijvoorbeeld beroep kunnen doen op de ervaring die werd opgedaan bij het organiseren van de survey Samenleven in Diversiteit.

Daaropvolgend kan men via het inschakelen van interviewers of enquêteurs ook waardevolle inzichten verkrijgen die men niet verwerft via andere *modes*. Zo kunnen de ervaringen die de interviewers beleven in het overtuigen en bereiken van de steekproefpersonen een zicht geven op waarom non respons plaatsvindt (Schockaert et al., 2012). Zo zetten Schockaert et al. (2012) een focusgroep op met enquêteurs van de EU-SILC om hiaten in het bevragings- en contactopnameproces te achterhalen. De meerderheid van deze interviewers beschikt over jarenlange ervaring met het bevragen van individuen en/of huishoudens. Over enkele knelpunten kwamen zij overeen. Zo is er een consensus over de moeilijkheidsgraad die verschillende aanvullende documenten hebben, vooral als het gaat om het taalgebruik. De instructies en bijkomende informatie moet zo beknopt en helder mogelijk omschreven worden. Bovendien begrepen de respondenten enkele begrippen van de vragenlijst niet altijd even goed, waardoor de item non respons niet noodzakelijk werd veroorzaakt door aarzeling of onwil om een antwoord te geven, maar evengoed met het niet verstaan van de vragen. Let op dat men hier werkt met interviewers waar respondenten onmiddellijk beroep op kunnen doen in geval men een begrip niet begrijpt. Wanneer men vragen beantwoordt via een postenquête, is het achterhalen van de betekenis van de voor hem of haar complexe begrippen nog moeilijker. Werken met face-to-face bevragingen kan de respondent een extra duwtje in de rug geven, niet enkel om algeheel deel te nemen, maar ook om zo compleet en correct mogelijk de vragenlijst te beantwoorden.

Parallel met de afname van de participatiegraad bij persoonlijke telefonische of face-to-face interviews, trad het optimisme naar de voorgrond dat het internet een nieuw middel zou worden om vragenlijsten rond te kunnen sturen. Vandaag is het afnemen **online enquêtes** van gemeengoed geworden in verschillende disciplines. Deze *mode* is de facto de goedkoopste en snelste manier om data te verzamelen. Bovendien is er ook hier geen sprake van interviewer effecten (i.e. effect op de respons als gevolg van de aanwezigheid van een interviewer). Een ander voordeel is de creativiteit die hierbij te pas kan komen: men kan internet enquêtes interactief maken met behulp van grafische toepassingen. Men kan de vragenlijst vaak invullen via een smartphone, een vaste computer of laptop, een iPad, en dergelijke (Couper, 2011). Daarnaast wordt de respondent van de ene vraag naar de andere vraag geleid, terwijl vervolgvragen in postenquêtes vaak als complex worden gezien en respondenten doen afhaken.

Toch is dit, omwille van verscheidene redenen, eveneens een instrument waar het responscijfer of de kwaliteit van de verzamelde gegevens niet automatisch hoog blijken te liggen. Ook hier blijft het een uitdaging om de doelpopulatie te kunnen bereiken. Het gebruik van het internet impliceert dat een huishouden daadwerkelijk toegang hiertoe heeft enerzijds, en anderzijds competenties bezit om dit medium te gebruiken (Stern, Adams, & Elsasser, 2009; Stern et al., 2014). De grootste valkuil in deze manier

van bevraging is dan ook de digitale kloof: de oververtegenwoordiging van personen die voldoende technologische competenties bezitten, versus de personen die botsen op een zekere drempel bij het gebruik van een computer en/of het internet en deze competenties niet bezitten. Evenals het gebruik van het internet, zijn ook hier de jongeren de respondenten die het snelst mee zijn met deze technologie. In geval de enquête zich richt op een jonge doelpopulatie, is een websurvey dus een adequaat instrument om data te verzamelen. In geval de enquête zich richt op een oudere of bredere leeftijdsgroep, is het enkelvoudige gebruik van een web survey mogelijks onvoldoende om een representatief beeld te schetsen van de doelpopulatie. Volgens Stern and Dillman (2006) zijn het voornamelijk de hooggeschoolde, welvarende, politiek en maatschappelijk geëngageerde personen die het grootste aandeel van de frequente Internet- gebruikers uitmaken. Rookey, Hanway, and Dillman (2008) toonden aan dat respondenten in hun huishoudpanel zonder internet over aanzienlijk lagere huishoudinkomens en opleidingsniveau beschikten vergeleken met de internetgebruikers. Deze laatste groep bleken ook vaker getrouwd te zijn, jonger en voltijds tewerkgesteld. In België heeft een aanzienlijk groot aandeel huishoudens geen toegang tot het internet. Volgens een enquête van de Algemene Directie Statistiek heeft 15% van de Belgische huishoudens thuis geen internetverbinding, waarvan 13% aangeeft dit nooit te hebben gehad. Let wel, deze percentages liggen voornamelijk hoog bij huishoudens zonder kinderen, eerder dan bij huishoudens met kinderen (Statbel, 2017). Een andere beperking van deze *mode* is dat er een gebrek is aan een compleet steekproefkader waarin men e-mailadressen voor desbetreffende gezinnen kan achterhalen.

3 Mixed-mode survey

Een 'mixed-method' of 'mixed-mode' survey maakt gebruik van meerdere *modes* om data te verzamelen. Hierin worden de voordelen van de eerder genoemde *modes* gecombineerd om een hoge responsgraad te bekomen (Dillman et al., 2009). Men kiest er bijvoorbeeld voor om eerst telefonisch te bevragen of de steekproefpersoon wenst deel te nemen aan een persoonlijke of schriftelijke bevraging. Of men stuurt een postenquête op en probeert op die manier de niet-respondenten te bereiken met een uitnodiging voor deelname aan een online survey, enzovoort. De keuze voor meerdere methoden in het bereiken van specifiek kwetsbare doelgroepen is niet onbekend in de literatuur. *“Geographically isolated groups may be easy to reach by telephone; mobile-only households may have no problem with a web-based survey; linguistic minorities may be quite happy to complete a questionnaire translated into their language; and the illiterate may enjoy a conversation with a face-to-face interviewer. Mixed-mode surveys aim to overcome the fact that different groups of people may be more or less easy to reach and more or less willing to respond in a particular mode. More generally, securing participation by the hard to reach or hard to survey will depend on the design and topic of the survey and the efforts one is willing to make (Stoop, 2014, p. 225).”*

Vandaag is het duidelijk dat geen enkele op zichzelf staande methodiek adequaat is voor alle segmenten van de doelpopulatie en voor elke context en/of situatie. De manier waarop we communiceren is en blijft voor verandering vatbaar, een gegeven dat onderzoekers moeten erkennen in het opstellen van een enquête. Verschillende doelgroepen vragen dan ook verschillende methodieken om hen te bereiken. Opteren voor verschillende manieren van dataverzameling- en verwerking reduceert bovendien de kans op de fouten die kunnen optreden (Stern et al., 2014). Vanuit deze kennis is het dan ook essentieel om voldoende aandacht te schenken aan het instrument van de dataverzameling, in dit geval de postenquête. Met het oog op vervolgenquêtes is het cruciaal dat men dit instrument uitbreidt of optimaliseert om

zodanig meerdere segmenten van de doelpopulatie te kunnen bereiken enerzijds en hierdoor de respons te verhogen anderzijds om preventief vertekening tegen te gaan.

Dillman et al. (2009) kenschetsen de voordelen van het gebruiken van een *mixed-mode survey design*. Ten eerste zou het een lager kostenplaatje betekenen. Voorbeelden in de literatuur omschrijven steeds één mode (bijvoorbeeld postenquête) waar respondenten na een eerste contactopname een andere *mode* (bijvoorbeeld online enquête) krijgen aangeboden. Ten tweede, en daarbij aansluitend, kan deze afwisseling ervoor zorgen dat meerdere personen worden bereikt. Een bekend voorbeeld is het starten met een online enquête, om nadien een postenquête aan te bieden voor de personen die er niet in slagen om de bevraging online te voltooien. Ten derde slaagt deze *mixed-mode* methode erin om de data te verzamelen in een versneld tempo. Men kan aanbieden om een persoonlijk interview af te nemen, of een telefonisch gesprek te voeren wanneer het voor de respondent het beste past. Een vierde, en meest besproken voordeel is de impact op de respons in geval men gebruik maakt van meerdere *modes*. Door het multiple gebruik van meerdere manieren om de doelpopulatie te bereiken kan men zich ook aanpassen aan de voorkeuren van de aangeschreven personen. In geval men op voorhand de preferentie van een (groep van) respondent(en) kan achterhalen en toepassen, neemt de snelheid van het dataverzamelingsproces toe. Desalniettemin moet men voornamelijk de aandacht vestigen op het verhogen van de respons van de personen of groepen die men niet bereikt met de initieel gehanteerde *mode*. Het gebruiken van een mixed-mode survey moet steeds in functie zijn, niet noodzakelijk van het zoveel mogelijk bereiken, maar ook van het zo breed mogelijk bereiken van de doelpopulatie (Groves, 2006). Ten vijfde zorgt het gebruik van meerdere methoden ook voor een reductie van meetfouten. Terwijl men vaak overgaat tot sociaal wenselijke antwoorden wanneer men persoonlijke interviews aflegt, zorgen post- of online enquêtes ervoor dat er minder item non respons is op gevoelige vragen. Een specifiek voorbeeld van zulk een gevoelige vraag, is de vraag naar de inkomens van de respondent. Deze vraag zou in feite weerstand oproepen in elke *mode*, maar in face-to-face-bevragingen voelt men zich net iets meer geneigd tot het sociaal wenselijk beantwoorden van vragen. Toch zagen enquêteurs van de EU-SILC (Schockaert et al., 2012) dit dan weer als een niet onoverkomelijke beperking: in geval men gebruik maakt van inkomensschrijven, zou dit de inkomensvraag minder complex maken voor respondenten die het exacte bedrag moeilijk weten te achterhalen, maar zou het ook minder weerstand kunnen oproepen om zulk een gevoelige vraag te beantwoorden.

Een voorbeeld van een survey die een mixed-method design hanteerde, is de Survey Samenleven in Diversiteit 2017. Allereerst kregen de respondenten een uitnodiging via een brief met daarop een weblink om de vragenlijst in te vullen. Indien ze niet deelnamen, kregen ze twee weken daarna een herinneringsbrief toegestuurd. Wanneer de steekproefpersoon ook niet inging op deze uitnodiging, ging men over tot een face-to-face bevraging. Uiteindelijk werd na één maand van eerste contactopname gestart met het afnemen van persoonlijke interviews. Dit kon twee kanten uitgaan: een harde weigering (i.e. de pogingen worden stopgezet), een zachte weigering (i.e. uitstellen van de bevraging of de respondent geeft aan dit online op te nemen) of geen enkel contact met de respondent. In de laatste twee gevallen ging men nogmaals over tot contactopname, waarbij men ook een niet-thuiskaartje achterliet met nogmaals een uitnodiging tot deelname (met daarin de optie op een CAPI- of CAWI- bevraging). Daaropvolgend kon men, afhankelijk van de reactie, nog tot vijf contactpogingen ondernemen. Voornamelijk de beschikbaarheid van een online enquête bleek een positief effect te hebben op de respons (Stuyck et al., 2018).

Ook de EU-SILC- enquête maakt gebruik van verschillende methodieken om haar steekproefpersonen te bereiken. De twee belangrijkste hierbij zijn face-to-face en telefonische contactopname. De procedures stellen immers dat de men de respondent minstens éénmaal persoonlijk moet benaderen. Uit een

focusgroep met de ervaren enquêteurs bleek ook dat deze manier van benaderen door hen als de meest ideale wordt omschreven. Respondenten zouden sneller deelnemen in geval iemand zich fysiek presenteert aan hem of haar. Het is namelijk gemakkelijker voor de aangeschreven persoon om een schriftelijk vragenboekje te negeren, of 'nee' te zeggen wanneer men telefonisch wordt gecontacteerd, dan wanneer de persoon die deze vraag stelt voor hem of haar staat. Daarnaast verwerft men inzicht in de sociale omgeving van de respondent, waardoor men beter kan inspelen op specifieke situaties die zich kunnen voordoen wanneer men contact opneemt.

Een derde voorbeeld van een mixed mode design is deze van de Enquête naar de Arbeidskrachten (EAK), vertaald vanuit de internationale *Labour Force Survey (LFS)* voor specifiek België. Via een introductiebrief nodigt men de steekproefpersoon uit tot deelname aan de bevraging via een persoonlijk interview (face-to-face via CAPI). Terwijl de eerste wave plaatsvindt via een face-to-face gesprek via CAPI, krijgt de respondent de keuze om in de tweede, derde en vierde wave een online of telefonische bevraging te verkrijgen. De onderzoekers geven in de eerste wave wel de kans aan huishoudens met personen van 65 jaar of ouder om telefonisch of schriftelijk deel te nemen. In een rapport over de hervorming van deze enquête omschreef men de postale enquête als een onvoldoende adequaat alternatief: dit zou een lage respons opleveren, zijn duur en vragen veel verwerkingstijd (Statbel, 2019a; Termote & Depickere, 2018).

Specifiek met het oog op moeilijk-te-bereiken bevolkingssegmenten stelt wetenschappelijke literatuur dat het hanteren van een *mixed-mode survey design* effectief een positief effect kan hebben om hen te includeren. Haan, Ongena, and Aarts (2014) concentreerden zich op jongvolwassenen, immigranten met een niet-Westerse herkomst, inwoners van grootsteden en voltijds tewerkgestelden. Een eerste groep werd face-to-face gecontacteerd en kregen de mogelijkheid om de vragenlijst in te vullen via een persoonlijk interview met computer assistentie (oftewel CAPI) of om dit online te doen. Men nam telefonisch contact op met een tweede groep en bood hen een online- of telefonische enquête aan. Een derde groep werd eveneens getelefoneerd met de vraag of men wilde deelnemen aan de vragenlijst, en kregen at random een *mode* toegewezen. Dit onderzoek toont duidelijk aan dat het zeer uitdagend is om specifieke groepen te bereiken. De benaderingstechnieken moeten zorgvuldig worden ingevuld naargelang de doelpopulatie die men wenst te bereiken. Ander onderzoek beklemtoont dat de onderzoekers de keuze niet mogen overlaten aan de respondent, omdat dit het besluitvormingsproces voor respondenten complexer maakt, resulterend in een lagere responsgraad (Dillman, Smyth, & Christian, 2014). Wanneer men de keuze overlaat aan de respondent op welke manier hij de vragenlijst afneemt, zou dit de responsgraad niet noodzakelijk ten goede komen (Dillman et al., 2009; Stern et al., 2014). Fulton and Medway (2012) besluiten dat het aanbieden van de keuze tussen een online- of post enquête het responscijfer zelf significant deed dalen. Met andere woorden, het combineren van verschillende methoden kan ervoor zorgen dat men een moeilijk te bereiken groep effectief bereikt. Voor welke combinatie men opteert is afhankelijk van het onderzoeksopzet en de praktische haalbaarheid.

4 Merging of knowledge en de extra steekproef

In deze sectie koppelen we de voorgaande hoofdstukken aan elkaar door het te hebben over '*merging of knowledge*'. Dit concept ontleen we van het Frans-Belgisch project van de internationale armoede-organisatie ATD (zoals omschreven door Bennett and Roberts (2004a)). Het project genaamd '*Merging of knowledge*' startte in 1996 en bracht twee jaar lang academici, fulltime ATD- vrijwilligers en personen in armoede samen om kennis uit te wisselen (Bennett & Roberts, 2004b). De idee van het 'samenbrengen van de kennis' met verschillende actoren werd in het voorbereidingsproces van de gezinsenquête reeds

volbracht door in verschillende werkgroepen experts rond de tafel te zetten om allerlei domeinen van het gezinsleven te bespreken. De totstandkoming van de gezinsenquête gebeurde, met andere woorden, door een samenwerking binnen en buiten het beleidsdomein. In de voorbereidende fase werden experts geraadpleegd voor het opstellen van de enquêtevragen, verspreid over vier werkgroepen gaande van (1) Opvoeding, (2) Gezinsvorming en –ontbinding, (3) Gezin, zorg en werk en (4) Andere, voor specifieke vragen zoals genderidentiteit, berekening weegcoëfficiënten, en dergelijke. Enkele leden uit deze werkgroepen zetelen in de algemene, begeleidende stuurgroep die het gehele proces opvolgde.

Samenwerkingsverbanden opzetten om expertise te verwerven kan om de vragenlijsten inhoudelijk te verrijken enerzijds, maar men kan ook beroep doen op expertise van organisaties om de doelpopulatie efficiënter te kunnen bereiken (Bircan et al., 2017). De gezinsenquête werkte zo samen met het Agentschap Binnenlands Bestuur (ABB), en specifiek dan met het team Integratie en Inburgering, om een extra steekproef te trekken van gezinnen met een niet-EU-herkomst. Vanuit de bezorgdheid dat deze specifieke gezinnen vaak ondervertegenwoordigd zijn in én schriftelijke én Nederlandstalige enquêtes, financierde het ABB een bijkomende steekproef van gezinnen met minstens één ouderfiguur met een niet-Europese herkomst. Zoals eerder vermeld in Hoofdstuk 2.2 vonden we ook bewijs van studies die extra steekproeven op basis van andere relevante kenmerken trokken, zoals de studie van het Centrum voor Sociaal Beleid bij het project ‘Zorgen voor kinderen in Vlaanderen’. Dit onderzoek besliste, om een breed beeld te kunnen schetsen van de leefsituatie van kinderen, twee extra steekproeven te trekken: één steekproef van kinderen met een handicap uit de databank van het Vlaams Agentschap voor Personen met een Handicap (VAPH) en een tweede steekproef van kansarme gezinnen uit de databank van Kind en Gezin. Aanverwante studies zoals deze kunnen voor de gezinsenquête niet enkel methodologisch inspirerend zijn in functie van de steekproeftrekking, maar eveneens in functie van het steekproefkader. In Vlaanderen komen er verscheidene databanken (zoals deze van Kind en Gezin of van het Vlaams Agentschap voor Personen met een Handicap) voor die reeds een brede waaier aan kennis verzamelde over het gezinsleven, kinderen, of kwetsbare gezinnen, die de gegevens van de gezinsenquête kunnen verrijken.

Reiken naar andere databronnen, gaat verder dan enkel de steekproeftrekking, maar kan ook van toepassing zijn op dataverzamelingsmodi. Er zijn namelijk kwetsbare bevolkingssegmenten die onderzoekers wantrouwen, zeker wanneer deze geassocieerd zijn met de overheid. Deze barrières kunnen weggewerkt worden door bijvoorbeeld samen te werken met invloedrijke leden van de gemeenschap of door het betrekken van enkele meertalige leden uit verschillende gemeenschappen bij het opzetten van de enquête (Harkness, Stange, Cibelli, Mohler, & Pennell, 2014). Zoals eerder vermeld is het inzetten van interviewers die op vlak van cultuur of taal gelijkenissen vertonen met de kwetsbare gezinnen ook een veelgebruikte en succesvolle manier om gezinnen te bereiken. Deze interviewers kunnen sneller een vertrouwensband creëren waardoor ze gezinnen aanmoedigen tot deelname en hen geruststellen (bijvoorbeeld met betrekking tot wat er gebeurt met hun gegevens) (De Leeuw et al., 2008; Harkness et al., 2014).

Een belangrijke vraag is dus of het zinvol is om voor een extra steekproef dezelfde mode te gebruiken, en of de vragen wel dezelfde moeten zijn. Hierboven bespraken we reeds de bevindingen dat een postenquête niet noodzakelijk de beste manier is om kwetsbare gezinnen te betrekken in het onderzoek. Betreffende de extra steekproef van personen met een herkomst buiten de Europese Unie van de gezinsenquête 2016 besliste men om deze doelgroep te benaderen met een iets kortere vragenlijst. Maar de vragen waren wel steeds dezelfde. Het is echter net de meerwaarde van een extra steekproef dat er ook een focus kan worden gelicht op thema's die inhoudelijk relevant zijn voor de doelgroep van de extra steekproef. In hoofdstuk 4 en 5 analyseren we in welke mate de basis en extra steekproef representatief zijn voor de Vlaamse gezinnen met kinderen en in hoeverre een extra steekproef bij gezinnen met een niet-EU

achtergrond bijdraagt aan een voldoende bereik van kwetsbare gezinnen. Als men in een volgende golf van de gezinsenquête specifiek wil inzoomen op de levensomstandigheden van kwetsbare gezinnen met kinderen, dan zou men bijvoorbeeld een extra steekproef kunnen trekken op basis van de databank van Kind en Gezin rond kansarmoede. De kinderen die geboren worden in kansarmoede (volgens de zogenaamde 'Kansarmoede-index') zijn immers ook beleidsmatig relevant. Men zou specifieke beleidsthema's rond de zes dimensies van kansarmoede, die worden gebruikt door Kind en Gezin om het risico op kansarmoede vast te stellen, kunnen bevragen. Voor die groep van gezinnen zou men dan een postenquête kunnen combineren met een face-to-face bevraging om de respons te vergroten. De specifieke samenwerking met Kind & Gezin biedt een belangrijk voordeel: bijna 97% van de Vlaamse gezinnen krijgt minstens één huisbezoek van K&G in de eerste drie levensmaanden. Zij hebben een quasi-universeel bereik van gezinnen met een pasgeboren kind dat leeft in kwetsbare omstandigheden. Een recent onderzoek naar de levensomstandigheden van gezinnen in kansarmoede in de stad Kortrijk op basis van de kansarmoede-index van Kind en Gezin toont ook aan dat de kansarmoede-index een valide instrument is om de kwetsbaarheid te operationaliseren (Van Lancker et al., 2020).

Samenvattend

Er zijn verschillende methoden mogelijk om gegevens te verzamelen bij gezinnen met kinderen, van mondelinge en schriftelijke methoden tot en met online enquêtes. Een brede waaier aan wetenschappelijke literatuur somt de voor- en nadelen op van het hanteren van deze verschillende methoden, waarbij wordt stil gestaan bij de praktische haalbaarheid, het kostenplaatje, de timing, de kwaliteit. Met een postenquête, de methode gehanteerd om data te verzamelen voor de gezinsenquête 2016, bestaat het risico dat de non respons verhoogt bij kwetsbare gezinnen. We staan ook stil bij het design van een vragenlijst, en bij de mogelijkheid om met roterende modules te werken.

Voor vervolgenquêtes van de gezinsenquête is het belangrijk om de mode van dataverzameling te bepalen aan de hand van de inhoudelijke thema's waarin men is geïnteresseerd. Een extra steekproef van kwetsbare gezinnen zou immers gebruik kunnen maken van een mixed mode enquête waarbij een postenquête gecombineerd wordt met een face-to-face bevraging. Zo'n extra steekproef zou getrokken kunnen worden uit de databanken van Kind en Gezin om een heel specifieke groep van kwetsbare kinderen voldoende te bereiken. Voor die groep zouden dan vragen gesteld kunnen worden gelinkt aan het beleid dat wordt gevoerd rond kansarmoede. Er is echter geen kant-en-klare oplossing om een specifieke doelpopulatie te bereiken via een combinatie van methodes.

Hoofdstuk 4

Respons en representativiteit

In dit hoofdstuk presenteren we empirische analyses naar de representativiteit van de gezinsenquête op basis van een koppeling met administratieve data uit het Datawarehouse. We gaan na of de gezinsenquête voldoende representatief is voor gezinnen met kinderen in Vlaanderen, en in hoeverre de extra steekproef het bereik van kwetsbare gezinnen heeft vergroot. De datakoppeling laat ons ook toe om een aantal van de scenario's en opties die we bespraken in de voorgaande hoofdstukken aan de realiteit te toetsen.

1 Responsgraad in de Gezinsenquête

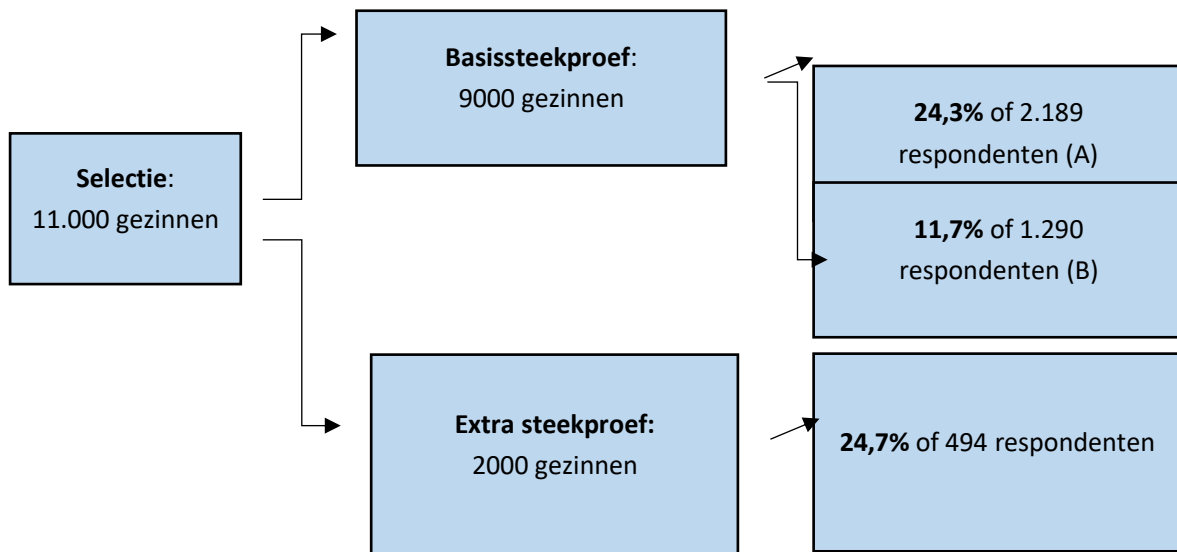
De gezinsenquête tracht via een postenquête voldoende gegevens te verzamelen over de geselecteerde steekproefeenheden. De steekproefpersonen kunnen echter weigeren deel te nemen, tijdens de deelname afhaken, of deelnemen maar weigeren enkele vragen te beantwoorden. We spreken van *unit non respons* wanneer een geselecteerde steekproefpersoon weigert om deel te nemen aan de vragenlijst; *item non respons* treedt op wanneer een aangeschreven steekproefpersoon wel deelneemt, maar weigert en/of er niet in slaagt alle vragen te beantwoorden. Ten slotte slaat *non-contact* op de personen die men daadwerkelijk niet heeft kunnen bereiken (Schockaert et al., 2012).

Er zijn verschillende redenen waarom unit non respons kan optreden (Lynn, 2008): (1) men slaagt er niet in om de respondent te identificeren, te lokaliseren of te bereiken; (2) men slaagt er niet in om contact op te nemen met de respondent; (3) de respondent is niet geschikt om te deel te nemen (bijvoorbeeld: door aanwezigheid, gezondheidsproblemen, overlijden, ...); (4) de respondent weigert om deel te nemen (bijvoorbeeld door een gebrek aan interesse of tijd); (5) de respondent ervaart moeilijkheden om te communiceren (bijvoorbeeld: doordat hij of zij de taal niet (voldoende) machtig is) en (6) de onderzoekers slagen er niet in om de data correct te verwerken (bijvoorbeeld: de vragenlijst geraakt zoek, fout in verwerking, en dergelijke)(Lynn, 2008).

Om preventief non respons tegen te gaan, werkte men voor de gezinsenquête een verzendschema uit volgens de principes van de *Total Design Method* van Dillman et al. (2014). Het verzenden van de vragenboekjes ging gepaard met een introductiebrief, een antwoordkaartje (waarop men kon aangeven niet te willen deelnemen) en een retourenvelop. Een week na de eerste verzending werd een herinnerings- en bedankkaartje uitgestuurd naar alle aangeschreven personen. Het veldwerk werd wekelijks opgevolgd. Acht zendingen bleken onbestelbaar, 8% of 723 personen gaven aan via het antwoordkaartje om niet te willen deelnemen. De voornaamste redenen die deze personen opgaven, waren: 'geen interesse' (23%); 'ik doe nooit mee met onderzoek' (23%) en 'te weinig tijd' (22%). Er kwamen 131 blanco vragenlijsten terug en er volgden ook 201 contactopnames (mails of telefoons) met vragen over het onderzoek, waarvan ruim een kwart zich afvroeg of men de vragenlijst in het Frans kon verkrijgen (Audenaert & Stuyck, 2018a).

De initieel opgestelde quota voor de respons bedroeg 45%, die nadien, op advies van het aangestelde veldwerkbureau, werd bijgesteld tot 25 à 33%. De finale participatie- of responsgraad van een enquête wordt berekend door het aantal deelnemende respondenten ten opzichte van het aantal geselecteerde personen te plaatsen. Van de 11.000 aangeschreven steekproefpersonen, bereikte de enquête 2.683

personen of **24,4%**. Met andere woorden, minder dan één op vier van alle geselecteerde personen nam deel aan de gezinsenquête. Opgesplitst zien we dat 24,3% van de basissteekproef en 24,7% van de extra steekproef werden bereikt. De basissteekproef kan opgesplitst worden in de respondenten die vragenboekjes A invulden en diens partners die vragenboekje B invulden (zie Figuur 2).



Figuur 2 Schematisch overzicht respons

Hoe doet deze responsgraad van 24,4% het in vergelijking met andere enquêtes? In de Gemeentemonitor 2017 had men een respons van 39%, in de Stadsmonitor 2017 was dit 36% (Agentschap Binnenlands Bestuur, 2018a)¹⁵. Een andere schriftelijke enquête is deze in het kader van Duurzame mantelzorg in Vlaanderen, waar de responsgraad neerkomt op 53,7% (Bronselae, Vandezande, Vanden Boer, & Boons, 2015). Andere voorbeelden van de responsgraad (maar dan voor face-to-face bevragingen) zijn deze van de MEqIN survey, de GGS en de EAK. In de MEqIN- survey die plaatsvond in opdracht van Belspo bedraagt de responsgraad voor alle aangeschreven huishoudens 37,9% (Maniquet et al., 2018), de Generations and Gender Survey (GGS) verkreeg een responsgraad van 43,1% (Lauwereys, Neels, & De Winter, 2011), en voor de Enquête naar Arbeidskrachten ligt de respons gemiddeld zelfs op 70% (Termote & Depickere, 2018). Volgens sociaalwetenschappelijke literatuur ligt een 'acceptabel' responscijfer bij postenquêtes echter rond de 50% (Babbie, 1973; Stedman, Connelly, Heberlein, Decker, & Allred, 2018). In vergelijking met de genoemde enquêtes en de literatuur kunnen we bijgevolg concluderen dat de respons van de gezinsenquête relatief laag ligt.

Maar wat is het effect van een lage responsgraad? Welke impact heeft deze op de resultaten? De responsgraad wordt geïnterpreteerd als een indicator van mogelijke *bias* of vertekening in de verkregen data. Men slaagt er namelijk (zo goed als) nooit in om data te verzamelen van alle getrokken steekprofeenheden, waardoor de uiteindelijke steekproef (netto steekproef) verschilt van de initieel getrokken steekproef (bruto steekproef). Volgens de principes van statistische inferentie streven we ernaar om uitspraken te maken over de doelpopulatie (in ons geval: de gezinnen in Vlaanderen en het Brusselse Gewest) door gebruik te maken van een steekproef waarvan we de selectiekans van de eenheden kennen. Non respons verstoort deze selectiekans, en kan ervoor zorgen dat bepaalde individuen of groepen niet vertegenwoordigd zijn in de steekproef waardoor men geen uitspraken kan doen over de gehele populatie zonder dit in rekening te brengen (Lynn, 2008). Met andere woorden, er

¹⁵ De gemeente- en stadsmonitor geven de keuze aan de respondent om de vragenlijst schriftelijk in te vullen en terug te zenden of het online in te vullen (Agentschap Binnenlands Bestuur, 2018b).

treedt vertekening op wanneer de groep van de deelnemende respondenten verschilt met de groep van niet-deelnemende respondenten. In deze sectie streven we naar het verkrijgen van inzicht in de profielen van de respondenten en niet-respondenten en in welke mate deze verschillen. Een ander belangrijk inzicht die men kan verwerven is het achterhalen van de redenen waarom een persoon niet deelneemt aan de bevraging. *“Met andere woorden, als we de persoonlijke redenen kennen van respondenten in specifieke socio-economische omstandigheden wordt het mogelijk te zoeken naar algemeen toepasbare oplossingen voor het remediëren van non respons (Schockaert et al., 2012, p. 39)”*. We hebben echter geen kennis van *wie* een reden opgaf om niet deel te nemen. We concentreren ons dan ook op de verscheidene achtergrondkenmerken van de gehele groep respondenten versus de groep niet-respondenten.

Veelal geldt het principe ‘Hoe meer observaties, hoe betrouwbaarder de gegevens zijn’. Let wel, een lage responsgraad is niet *an sich* problematisch (Groves & Peytcheva, 2006). Het is echter een probleem wanneer de niet-respondenten andere achtergrondkenmerken bezitten dan de respondenten. Of deze non respons selectief is, gaan we dus na door de achtergrondgegevens van de (niet-)respondenten in kaart te brengen (Schockaert et al., 2012; Schouten, 2004). We hebben hiervoor een zicht nodig op de gegevens alle geselecteerde personen. Via een aanvraag bij het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB) verkregen we enerzijds geanonimiseerde individuele gegevens over de geselecteerde steekproefpersonen (respondenten en niet-respondenten) en geaggregeerde gegevens van de doelpopulatie. In wat volgt, beschrijven we deze verkregen administratieve gegevens en staan we stil bij de responsgraad, de representativiteit van de gerealiseerde steekproef ten aanzien van populatiegegevens, de representativiteit van de gerealiseerde steekproef ten aanzien van de geselecteerde steekproef, de meerwaarde van de extra steekproef en de item non respons.

2 Representativiteit van de gerealiseerde steekproef

In deze sectie gaan we na wat de achtergrondkenmerken van de personen zijn die werden geselecteerd voor de enquête (bruto steekproef) en plaatsen deze tegenover de deelgenomen respondenten (netto steekproef). Het is mogelijk om deze vergelijking te maken doordat we administratieve gegevens verkregen vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB). Door de gegevens te koppelen, krijgen we een zicht op de geselecteerde referentiepersonen van de basissteekproef (n= 9.000), de geselecteerde personen van de extra steekproef (n= 2.000) en de personen die werkelijk hebben deelgenomen (respectievelijk n= 2.189 en n= 493) (zie Tabel 7).

Tabel 7 Overzicht van de verkregen gegevens van respondenten en niet-respondenten voor de basis en extra steekproef (Bron: DWH AM&SB)

	Basissteekproef	Extra steekproef	Totaal
Geselecteerde steekproef (of selectie)	8.984	1.986	10.970
Individuele gegevens over de gezinsleden van de selectie	33.985	8.288	42.273
Gerealiseerde steekproef (of “respondenten”)	2.189	493	2.682
Niet-deelnemende steekproefpersonen (of “niet-respondenten”)	6.795	1.493	8.270
Respons	24,3%	24,6%	24,4%
Non respons	75,7%	75,4%	75,6%

Van de 11.000 geselecteerde steekproefpersonen (basis- en extra steekproef), namen 2.683 personen deel (vragenboekje A). Dat komt neer op een respons van 24%. Specifiek voor dit project verkregen we gegevens vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming over deze aangeschreven personen. Het ging om individuele- en gezinsgegevens, wat betekent dat we naast de 9.000 individueel aangeschreven steekproefpersonen ook informatie verkregen over diens huishouden. Met andere woorden, de brutosteekproef van 9.000 geselecteerde referentiepersonen werd nu uitgebreid met 24.985 gezinsleden die niet expliciet werden aangeschreven. In totaal gaat het om een verzameling van 42.273 gezinsleden waarvan 8984 huishoudens (16 *missings vanuit de KSZ*) voor de basissteekproef en 1.986 huishoudens (14 *missings vanuit de KSZ*) voor de extra steekproef. Daaropvolgend verkregen we uiteindelijk informatie over 2.189 respondenten van de basissteekproef en 493 respondenten van de extra steekproef. Dit leidt tot een totaal van 2.682, wat verschilt met het totaal van de gezinsenquête gegevens (n= 2.683).

Waarom is het belangrijk om inzicht te verwerven in de achtergrondgegevens van de personen die niet hebben deelgenomen aan de gezinsenquête? Deze verdieping is interessant om te kijken of de respons *selectief* plaatsvond. Het uitspitten van deze vergelijking tussen de respondenten en niet-respondenten vinden we eveneens terug in ander enquête-onderzoek. De Keulenaer (2002) analyseerde de socio-economische achtergrondkenmerken van personen die niet deelnamen aan de tweede golf van de Panelstudie van Belgische Huishoudens (PSBH). Hieruit bleek dat bepaalde groepen een grotere kans kenden om niet te worden bereikt en dus ondervertegenwoordigd waren in de studie. Het ging over gezinnen met niet-Belgische gezinshoofden, jongeren, alleenstaanden, gezinnen in het Brussels Hoofdstedelijk Gewest en gezinnen die verhuisden tussen de twee golven. Daarnaast bleken in de PSBH zowel huishoudens met een laag beschikbaar inkomen, als personen die het moeilijk hadden om rond te komen (subjectieve armoede-indicator) ondervertegenwoordigd. Ten slotte kwamen de onderzoekers tot de conclusie dat men huishoudens die huren of wonen in een huisvesting met gebrek aan comfort niet kon bereiken (De Keulenaer & Levecque, 2004; Schockaert et al., 2012).

Andere studies die de relatie tussen respons en achtergrondkenmerken bekijken, geven echter vaak gemengde bevindingen weer. De hypothese dat, bij wijze van voorbeeld, etnisch-culturele minderheden de facto moeilijker te bereiken zijn is niet eenduidig terug te vinden. Zo beargumenteren Schmeets and Van der Bie (2006) dat de relatie tussen de respons en herkomst berust op een schijnverband: herkomst hangt namelijk samen met een zwakke arbeidspositie en het wonen in een stad, en het is eerder dit verband dat een effect blijkt te hebben op de respons. Dit impliceert dat respons survey per survey moet bekeken worden (Schockaert et al., 2012). Ook de relatie tussen de respons en financiële armoede zou gebaseerd zijn op een schijnverband. Schockaert et al. (2012) onderzochten de uitval of attritie van een 'verborgen groep armen' in de EU-SILC. In hun rapport stellen ze dat financiële armoede *an sich* geen significante invloed blijkt te hebben op de respons. Het bleken voornamelijk andere indicatoren van sociale uitsluiting te zijn die een effect hebben op non respons. Het gaat om (1) de lage scholingsgraad (van de ouders), (2) het ontbreken van een arbeidsinkomen in het huishouden en (3) het huren van een woning. Doordat een lage scholingsgraad en het huren van een woning gecorreleerd zijn met armoede, stelt men wel degelijk dat er een selectieve uitval plaatsvindt van arme huishoudens uit de EU-SILC (Schockaert et al., 2012).

Diezelfde oefening passen we nu toe op de gezinsenquête. In deze sectie beschrijven we enkele achtergrondkenmerken die we verkregen vanuit het DWH AM&SB. Op die manier kunnen we nagaan of de gerealiseerde steekproef de geselecteerde steekproef goed weergeeft en dus in welke mate de

gezinsenquête een correcte representatie van de doelpopulatie afbeeldt. Via beschrijvende statistieken op basis van kruistabellen omschrijven we de voornaamste bevindingen.

2.1 Representativiteit ten aanzien van de populatie

In Tabel 8 geven we een overzicht van de gerealiseerde *basissteekproef* [1] en de gerealiseerde *extra steekproef* [2]; de geselecteerde *basissteekproef* [3] en de geselecteerde *extra steekproef* [4]; de *totale gerealiseerde steekproef* [5] en de populatiegegevens [6]. Hierbij tonen we twee variabelen waarvoor we de verdeling van de populatie verkregen vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming, namelijk gezinstype en herkomst.

Vooreerst bekijken we de verdeling van de herkomstvariabelen in de geselecteerde en gerealiseerde steekproef. De populatiegegevens [kolom 6] rapporteren de verdeling van de doelpopulatie naar herkomst. Van alle Vlaamse gezinnen met minstens één kind jonger dan 25 jaar blijken er 77,3% een Belgische herkomst te hebben, 9,8% een EU-herkomst en 12,7% een niet-EU-herkomst. Van degene die men uiteindelijk heeft geselecteerd voor de basissteekproef [3] is slechts 71,7% van Belgische herkomst, 10,0% van EU en 13,4% van niet-EU-herkomst. Het aandeel van personen waarvan de herkomst niet bekend is, is relatief groot in de basisselectie (4,9%) en de extra selectie (1,9%). Het gevolg is dat de geselecteerde basissteekproef [3] significant verschilt van de populatiegegevens [6]. Dit betekent dat er in de selectie een vertekening zit, met daarbij een beperkte oververtegenwoordiging van personen met een niet-EU-herkomst en een sterke ondervertegenwoordiging van personen met een Belgische herkomst. Daaropvolgend kijkt ook de herkomstverdeling van de *gerealiseerde* basissteekproef [1] af met de populatieverdeling [6]: 72,8% van de respondenten is van Belgische origine, 9,4% een EU- en 12,9% hebben een niet-EU-herkomst. Dit is niet geheel onverwacht: wanneer de selectie geen representatief beeld weergeeft van de populatie, is er ook een grote kans dat de uiteindelijk bereikte steekproef dit niet doet.

Dit moet echter genuanceerd worden. Het grote aandeel van mensen met een herkomst die 'niet beschikbaar' is in de administratieve data van het Datawarehouse vertekent het beeld. De herkomst werd door het Datawarehouse Arbeidsmarkt en Sociale Bescherming zo gecodeerd dat deze voldoet aan de herkomstvariabele zoals omschreven in het Vlaamse Integratie en Inburgeringsbeleid¹⁶. Personen met een vreemde herkomst worden gedefinieerd als personen met een *huidige* vreemde nationaliteit, personen met een vreemde *eerste* maar huidig Belgische nationaliteit, personen met een huidige en eerste Belgische nationaliteit maar met een moeder en/of vader met een vreemde eerste nationaliteit. Wanneer bijvoorbeeld de eerste nationaliteit van de moeder of vader onbekend is, worden deze gecodeerd als zijnde 'onvolledige informatie'. Op basis van de datakoppeling tussen de gezinsenquête en de gegevens van het Datawarehouse kunnen we voor de gerealiseerde steekproef echter nagaan wat de zelfgerapporteerde herkomst is van de groep respondenten met een herkomst die 'niet beschikbaar' is. 82% van deze groep geeft aan Belg te zijn, 11% geeft aan een EU-herkomst te hebben, en 7% een niet-EU-herkomst. Deze resultaten suggereren dat het gros van de personen met ontbrekende herkomstgegevens eigenlijk de Belgische nationaliteit bezit, waardoor de verdeling van de herkomst in de steekproeven dichter aansluiten bij de verdeling van herkomst in de populatiegegevens.

Aansluitend vinden we geen significant verschil tussen de *gerealiseerde* basissteekproef [1] en de *geselecteerde* basissteekproef [3]: dit betekent dat er genoeg personen van de drie herkomstgroepen

¹⁶ Zoals omschreven in Artikel 3 van het Decreet van 28 april 1998

uiteindelijk werden bevestigd zoals men het voor ogen had in de selectie. Hierdoor heeft de gezinsenquête uiteindelijk ook voldoende personen van niet-EU-herkomst kunnen bevragen. Dit wordt bevestigd door het significant verschil tussen de totaal gerealiseerde steekproef (basis én extra) [5] en de basisselectie [3]: doordat een extra steekproef van personen met een niet-EU-herkomst werd getrokken, krijgt de gezinsenquête –zoals verwacht– een oververtegenwoordiging van personen met een vreemde herkomst. De extra steekproef, zoals vermeld in Hoofdstuk 2.2, kwam tot stand via een samenwerking met het Agentschap Binnenlands Bestuur (ABB) vanuit de bezorgdheid dat personen afkomstig van buiten de Europese Unie vaak ondervertegenwoordigd zijn in enquêtes. Maar, en dat is belangrijk, uiteindelijk geven de basisselectie en basisrespondenten een voldoende representatief beeld van deze bevolkingsgroep, waardoor de extra steekproef op basis van deze herkomstcriteria eigenlijk onnodig was. Kortom, het toevoegen van de extra steekproef trok de verdeling van de totale gerealiseerde steekproef [5] scheef qua herkomst: 27,9% van de gerealiseerde steekproef heeft een niet-EU-herkomst terwijl deze groep in de populatie 12,7% uitmaakt. Om een representatief beeld van de Vlaamse gezinnen met kinderen naar herkomst te verkrijgen was de extra steekproef niet nodig. Als de gegevens geanalyseerd worden, moet er dus rekening gehouden worden met deze vertekening.

We toetsen ook de verdeling van de gezinstypes met de verkregen populatiegegevens. Van alle Vlaamse gezinnen met minstens één kind jonger dan 25 jaar zijn 81,2% geclassificeerd als gehuwd of ongehuwd en 18,8% zijn alleenstaand. Hier zien we geen significant verschil tussen de gerealiseerde [1] en geselecteerde [3] basissteekproef, in vergelijking met de populatieverdeling [6]. Dit betekent dat er voldoende (on)gehuiden en voldoende alleenstaanden werden bereikt in de steekproeven om representatief te zijn voor de populatie van Vlaamse gezinnen met kinderen (jonger dan 25 jaar). De geselecteerde basissteekproef is bijgevolg representatief voor de doelpopulatie.

Tabel 8 Overzicht gerealiseerde en geselecteerde (basis en extra) steekproef ten aanzien van populatiegegevens (Bron: Ongewogen cijfers vanuit het DWH AM&SB)

	Gerealiseerde steekproef Basis [1]	Gerealiseerde steekproef Extra [2]	Geselecteerde steekproef Basis [3]	Geselecteerde steekproef Extra [4]	Gerealiseerde steekproef Totaal [5]	Populatie gegevens [6]
Herkomst						
BE	72,8%	0,0%	71,7%	0,2%	59,6%	77,3%
EU	9,4%	1,9%	10,0%	1,9%	8,0%	9,8%
Niet-EU ¹⁷	12,9%	95,6%	13,4%	96,0%	27,9%	12,7%
Overige/ Niet beschikbaar	4,9%	2,5%	4,9%	1,9%	4,5%	0,2%
<i>n</i>	2.186 ^{1,3}	483	8.986 ^{2,4}	1.976	2.669	904.954
Gezinstype						
(on)gehuwd	82,0%	83,4%	81,6%	83,8%	82,3%	81,2%
alleenstaand	18,0%	16,6%	18,4%	16,2%	17,7%	18,8%
<i>n</i>	2.175	477	8.941	1.942	2.652	904.954

¹Significant verschil (Pearson $\chi^2= 23000$, $p<0,001$) tussen [1] en [6]; ² Significant verschil (Pearson $\chi^2= 8000$, $p<0,001$) tussen [3] en [6];

³Significant verschil (Pearson $\chi^2= 169,35$, $P<0,001$) tussen [1] en [5]; ⁴Significant verschil (Pearson $\chi^2= 324,17$, $p<0,001$) tussen [3] en [5].

¹⁷ Voor de afbakening van personen met niet-EU-herkomst: zie bijlage 3.

2.2 Representativiteit ten aanzien van de geselecteerde steekproef

In Tabel 9, 10 en 11 tonen we of en in welke mate de gerealiseerde steekproef afwijkt van de geselecteerde steekproef op een aantal sociale, demografische en economische kenmerken. We geven een overzicht van de gerealiseerde *basissteekproef* [1] en de gerealiseerde *extra steekproef* [2]; de geselecteerde *basissteekproef* [3] en de geselecteerde *extra steekproef* [4]; alsook van de totale gerealiseerde steekproef [5]. Het gaat om ongewogen cijfers verkregen vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming. De vergelijking met de werkelijke populatie kunnen we hier niet meer maken; we vergelijken telkens met de geselecteerde basissteekproef die, zoals hierboven besproken, een goede weergave is van de werkelijke doelpopulatie. Waar de verschillen significant zijn, bespreken we ook de verdeling van niet-respondenten over kenmerken: de verdeling van het aandeel van de aangeschreven selectie dat de enquête niet heeft ingevuld.

Allereerst bekijken we enkele **sociale en demografische gegevens**: het geslacht, de leeftijd, het gezinstype, de gezinsgrootte, het aantal kinderen, de provincie en de herkomst. Algemeen stellen we op het eerste zicht vast dat de gerealiseerde *basissteekproef* [1] een degelijke afspiegeling is van de geselecteerde *basissteekproef* [3]. Hoewel de gezinsenquête kampt met een lage responsgraad, is de verdeling van de *basis*respondenten qua sociodemografische kenmerken gelijkaardig aan de verdeling van de *basis*selectie. We vinden enkel een significant verschil tussen de gerealiseerde en geselecteerde basissteekproef voor herkomst (oververtegenwoordiging van personen met een Latijns-Amerikaanse herkomst). Een tweede vaststelling is dat de extra gerealiseerde steekproef *geen* goede afspiegeling is van de extra geselecteerde steekproef. De 494 deelnemende respondenten met niet-EU-herkomst zijn, met andere woorden, geen representatief staal van de gezinnen met niet-EU-herkomst in Vlaanderen en/of het Brussels Hoofdstedelijk Gewest. Tussen beide steekproeven vinden we significante verschillen bij het gezinstype en de herkomst. Een derde, daaropvolgende, vaststelling is de mate waarin de extra steekproef van personen met een niet-EU-herkomst de totaal gerealiseerde steekproef vertekent. In vergelijking met de basissteekproef wordt deze extra steekproef gekenmerkt door vrouwelijke referentiepersonen met een hoger aantal kinderen. De extra steekproef beschikt over een hoger aandeel gehuwden, en een hoger aandeel van gezinnen in het Brussels Hoofdstedelijk Gewest. Daaropvolgend vinden we meerdere significante verschillen tussen de totale gerealiseerde steekproef [5] en de basisselectie [3] als we het gezinstype, de gezinsgrootte, de provincie en de herkomst onder de loep nemen. Door het toevoegen van de extra steekproef zijn gehuwde ouders (63,2%) oververtegenwoordigd, in vergelijking met de basisselectie (60,4%). Gezinnen in het Limburgse zijn licht ondervertegenwoordigd, gezinnen met twee leden licht oververtegenwoordigd. Zoals verwacht zijn er eveneens significante verschillen in herkomst tussen de totale gerealiseerde steekproef en de basisselectie: referentiepersonen met een Belgische herkomst zijn sterk ondervertegenwoordigd (59,6%) vergeleken met de basisselectie (71,7%), personen met een Europese herkomst licht ondervertegenwoordigd (8,0% versus 10,0%). Het aandeel personen met een niet-Europese herkomst is verdubbeld in de gerealiseerde steekproef (27,9%), vergeleken met de basisselectie (13,4%), en dus sterk oververtegenwoordigd. Dat bevestigt onze eerdere vergelijking met de populatiegegevens.

Tabel 9 Overzicht gerealiseerde en geselecteerde (basis en extra) steekproef tegenover de totale gerealiseerde steekproef voor enkele sociodemografische kenmerken (Bron: Ongewogen cijfers uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming)

Sociale en demografische kenmerken

	Gerealiseerde steekproef Basis [1]	Gerealiseerde steekproef Extra [2]	Gerealiseerde steekproef Totaal [5]	Geselecteerde steekproef Basis [3]	Geselecteerde steekproef Extra [4]
Geslacht					
Man	45,0%	37,2%	43,6%	44,7%	41,0%
Vrouw	55,0%	62,8%	56,4%	55,3%	59,0%
<i>n</i>	2.186	486	2.672	8.987	1.981
Leeftijd					
<30	6,0%	10,1%	6,7%	7,2%	13,0%
30-39	30,9%	43,8%	33,2%	31,5%	42,4%
40-49	37,0%	34,2%	36,4%	37,7%	31,3%
>=50	26,2%	11,9%	23,6%	23,7%	13,3%
<i>n</i>	2.186	486	2.672	8.987	1.981 ¹
Gezinstype					
Gehuwd	61,9%	68,9%	63,2%	60,4%	68,6%
Ongehuwd	19,7%	13,0%	18,5%	20,8%	13,6%
Alleenstaand	17,9%	16,3%	17,6%	18,3%	15,9%
Andere	0,5%	1,8%	0,8%	0,5%	2,0%
<i>n</i>	2.186	486 ²	2.673 ³	8.987	1.981
Gezinsgrootte					
1	0,1%	0,6%	0,2%	0,2%	0,9%
2	8,9%	10,0%	9,1%	9,0%	8,4%
3	33,3%	30,1%	32,7%	32,7%	24,3%
4	37,2%	31,1%	36,1%	38,1%	31,0%

	5	15,2%	14,5%	15,0%	14,1%	18,6%
	6+	5,4%	13,7%	6,9%	5,9%	16,6%
	<i>n</i>	2.185	482	2.667	8.984	1.968
Aantal kinderen						
	0	1,0%	5,7%	1,9%	1,0%	5,2%
	1	37,4%	32,0%	36,4%	36,8%	27,2%
	2	41,3%	36,1%	40,3%	42,4%	34,8%
	3	15,5%	14,6%	15,3%	14,8%	19,3%
	4+	4,8%	11,6%	6,0%	5,0%	13,6%
	<i>n</i>	2.189	493	2.682	9.000	2.000
Provincie						
	Antwerpen	28,4%	35,6%	29,7%	27,4%	35,1%
	Brussel	0,8%	3,9%	1,4%	0,9%	3,7%
	Limburg	12,8%	10,5%	12,4%	13,9%	12,2%
	Oost-Vlaanderen	23,3%	23,0%	23,2%	23,3%	20,3%
	Vlaams-Brabant	17,2%	18,7%	17,5%	17,8%	19,8%
	West-Vlaanderen	17,4%	7,4%	15,6%	16,6%	8,1%
	Andere/onbekend	0,1%	0,8%	0,2%	0,0%	0,7%
	<i>n</i>	2.186	486	2.672	8.987⁴	1.981⁵
Herkomst (alle clusters)						
	BE	72,8%	0,0%	59,6%	71,7%	0,2%
	Buurlanden	4,2%	0,0%	3,5%	4,8%	0,0%
	West-EU	0,7%	0,0%	0,6%	0,7%	0,0%
	Zuid-EU	2,6%	0,0%	2,2%	2,4%	0,1%
	EU-12	1,8%	1,9%	1,8%	2,0%	1,7%
	Centraal- en Oost- EU	2,0%	12,0%	3,8%	1,8%	12,9%
	Turkije	2,5%	12,6%	4,4%	3,0%	17,9%
	Noord-Afrika	3,8%	22,0%	7,1%	3,9%	27,6%

Centraal- en Zuid- Afrika	1,8%	14,5%	4,1%	2,0%	12,8%
Noord- Amerika	0,1%	2,3%	0,5%	0,2%	1,2%
Latijns-Amerika	0,6%	9,1%	2,1%	0,4%	5,8%
Azië	2,0%	23,2%	5,8%	2,2%	17,9%
Overige of onvolledig	4,9%	2,5%	4,5%	4,9%	1,9%
n	2.186 ⁶	483 ⁸	2.669 ⁷	8.986	1.976

¹ Significant verschil (Pearson $\chi^2=146,4$, $p=0,000$) tussen totaal gerealiseerde steekproef [5] en extra geselecteerde steekproef [4];

² Significant verschil (Pearson $\chi^2=19,98$, $p=0,000$) tussen extra gerealiseerde steekproef [2] en extra geselecteerde steekproef [4];

³ Significant verschil (Pearson $\chi^2=11,6$, $p=0,009$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3];

⁴ Significant verschil (Pearson $\chi^2=25,99$, $p=0,000$) tussen de totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3];

⁵ Significant verschil (Pearson $\chi^2=102,85$, $p=0,000$) tussen de totaal gerealiseerde steekproef [5] en de extra geselecteerde steekproef [4];

⁶ Significant verschil (Pearson $\chi^2= 171,48$, $p=0,000$) tussen basis gerealiseerde steekproef [1] en totaal gerealiseerde steekproef [5];

⁷ Significant verschil (Pearson $\chi^2=551,52$, $p=0,000$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3];

⁸ Significant verschil (Pearson $\chi^2=30,40$, $p=0,002$) tussen extra gerealiseerde steekproef [2] en extra geselecteerde steekproef [4].

Waar we vervolgens nog aandacht aan willen schenken, is de verdeling van het **opleidingsniveau** binnen de beide steekproeven. We voegen dit kenmerk niet toe aan bovenstaande tabel omdat er slechts beperkte administratieve gegevens beschikbaar zijn over het scholingsniveau van de geselecteerde steekproefpersonen. In Tabel 10 vergelijken we dus enkel de basis, extra en totale gerealiseerde steekproef, gebaseerd op de enquêtegegevens.

Tabel 10 Overzicht opleidingsniveau gerealiseerde steekproeven (Basis - Extra - Totaal) op basis van enquêtegegevens (Bron: ongewogen cijfers uit de gezinsenquête 2016)

	Gerealiseerde steekproef BASIS [1]	Gerealiseerde steekproef EXTRA [2]	Gerealiseerde steekproef TOTAAL [3]
Laag	13,5%	29,3% ²	16,5%
Midden	29,7%	31,6%	30,0%
Hoog	56,8%	39,1%	53,5%
<i>n</i>	2.096 ^{1,3}	481 ^{2,3}	2.577

¹ Significant verschil tussen [1] en [3]: Pearson $\chi^2=206,83$, $p=0,000$

² Significant verschil tussen [2] en [3]: Pearson $\chi^2=53,08$, $p=0,000$

³ Significant verschil tussen [1] en [2]: Pearson $\chi^2=82,96$, $p=0,000$

Meer dan de helft van de *basis*respondenten geeft aan hooggeschoold te zijn (56,8%). 13,5% antwoordden laaggeschoold te zijn. De verdeling van de geschooldheid bij de *extra* respondenten wijkt hier sterk van af: 39,1% zeggen hoogopgeleid te zijn en 29,3% zijn laagopgeleid. Opnieuw zien we dat, door de integratie van de extra steekproef, het aandeel laagopgeleiden in de totaal gerealiseerde steekproef wordt opgetrokken, het aandeel van de hoogopgeleiden wordt kleiner.

Tabel 11 Overzicht gerealiseerde en geselecteerde (basis en extra) steekproef tegenover de totale gerealiseerde steekproef voor enkele socio-economische kenmerken (Bron: Ongewogen cijfers uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming).

Socio-economische kenmerken

	Gerealiseerde steekproef Basis [1]	Gerealiseerde steekproef Extra [2]	Gerealiseerde steekproef Totaal [5]	Geselecteerde steekproef Basis [3]	Geselecteerde steekproef Extra [4]
Bruto belastbaar gezinsinkomen (uit (zelfstandige) arbeid, excl. uitkeringen)					
<15.000	9,0%	16,6%	10,3%	9,4%	21,1%
15.000-30.000	12,5%	26,5%	14,9%	13,5%	26,9%
30.000-45.000	18,1%	23,3%	19,0%	17,6%	23,1%
45.000-60.000	17,9%	12,4%	17,0%	18,0%	14,2%
60.000-75.000	15,8%	11,8%	15,1%	15,3%	7,0%
>75.000	26,7%	9,4%	23,7%	26,2%	7,8%
<i>n</i>	2.102	434 ³	2.536 ^{1,2}	8.577	1.741

Bruto belastbaar gezinsinkomen (incl. uitkeringen)						
<15.000	6,1%	13,8%	7,5%	6,4%	16,6%	
15.000-30.000	11,3%	27,9%	14,2%	12,0%	25,1%	
30.000-45.000	17,3%	21,8%	18,1%	17,2%	24,5%	
45.000-60.000	19,8%	15,1%	19,0%	19,5%	16,8%	
60.000-75.000	16,6%	11,4%	15,7%	16,7%	8,1%	
>75.000	28,9%	9,9%	25,6%	28,1%	8,9%	
<i>n</i>	2.149	463	2.612 ^{4,5}	8.810	1.882	

¹Significant verschil (Pearson $\chi^2=12,5$, $p=0,029$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3];

²Significant verschil (Pearson $\chi^2=12,2$, $p=0,03$) tussen totaal gerealiseerde steekproef [5] en basis gerealiseerde steekproef [1];

³Significant verschil (Pearson $\chi^2=15,5$, $p=0,008$) tussen extra gerealiseerde steekproef [2] en extra geselecteerde steekproef [4].

⁴Significant verschil (Pearson $\chi^2= 17,8$, $p=0,003$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3]

⁵Significant verschil (Pearson $\chi^2=17,6$, $p=0,004$) tussen totaal gerealiseerde steekproef [5] en basis gerealiseerde steekproef [1]

	Gerealiseerde steekproef Basis [1]	Gerealiseerde steekproef Extra [2]	Gerealiseerde steekproef Totaal [5]	Geselecteerde steekproef Basis [3]	Geselecteerde steekproef Extra [4]
Nomenclatuur					
Werknemer	67,8%	55,6%	65,5%	66,9%	49,7%
Zelfstandige of helper	14,3%	6,9%	12,9%	14,2%	8,5%
Werkzoekende	4,5%	9,3%	5,4%	4,7%	10,4%
Loopbaanonderbreking of tijdkrediet	0,3%	0,6%	0,3%	0,4%	0,8%
Leefloon	0,3%	2,0%	0,7%	0,6%	2,5%
Pensioen	2,0%	0,2%	1,7%	1,9%	1,0%
Arbeidsongeschiktheid	2,3%	2,0%	2,2%	2,8%	4,4%
Andere	8,5%	23,3%	11,2%	8,5%	22,8%
<i>n</i>	2.189	493	2682 ^{6,7}	9.000	2.000
Type contract					
Voltijds	41,2%	28,8%	39,0%	40,6%	27,8%
Deeltijds	25,5%	23,5%	25,1%	25,2%	19,0%
Speciaal contract	1,2%	3,2%	1,6%	1,3%	3,2%
Technisch werkloos	0,0%	0,0%	0,0%	0,0%	0,0%
Geen contract/ geen info	32,0%	44,4%	34,3% ⁸	32,7%	50,2%

<i>n</i>	2.189	493	2.682	9.000	2.000
Inkomensbronnen					
Uit arbeid	91,0%	83,6%	89,6%	90,6%	81,9%
Uit zelfstandige activiteit	23,2%	12,0%	21,1%	21,4%	13,6%
(Equivalent) leefloon of bestaansminimum	1,4% ⁹	8,5%	2,7%	2,2%	7,8%
Uitkering voor werkloosheid, activering, loopbaanonderbreking of tijdskrediet	42,6%	44,8%	43,2%	42,5%	50,7%
Uitkering voor arbeidsongeschiktheid, (beroeps)ziekte, invaliditeit of arbeidsongeval	28,7%	26,3%	27,0%	29,2%	32,4%
Pensioenen	5,1%	3,2%	4,8%	5,3%	5,8%
Hulp aan bejaarden, IVT, IT	2,0%	2,9%	2,2%	2,2%	3,5%
<i>n</i>	2.185	482	2.667	8.984	1.968

⁶Significant verschil (Pearson $\chi^2=17,1$, $p=0,017$) tussen totaal gerealiseerde steekproef [5] en basis gerealiseerde steekproef [1];

⁷Significant verschil (Pearson $\chi^2=25,1$, $p=0,001$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3];

⁸Significant verschil (Pearson $\chi^2=9,68$, $p=0,046$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3];

⁹Significant verschil (Pearson $\chi^2=5,4$, $p=0,02$) tussen basis gerealiseerde steekproef [1] en basis geselecteerde steekproef [3].

	Gerealiseerde steekproef Basis [1]	Gerealiseerde steekproef Extra [2]	Gerealiseerde steekproef Totaal [5]	Geselecteerde steekproef Basis [3]	Geselecteerde steekproef Extra [4]
Werkintensiteit (def 1)					
Zeer lage werkintensiteit	6,6%	15,4%	8,2%	7,5%	20,1%
Lage werkintensiteit	6,5%	10,0%	7,2%	6,1%	12,1%
Medium werkintensiteit	11,8%	29,2%	14,9%	11,4%	23,7%
Hoge werkintensiteit	23,1%	19,0%	22,4%	21,3%	20,0%
Zeer hoge werkintensiteit	51,9% ¹	26,5%	47,3%	53,7%	24,1%
<i>n</i>	2.174	480	2.654 ^{9,10}	8.934	1.962 ¹¹
Werkintensiteit (def 2)					
Zeer lage werkintensiteit	8,1%	18,1%	9,9%	9,1%	23,4%
Lage werkintensiteit	9,6%	15,2%	10,6%	8,5%	16,7%
Medium werkintensiteit	11,8%	25,8%	14,4%	12,1%	20,7%
Hoge werkintensiteit	28,9%	20,4%	27,4%	27,8%	22,0%
Zeer hoge werkintensiteit	41,6%	20,4%	37,8%	42,5%	17,1%
<i>n</i>	2.174	480	2.654 ^{12,13}	8.934 ¹⁴	1.962

⁹Significant verschil (Pearson $\chi^2=18,9$, $p=0,001$) tussen totaal gerealiseerde steekproef [5] en basis gerealiseerde steekproef [1]; ¹⁰Significant verschil (Pearson $\chi^2=42,0$, $p=0,000$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3]; ¹¹Significant verschil (Pearson $\chi^2=11,6$, $p=0,021$) tussen extra gerealiseerde steekproef [2] en extra geselecteerde steekproef [4].

¹²Significant verschil (Pearson $\chi^2=16,8$, $p=0,002$) tussen totaal gerealiseerde steekproef [5] en basis gerealiseerde steekproef [1]; ¹³Significant verschil (Pearson $\chi^2=30,8$, $p=0,000$) tussen totaal gerealiseerde steekproef [5] en basis geselecteerde steekproef [3]; ¹⁴Significant verschil (Pearson $\chi^2=12,8$, $p=0,012$) tussen extra gerealiseerde steekproef [2] en extra geselecteerde steekproef [4].

Vervolgens bespreken we een aantal **socio-economische gegevens** (Tabel 11). Het gaat hier om de bruto belastbare gezinsinkomens, de nomenclatuur van de socio-economische positie, het contracttype, de inkomensbronnen en de werkintensiteit.

Via de koppeling met gegevens vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming verkregen we individuele en huishoudelijke gegevens om de **bruto belastbare inkomsten** van de geselecteerde en deelgenomen steekproefpersonen te vergelijken. In Tabel 11 tonen we twee soorten inkomens: het bruto belastbare gezinsinkomen uit (zelfstandige) arbeid (exclusief uitkeringen) en het totale bruto belastbare gezinsinkomen (inclusief uitkeringen). Voor beide inkomensverdelingen vinden we geen significante verschillen tussen de gerealiseerde basissteekproef en de geselecteerde basissteekproef. Ook op het vlak van bruto belastbare inkomens is er geen vertekening in de gerealiseerde steekproef in vergelijking met de selectie. We kunnen dus stellen dat zowel de lage-, midden- als hoge inkomensgezinnen adequaat vertegenwoordigd zijn in de gerealiseerde basissteekproef.

We stellen daarentegen sterke verschillen vast tussen de basissteekproef en de extra steekproef. De steekproef van niet-EU-persoonen worden voornamelijk gekenmerkt door lage inkomens: 48,0% van de extra geselecteerde steekproef heeft een gezinsinkomen uit (zelfstandige) arbeid dat lager ligt dan 30.000 euro per jaar; in de basisselectie is dat minder dan de helft (22,9%). Dit aandeel ligt lager wanneer we de uitkeringen toevoegen. Rekening houdend met uitkeringen heeft 41,7% van de gezinnen met niet-Europese herkomst een gezamenlijk belastbaar inkomen minder dan 30.000 euro per jaar; in de basisselectie is dat 18,4%. De keerzijde hiervan is dat meer respondenten in de basisselectie hoge inkomens laten optekenen: 31,5% heeft een gezamenlijk belastbaar inkomen hoger dan 60.000 euro per jaar. Voor de extra selectie is dat 14,8%. Met andere woorden, de inkomensprofielen van beide geselecteerde –maar ook gerealiseerde- steekproeven wijken sterk van elkaar af. Onderzoek van Corluy and Verbist (2010) stelt dat er een inkomens kloof heerst tussen personen met een Belgische en personen met een niet-EU-herkomst. Het gemiddelde arbeidsinkomen van deze laatste groep zou bijna een derde lager liggen dan van de eerste groep. Tot een gelijkaardige vaststelling komen we bij het vergelijken van de bruto belastbare gezinsinkomens¹⁸ van de geselecteerde steekproefpersonen in de basis en extra steekproef. Terwijl we geen significante verschillen vinden voor de *basis*respondenten ten aanzien van de *basis*selectie wat betreft hun inkomens, zien we wel verschillen tussen de totaal gerealiseerde steekproef en de basisselectie. Dit betekent dat de extra steekproef voor een lichte oververtegenwoordiging zorgde van de laagste inkomensgroepen (<15.000 euro en 15.000-30.000 euro per jaar) en een lichte ondervertegenwoordiging van de hoogste inkomensgroep (>75.000 euro).

De arbeidsmarkt is het belangrijkste domein voor gezinnen om een inkomen te verwerven om te voorzien in hun bestaanszekerheid. Een eerste manier om deze arbeidsmarkt te bestuderen is via de **nomenclatuur van de socio-economische positie**. Dit is de arbeidsmarktpositie van het individu op de laatste dag van het kwartaal¹⁹. De basisrespondenten bestaan voornamelijk uit werknemers (67,8%) en zelfstandigen (14,3%) en ‘andere’ (8,5%). Van de extra respondenten zijn iets meer dan de helft werknemers (55,6%), zijn er minder zelfstandigen (6,9%) en meer extra respondenten behoren tot de

¹⁸ Wat niet wordt opgenomen in de bruto belastbare gezinsinkomens is het gebruik van een bedrijfswagen, maaltijdcheques, aanvullingen op het dubbel vakantiegeld, vergoedingen voor het stopzetten van de opzegtermijn op onrechtmatige wijze en vergoedingen bij ontslag van beschermde werknemers.

¹⁹ Er wordt slechts één nomenclatuurpositie per individu toegekend. Wanneer steekproefpersonen meerdere arbeidsmarktposities innemen, worden ze toegewezen aan de positie die het dichtst aansluit bij de arbeidsmarkt. Zo worden, bijvoorbeeld, personen gerekend tot ‘werkend’, ook wanneer ze deeltijds tijdscrediet zouden opnemen, of deeltijds werkzoekend zijn. Analoog worden ook deeltijds werkende gepensioneerden ingedeeld als ‘werkend’ (en niet bij de gepensioneerden). (Meer informatie hierover via www.ksz-bcss.fgov.be).

restcategorie (23,3%). Bovendien zijn 9,3% van de extra respondenten werkzoekend, versus 4,5% bij de basisrespondenten. We stellen vast dat de basisrespondenten geen goede weerspiegeling is van de basisselectie met betrekking tot de nomenclatuur (i.e. we vinden een significant verschil tussen de gerealiseerde en geselecteerde basisteekproef) als gevolg van een lichte oververtegenwoordiging van het aandeel werknemers en lichte ondervertegenwoordiging van respondenten die een leefloon ontvangen. Ook de totaal gerealiseerde steekproef wijkt op betekenisvolle wijze af van de basisselectie: door de extra steekproef, zien we dat werkenden plots licht ondervertegenwoordigd zijn, en dat de restcategorie sterk toeneemt.

Een tweede manier om naar de arbeidsmarkt te kijken, is via het **contracttype**. In de gezinsenquête is de vraag 'In uw huidige job, heeft u...' opgenomen, met contracttypes als mogelijke antwoorden. Het gros van de respondenten (95%) blijkt over een vast contract te beschikken: 64,8% geeft aan over een contract onbepaalde duur te beschikken en 30,4% zou een vaste benoeming hebben. Daartegenover heeft vijf procent een tijdelijke overeenkomst: 3% stelt te beschikken over een contract bepaalde duur, 1,6% over een interim contract, 0,2% over een stagecontract en 0,1% over geen schriftelijk arbeidsovereenkomst. Deze variabele blijkt echter moeilijk te repliceren op basis van de gegevens uit het Datawarehouse. Er is namelijk geen informatie beschikbaar over het contracttype bij 34,3% van de respondenten van de gezinsenquête: dit betekent dat er ofwel een groot aandeel is zonder contract ofwel dat deze informatie niet werd aangeleverd. Van degene waarover we informatie verkregen, konden we een opsplitsing maken tussen (1) voltijds of (2) deeltijds werkend, (3) werkend onder een speciaal contract²⁰, (4) technisch werkloos²¹ zijn en (5) het niet hebben van een contract of niet gekend in de dataset.

²⁰ Het hebben van een 'speciaal' contract betekent dat deze persoon korte of onregelmatige contracten aangaat zoals een interim job, seizoensarbeid, gelegenheidsarbeid in de horeca of gelegenheidsarbeid in land- en tuinbouw.

²¹ Vanaf 2011 kent het DWH AM&SB een extra waarde toe aan het contracttype, namelijk 'Afwezigheid wegens (langdurige) ziekte'. Het gaat om een beperkt aantal personen die als het ware technisch werkloos zijn, bijvoorbeeld doordat deze persoon een langdurige ziekte heeft of personen 'in afwezigheid van een passende betrekking' en werkend in de publieke sector (Meer informatie via www.ksz-bcss.fgov.be).

Van de respondenten die deelnamen aan de gezinsenquête beschikken 39,0% over een voltijds arbeidscontract en 25,1% over een deeltijds contract. Daarnaast heeft 1,6% een 'speciaal' contract, één respondent is technisch werkloos²². Let wel, van 919 respondenten of 34,3% hebben we geen informatie. Opmerkelijk is klaarblijkelijk het aandeel respondenten met een voltijds contract die beslisten om deel te nemen aan de *basissteekproef* (41,2%) vergeleken met de *extra steekproef* (28,8%). De verhouding van de contracttypes liggen anders in de extra gerealiseerde steekproef: hier is het aandeel van personen met een speciaal contract hoger (3,2%) dan bij de basisrespondenten (1,3%).

In Tabel 11 geven we eveneens een beknopt overzicht van de verschillende **inkomensbronnen** van de gezinnen in de gezinsenquête. We maken een onderscheid tussen: 1) de inkomens uit arbeid; 2) de inkomens uit zelfstandige activiteit; 3) het leefloon of bestaansminimum; 4) de uitkering voor werkloosheid of activering, loopbaanonderbreking of tijdskrediet; 5) de uitkering voor arbeidsongeschiktheid, (beroeps)ziekte, invaliditeit of arbeidsongeval; 6) het pensioen; en 7) de tegemoetkoming voor hulp aan bejaarden (THAB), de integratietegemoetkoming (IT), de inkomensvervangende tegemoetkoming (IVT) en de gewone en bijzondere tegemoetkoming aan mindervaliden. Omdat enkele inkomensbronnen verspreid zitten over instellingen die gegevens aanleveren aan het Datawarehouse, bundelen we enkele bronnen²³. De proportie huishoudens die deze bronnen ter beschikking hebben telt niet op tot 100%. Een huishouden kan namelijk meerdere inkomsten combineren. We stellen vast dat het inkomen uit arbeid voor de geselecteerden de hoofdcomponent van het totale inkomen inhoudt. Het aandeel geselecteerden uit de extra steekproef met een inkomen uit arbeid is echter beduidend lager dan deze voor de basissteekproef. Respectievelijk halen 90,6% van de basis geselecteerden een inkomen uit arbeid terwijl dit voor de extra geselecteerden 81,9% is.

Wat betreft inkomens uit zelfstandige activiteit in bij- of hoofdberoep put het gros van de deelnemers hier geen inkomens uit. Als inkomensbestanddeel lijkt dit voornamelijk een aanzienlijk bedrag te betekenen voor enkele basis geselecteerden. Algemeen ligt het aandeel van basis geselecteerden met een bij- of hoofdberoep (21,4%) aanzienlijk hoger dan deze van de extra geselecteerden (13,6%). Maar daarnaast ligt dit specifieke inkomen ook veel hoger in de basisselectie: 16,3% van de basisrespondenten verdient meer dan 60.000 euro per jaar door deze activiteit, voor de extra respondenten is dit slechts 6,8%.

Omgekeerd zien we een hoger aandeel van de niet-EU-steekproefpersonen van de extra steekproef bij de inkomens uit uitkeringen: 7,9% van de extra steekproef neemt een leefloon of bestaansminimum op (*zie verder*), meer dan de helft (50,7%) beschikt over een uitkering voor werkloosheid of loopbaanonderbreking en bijna één op drie (32,4%) heeft een uitkering voor arbeidsongeschiktheid of beroepsziekte. Voor de basis geselecteerden is dat respectievelijk 2,2%, 42,5% en 29,2%.

Van de basis geselecteerde steekproefpersonen ontvangen 2,2% een leefloon. Opnieuw wijkt de extra steekproef sterk af van de basissteekproef. Dit cijfer van de basisselectie verschilt significant met het

²² Vanaf 2011 kent het DWH AM&SB een extra waarde toe aan het contracttype, namelijk 'Afwezigheid wegens (langdurige) ziekte. Het gaat om een beperkt aantal personen die als het ware technisch werkloos zijn, bijvoorbeeld doordat deze persoon een langdurige ziekte heeft of personen 'in afwezigheid van een passende betrekking' en werkend in de publieke sector (Meer informatie via www.ksz-bcss.fgov.be).

²³ Het inkomen uit arbeid leiden we bijvoorbeeld af uit het Rijksregister Sociale Zekerheid en de Rijksdienst voor Sociale Zekerheid van de Provinciale en Plaatselijke Overheden. De uitkering voor arbeidsongeschiktheid, (beroeps)ziekte, invaliditeit of arbeidsongeval is een samengestelde inkomensbron vanuit vier verschillende instellingen: het Nationaal Intermutualistisch College (NIC), het Fonds voor Arbeidsongevallen (FAO), het Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV) en het Fonds voor Beroepsziekten (FBZ).

aandeel personen die men effectief heeft bereikt. Slechts 1,4% van de *basis*respondenten ontving in 2015 deze uitkering. Bijkomende berekeningen bekeken het profiel van de basis respondent met het profiel van de niet-respondenten van deze basissteekproef. Ook hier wijken de cijfers significant af: 1,4% van de respondenten is leefloongerechtigd, voor de niet-respondenten is dat 2%. Hieruit blijkt de potentiële meerwaarde van de extra steekproef. Omdat een hoger aandeel van de steekproefpersonen met een niet-EU-herkomst inkomen uit uitkeringen ontvangt, compenseert de extra steekproef de ondervertegenwoordiging van leefloonontvangers in de basissteekproef. 7,9% van de extra geselecteerden is namelijk gekend bij POD MI, wat bijgevolg zelfs opmerkelijk resulteert in een lichte oververtegenwoordiging bij de *extra* respondenten (8,5%). De extra steekproef wordt, met andere woorden, niet enkel gekenmerkt door lage inkomensgezinnen, maar ook door een hoger aandeel aan gezinnen die met een leefloon moeten rondkomen.

Een andere, en laatste, manier om naar de tewerkstelling van de respondenten en niet-respondenten te kijken is door de **arbeidsintensiteit** onder de loep te nemen. Deze *work intensity* berekent de tijd die de volwassen gezinsleden op actieve leeftijd *effectief* (in de praktijk) besteedden aan betaalde arbeid in verhouding tot de tijd die ze *mogelijk* (in theorie) hadden kunnen besteden. Men zet dus het werkelijk gepresteerde jaarlijkse arbeidsvolume op huishoudniveau af tegenover het potentiële jaarlijkse arbeidsvolume. Concreet betekent dit dat als alle volwassenen ononderbroken en voltijds werkten, dit gezin een werkintensiteit van 100% werd toegekend en als ‘werkrijk’ kan benoemd worden. Indien géén van de gezinsleden werkten, heeft dit ‘werkarm’ gezin een werkintensiteit van 0%. Eurostat (2018) maakt een onderscheid tussen vijf categorieën: een zeer lage werkintensiteit (0-20%), een lage werkintensiteit (20-45%), een medium werkintensiteit (45-55%), een hoge werkintensiteit (55-85%) en een zeer hoge werkintensiteit (85-100%). Een persoon leeft dus in een huishouden met een zeer lage werkintensiteit wanneer het feitelijke aantal maanden dat in het huishouden wordt gewerkt, kleiner is dan 20% van het maximale aantal maanden waarin alle volwassen gezinsleden hadden kunnen werken.

Het Datawarehouse maakt een onderscheid tussen twee definities van deze arbeidsintensiteit. De eerste definitie sluit zo nauw mogelijk aan bij dezelfde definitie die de EU-SILC hanteert, terwijl de tweede definitie geconstrueerd werd op basis van de beschikbare informatie uit het Datawarehouse zelf (Braes & Herremans, 2012)²⁴. In dit rapport geven we steeds beide definities weer. Voor beiden stellen we namelijk duidelijke patronen vast wat betreft de verschillen tussen de twee getrokken steekproeven enerzijds en de deelnemende respondenten anderzijds.

Tabel 11 toont duidelijk aan dat de werkrijke gezinnen sterk vertegenwoordigd zijn binnen de groep van Vlaamse gezinnen met kinderen: 53,7% (eerste definitie) of 42,5% (tweede definitie) van de basisselectie heeft een zeer hoge werkintensiteit. In de gerealiseerde basissteekproef kennen ‘slechts’ 6,6% resp. 8,1% van de deelnemende respondenten een zeer lage werkintensiteit. Deze verdeling van de gerealiseerde basissteekproef wijkt niet significant af van de basisselectie. In lijn met de verschillen op het vlak van de inkomens zien we ook nu een discrepantie tussen de basis- en extra steekproef. De extra steekproef kent een aanzienlijk hoger aandeel geen- of eenverdienersgezinnen in vergelijking met de

²⁴ De eerste definitie maakt gebruik van het arbeidsvolume zoals bepaald in het statuut op de laatste dag van het kwartaal (en de maatman vastgelegd op 35 uur). De tweede definitie maakt gebruik van het reëel gepresteerde arbeidsvolume in het kwartaal (en de reële maatman zoals gekend in de onderneming of bedrijfstak). Bijgevolg ligt de arbeidsintensiteit volgens deze tweede definitie lager dan de intensiteit op basis van de eerste definitie (Braes & Herremans, 2012).

basissteekproef. Het gevolg is dat de totaal gerealiseerde steekproef voor beide definities van werkintensiteit een sterke ondervertegenwoordiging van de gezinnen met een zeer hoge werkintensiteit kent ten opzichte van de basisselectie.

In Tabel 12 vergelijken we het profiel van de basisrespondenten met het profiel van de basis niet-respondenten. Bij de niet-respondenten zien we ten opzichte van de respondenten een oververtegenwoordiging van gezinnen met een zeer lage en een zeer hoge werkintensiteit. Tegelijkertijd zien we een lichte ondervertegenwoordiging van gezinnen met een hoge werkintensiteit. Gebaseerd op de eerste definitie zijn de gezinnen met een zeer lage werkintensiteit ondervertegenwoordigd in de gezins-enquête. Men slaagde er slechts in om 6,6% van deze gezinnen te bereiken. Opmerkelijk is daarnaast eveneens de ondervertegenwoordiging van de jobrijke gezinnen: 51,9% van de basisrespondenten leeft in een gezin met een zeer hoge werkintensiteit, terwijl dit aandeel hoger ligt bij de niet-respondenten. Gezinnen met een hoge werkintensiteit zijn dan weer licht oververtegenwoordigd in de gezins-enquête, wat algemeen werkrijke gezinnen in evenwicht brengt. De werkintensiteit op basis van de tweede definitie toont eveneens dat een groter aandeel van de niet-respondenten een zeer lage werkintensiteit rapporteren.

Tabel 12 Werkintensiteit van basisrespondenten versus basis niet-respondenten (Bron: DWH AM&SB)

	Gerealiseerde steekproef BASIS	Niet-gerealiseerde steekproef BASIS	Gerealiseerde steekproef TOTAAL
Werkintensiteit (def. 1)			
Zeer lage werkintensiteit	6,6%	7,8%	8,2%
Lage werkintensiteit	6,5%	5,9%	7,2%
Medium werkintensiteit	11,8%	11,3%	14,9%
Hoge werkintensiteit	23,1%	20,7%	22,4%
Zeer hoge werkintensiteit	51,9%	54,2%	47,3%
<i>n</i>	2.174 ¹	6.760	2.654 ^{2,3}
Werkintensiteit (def. 2)			
Zeer lage werkintensiteit	8,1%	9,3%	9,9%
Lage werkintensiteit	9,6%	8,1%	10,6%
Medium werkintensiteit	11,8%	12,1%	14,4%
Hoge werkintensiteit	28,9%	27,4%	27,4%
Zeer hoge werkintensiteit	41,6%	42,7%	37,8%
<i>n</i>	2.174	6.760	2.654 ^{4,5}

¹Significant verschil (Pearson $\chi^2=10,4$, $p=0,034$) tussen gerealiseerde basissteekproef en niet-gerealiseerde basissteekproef. ²Significant verschil (Pearson $\chi^2=44,5$, $p=0,000$) tussen niet-gerealiseerde basissteekproef en totaal gerealiseerde steekproef; ³Significant verschil (Pearson $\chi^2=18,9$, $p=0,001$) tussen gerealiseerde basissteekproef en totaal gerealiseerde steekproef. ⁴Significant verschil (Pearson $\chi^2=31,7$, $p=0,000$) tussen niet-gerealiseerde basissteekproef en totaal gerealiseerde steekproef. ⁵Significant verschil (Pearson $\chi^2=16,8$, $p=0,002$) tussen gerealiseerde basissteekproef en totaal gerealiseerde steekproef.

3 Multivariate analyse

Naast de beschrijvende analyse van de verdeling van sociale, demografische en economische achtergrondkenmerken van de respondenten in de gezinsenquête, gaan we in deze sectie door middel van een multivariate analyse na welke achtergrondkenmerken het al dan niet deelnemen aan de enquête hebben beïnvloed. We maken gebruik van een logistische regressie waarbij we de kans inschatten van deelname (of niet) aan de gezinsenquête aan de hand van een aantal achtergrondkenmerken. De centrale vraag die we hierbij stellen is of gezinnen met bepaalde achtergrondkenmerken een grotere kans hebben om niet deel te nemen in vergelijking met gezinnen die deze kenmerken niet bezitten.

Op basis van de voorgaande beschrijvende analyse formuleren we enkele hypothesen over de verschillende determinanten die de non respons in de hand kunnen werken. We operationaliseerden een dummyvariabele 'deelname' die aangeeft of de geselecteerde steekproefpersoon van de basisselectie (Model 1), van de extra selectie (Model 2) of de gehele selectie (Model 3) deelneemt of niet (deelname = 1). Kortom, we splitsen deze analyse op voor de basis- en extra steekproef om een zicht te krijgen op welke determinanten in beide steekproeven een effect kunnen hebben op de participatiegraad. De onafhankelijke variabelen zijn als volgt geoperationaliseerd:

- Huishoudtype (3 categorieën): gehuwd, ongehuwd, alleenstaand;
- Arbeidscontract (5 categorieën): voltijds, deeltijds, speciaal, technisch werkloos, geen contract of geen informatie beschikbaar;
- Bruto belastbare gezinsinkomens (5 categorieën): <15.000, 15.000-30.000, 30.000-45.000, 45.000-60.000, 60.000-75.000, >75.000;
- Werkintensiteit (5 categorieën): zeer laag, laag, medium, hoog of zeer hoog.

Tabel 13 toont de schatting van het model, uitgedrukt in odds ratio's (OR). Een OR groter dan 1 betekent hier dat de kans van deelname van dit kenmerk groter is dan de kans op deelname bij de referentiecategorie. Een OR kleiner dan 1 betekent het omgekeerde, namelijk dat de kans op deelname van de referentiecategorie groter is dan deze van de niet-referentie categorie.

In het eerste model schatten we de waarschijnlijkheid van de deelname van de geselecteerden in de geselecteerde basissteekproef in de gerealiseerde basissteekproef. Algeheel blijkt dit model niet significant te zijn. Zoals eerder vermeld in dit hoofdstuk is de geselecteerde basissteekproef een aselechte steekproef van de doelpopulatie, en is de gerealiseerde basissteekproef een goede weerspiegeling te zijn van deze specifieke selectie. Kortom, deze analyse bevestigt wat we eerder zagen in de beschrijvende tabellen. We vinden geen evidentie dat bepaalde kwetsbare socio-economische profielen een lagere kans op deelname zouden hebben. Gezinnen met een laag bruto belastbaar gezinsinkomen (<15.000 euro per jaar) van deze basisselectie hebben een kleinere kans om op deelname, maar dit blijkt niet significant te zijn. We vinden enkel significante afwijkingen voor gezinnen met een bruto belastbaar gezinsinkomen dat tussen 15.000 en 30.000 euro per jaar ligt: zij hebben een betekenisvol lagere kans op deelname in de gerealiseerde basissteekproef. Daarnaast hebben gezinnen met een hoge werkintensiteit een significant hogere kans op deelname (vergeleken met de referentiegroep: gezinnen met een zeer hoge werkintensiteit). Deze logistische regressie bevestigt, met andere woorden, dat de gerealiseerde basissteekproef algemeen een accurate weerspiegeling is van de geselecteerde basissteekproef.

Tabel 13 Overzicht resultaten logistische regressie van de deelname van de gerealiseerde basis, extra en totale steekproef (Bron: DWH AM&SB)

	Deelname <u>basis</u> gezinsenquête [Model 1] n= 8.977 n.s.		Deelname <u>extra</u> gezinsenquête [Model 2] n= 1.876 ***		Totale deelname gezinsenquête [Model 3] n= 10.280 **	
	Odds ratio	St. fout	Odds ratio	St. fout	Odds ratio	St. fout
Gezinstype						
Ongehuwd	0,9	0,0	0,9	0,2	0,9	0,0
Alleenstaand	1,0	0,1	1,2	0,2	1,1	0,1
Andere	0,7	0,3	0,8	0,4	0,9	0,3
<i>Ref.: gehuwd</i>						
Type contract						
Deeltijds	1,0	0,1	1,5*	0,2	1,0	0,1
Speciaal	0,9	0,2	1,1	0,4	1,0	0,2
Technisch werkloos	0,9	1,0	-	-	0,9	1,0
Geen contract/geen info	1,0	0,1	0,9	0,1	1,0	0,0
<i>Ref.: voltijds</i>						
Werkintensiteit						
Zeer lage WI	1,1	0,2	0,6	0,1	1,0	0,1
Lage WI	1,3	0,1	0,7	0,1	1,2	0,1
Medium WI	1,2	0,1	1,2	0,2	1,3***	0,1
Hoge WI	1,2**	0,1	0,7*	0,1	1,1*	0,1
<i>Ref.: zeer hoge WI</i>						
Gezinsinkomen						
<15.000	0,8	0,1	0,9	0,1	0,8*	0,1
15.000-30.000	0,8**	0,1	1,0	0,2	0,8**	0,1
30.000-45.000	0,9	0,1	0,8	0,2	0,9	0,1
45.000-60.000	0,9	0,1	0,7	0,1	0,9	0,1
60.000-75.000	1,0	0,1	1,4	0,3	1,1	0,1
<i>Ref.: >75.000</i>						
Constant	0,3***	0,0	0,4***	0,1	0,3***	0,0

* $p < 0.05$, Model 1: LR $Chi^2 = 21,5$, $p = 0,20$, pseudo $R^2 = 0,0022$
 ** $p < 0.01$, Model 2: LR $Chi^2 = 48,2$, $p = 0,02$, pseudo $R^2 = 0,0231$
 *** $p < 0.001$, Model 3: LR $Chi^2 = 31,7$; $p = 0,01$, Pseudo $R^2 = 0,0028$

In het tweede model schatten we de waarschijnlijkheid van de deelname van de geselecteerde extra steekproef in de gerealiseerde extra steekproef. Hiervoor houden we rekening met dezelfde kenmerken als in model 1. Dit tweede model telt 1.876 observaties en is wel significant. We stellen vast dat het gezinstype geen significant effect heeft op de kans op deelname. Wel significant is de relatie tussen arbeidscontract, en met name deeltijds werken, en deelname. De kans dat een deeltijds werkende steekproefpersoon met niet-EU-herkomst deelneemt is hoger dan de kans dat een voltijds werkende

steekproefpersoon met niet-EU-herkomst participeert. Ook de hoge werkintensiteit beïnvloedt op significante wijze of een steekproefpersoon de enquête invult. De kans dat een persoon die leeft in een huishouden met een hoog werkintensiteit de enquête invult is lager dan een persoon die leeft in een huishouden met een zeer hoge werkintensiteit. Tot slot heeft een gezin met een jaarlijks inkomen dat lager ligt dan 15.000 euro een kleinere kans op deelname dan een inkomensrijk gezin, maar de impact van deze inkomens is niet significant op de deelname van deze extra respondenten.

In het derde model bekijken we de kans op deelname van de *totale* selectie in de *totale* gerealiseerde steekproef. Ook in dit model heeft het gezinstype geen significante invloed op de deelname. In vergelijking met de extra steekproef blijkt het arbeidscontract hier geen effect te hebben. De werkintensiteit van het gezin heeft wel een effect op participatie. De kans op deelname voor huishoudens met een medium of hoge werkintensiteit is significant hoger dan deelname van huishoudens met een zeer hoge werkintensiteit. Ook het gezinsinkomen heeft een significant effect, voornamelijk wanneer we de onderkant van de inkomensverdeling bekijken. De kans dat een huishouden met een laag gezinsinkomen (>30.000) deelneemt is kleiner dan de kans dat een huishouden uit de hoogste inkomensklasse deelneemt.

Algemeen concluderen we dat deze analyse een bevestiging is van de daarvoor beschreven beschrijvende tabellen. Deze logistische regressie toont aan dat de gerealiseerde basissteekproef een accurate weerspiegeling is van de geselecteerde basissteekproef. Vertrekkend vanuit de assumptie dat bepaalde zwakkere socio-economische profielen ondervertegenwoordigd zouden zijn, kunnen we stellen dat dit niet het geval is. Gezinnen met een laag bruto belastbaar gezinsinkomen (<15.000 euro per jaar) van deze basisselectie hebben een kleinere kans op deelname, maar dit blijkt niet significant te zijn. We vinden enkel significante afwijkingen voor gezinnen met een bruto belastbaar gezinsinkomen dat tussen 15.000 en 30.000 euro per jaar ligt: zij hebben een significant lagere kans op deelname in de gerealiseerde *basissteekproef*. Daarnaast hebben gezinnen met een hoge werkintensiteit een significant hogere kans op deelname.

4 Item non respons

In dit hoofdstuk kregen we reeds een zicht op de *unit non respons*, met andere woorden wie wel en wie niet heeft deelgenomen. Dit deden we door administratieve gegevens onder de loep te nemen van zowel de gerealiseerde als de geselecteerde steekproeven. In deze sectie bestuderen we de *item non respons* van de gezinsenquête. Bij deze specifieke non respons gaat het om steekproefpersonen die wel deelnemen aan de bevraging, maar weigeren en/of er niet in slagen om op alle vragen een antwoord te geven. In dat geval krijgt de uiteindelijk verzamelde dataset enkele ontbrekende antwoorden op bepaalde vragen. In de literatuur maakt men dan ook een onderscheid tussen *willekeurig* ontbrekende gegevens en *systematisch* ontbrekende gegevens (Huisman, 2007; Little & Rubin, 2002). Rubin (1976) splitst deze mechanismen op: 1) *missing at random* (MAR) en 2) *missing not at random* (MNAR). MAR betekent dat het beantwoorden van bepaalde vragen zeer willekeurig gebeurt, MNAR betekent dat er bepaalde trends te vinden zijn van profielen die weigeren op vragen te beantwoorden. Deze MNAR kan vertekening in de hand werken wanneer de respondenten die de vraag wel beantwoordden afwijken van de respondenten die dit niet deden. Er is, bijvoorbeeld, sprake van MNAR wanneer het voornamelijk personen met een laag inkomen zijn die de inkomensvraag blanco laten. Wanneer we een hoge item non respons hebben op deze vraag, getekend door voornamelijk lage inkomensgezinnen, verkrijgen we een vertekend beeld van de inkomens in de gezinsenquête. Het is deze potentiële vertekening die we willen onderzoeken in deze sectie.

Allereerst gaan we beschrijvend te werk door de item non respons van een selectie aan kenmerken te bekijken. De variabelen die we hier beschrijven, zijn opleidingsniveau, materiële deprivatie, gezinsinkomen, tewerkstelling, gezondheid en de woonsituatie. Tabel 14 geeft een beknopt overzicht van de percentages van personen die de vraag niet beantwoordden voor de geselecteerde items. Het valt op dat de item non respons over het algemeen beperkt is.

Tabel 14 Item non respons voor een selectie aan items (Bron: ongewogen cijfers uit de gezinsenquête 2016)

	Item non respons in de <u>basis-</u> gerealiseerde steekproef (n= 2.189)	Item non respons in de <u>extra</u> gerealiseerde steekproef (n= 493)	Item non respons in de <u>totale</u> gerealiseerde steekproef (n= 2.682)	Item respons in de <u>totale</u> gerealiseerde steekproef (n= 2.682)
Opleidingsniveau	2,5%	1,8%	2,4%	97,6%
Niet kunnen rondkomen	4,1%	4,0%	4,1%	95,9%
Kunnen veroorloven van jaarlijkse vakantie	2,9%	3,7%	3,0%	97,0%
Inkomens	5,4%	7,3%	5,8%	94,2%
Gezondheid	1,6%	0,4%	1,5%	98,5%

Vervolgens gaan we verder in op de item non respons van enkele genoemde items en hoe het profiel van deze item non respondenten eruitziet. In elk overzicht plaatsen we ook een extra kolom waar we de verdeling van dit kenmerk in de gezinsenquête terugvinden, om een eerste zicht te verkrijgen op de potentiële afwijking van de profielen tussen de item respondenten en item niet-respondenten.

Opleidingsniveau

Van de 2.682 personen in de gezinsenquête²⁵ vulden 64 personen (2,4%) de vraag over het hoogst behaalde diploma niet in: verder opgesplitst gaat het om 55 personen van de basissteekproef, en 9 personen uit de extra steekproef (Tabel 15). Wie zijn dan deze personen die de vraag onbeantwoord laten?

Tabel 15 Profielen van item non respondenten voor het opleidingsniveau (Bron: ongewogen cijfers via DWH AM&SB en de gezinsenquête 2016)

	INR basis (n= 55)	IR basis (n=2.134)	INR extra (n= 9)	IR extra (n=484)	INR totaal (n= 64)	IR totaal (n=2.618)
Gezinsinkomen (incl. uitkeringen)						
<30.000	29,1%	20,4%	55,6%	38,8%	31,2%	20,9%
30.000-60.000	29,1%	34,7%	33,3%	34,7%	29,7%	36,2%
>60.000	40,0%	40,8%	0,0%	20,5%	34,4%	40,3%
Geen informatie	1,8%	4,0%	11,1%	6,0%	4,7%	2,6%
Werkintensiteit (WI)¹						
(Zeer) lage WI	16,4%	13,1%	55,6%	24,8%	21,9%	15,2%
Medium WI	10,9%	11,8%	44,4%	28,9%	15,6%	14,9%
(Zeer) Hoge WI	72,7%	75,1%	0,0%	46,3%	62,5%	69,9%
Geen informatie	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

¹Significant verschil (Pearson $\chi^2=6,2$, $p=0,04$) tussen de item non respons van de extra gerealiseerde steekproef en de item respons van de extra steekproef.

Volgens de beschrijvende cijfers in Tabel 15 is het, bij de basissteekproef, voornamelijk een hoog-opgeleide doelgroep met hoog inkomen en een (zeer) hoge werkintensiteit die de vraag onbeantwoord laat. Maar de item respons van diezelfde steekproef toont aan dat de verdeling gelijkaardig is aan de item respons. In de extra steekproef gaat het daarentegen om lage inkomensgezinnen met een lage tot zeer lage werkintensiteit die het item openlaten. Hier zien we een significant verschil met de groep die de vraag wel invult: 46,3% van de respondenten met een niet-EU-herkomst die leven in een werkdijk

²⁵ We bekijken deze item non respons vanuit beschikbare administratieve gegevens. Met andere woorden, we werken verder met de beschikbare informatie over alle respondenten (n= 2.682) in plaats van met de enquêtegegevens (n= 2.683) doordat we in deze sectie de profielen bekijken aan de hand van gegevens van het DWH AM&SB.

gezin beantwoordden wel degelijk de vraag naar het behaalde diploma. Hier lijken dus voornamelijk gezinnen uit een werkarm gezin hun diploma niet aan te geven (maar het gaat hier slechts om een zeer laag aantal van negen respondenten).

Materiële deprivatie

De gezinsenquête peilde bij haar respondenten in welke mate ze zelf subjectief de eigen leefsituatie inschatten. Dit deed men aan de hand van enkele items, zoals 1) de vraag of men het gevoel heeft met het beschikbare inkomen te kunnen rondkomen en 2) of de respondent en diens huishouden een jaarlijkse vakantie kunnen veroorloven. Hoewel het om een vraag gaat die in theorie elke respondent kan beantwoorden, is de item non respons relatief hoog.

Vooreerst beantwoordden 110 deelnemers (of 4,1%) de vraag niet met betrekking tot **het kunnen rondkomen met het beschikbare gezinsinkomen**. In Tabel 16 splitsen we deze item non respons op in de respondenten van de *basissteekproef*, de *extra steekproef*, om in de laatste kolom de totale (non) respons van het item te bekijken voor alle deelnemers samen. Opnieuw bekijken we op beschrijvende wijze kort welke profielen samenhangen met het niet beantwoorden van deze vragen, om potentiële trends te kunnen vaststellen.

Tabel 16 Profielen van item non respondenten voor materiële deprivatie (1: het kunnen rondkomen met het beschikbare gezinsinkomen)(Bron: ongewogen cijfers via DWH AM&SB).

	INR basis (n= 90)	IR basis (n=2.099)	INR extra (n= 20)	IR extra (n= 473)	INR totaal (n= 110)	IR totaal (n= 2.572)
Gezinsinkomen (incl. uitkeringen)						
<30.000	17,8%	17,0%	40,0%	39,1%	21,8%	21,1%
30.000-60.000	30,0%	36,7%	45,0%	34,3%	32,7%	36,2%
>60.000	52,2%	44,4%	10,0%	20,5%	44,5%	40,0%
Geen informatie	0,0%	1,9%	5,0%	6,1%	1,0%	2,7%
Werkintensiteit (WI)						
Zeer (lage) WI	14,4%	13,0%	35,0%	24,3%	18,2%	15,1%
Medium WI	7,8%	11,9%	25,0%	28,5%	10,9%	14,9%
Zeer (hoge) WI	76,7%	74,5%	40,0%	44,4%	70,0%	68,9%
Geen informatie	1,1%	0,7%	0,0%	0,0%	0,9%	1,1%
Opleidingsniveau*^{1,2}						
Laag	22,2%	12,5%	35,0%	28,3%	24,5%	15,4%
Midden	22,2%	28,7%	30,0%	30,9%	23,6%	29,1%
Hoog	43,3%	54,9%	35,0%	38,3%	41,8%	51,8%
Geen informatie	12,2%	3,9%	0,0%	2,5%	10,0%	3,7%

*Opleidingsniveau is gebaseerd op enquêtegegevens (Bron: gezinsenquête 2016). ¹Significant verschil (Pearson $\chi^2=23,7$, $p=0,000$) tussen de item non respons en item respons van de basissteekproef; ²Significant verschil (Pearson $\chi^2=19,5$, $p=0,000$) tussen de item non respons en item respons van de totaal gerealiseerde steekproef.

Het niet beantwoorden van deze vraag hangt op het eerste zicht niet samen met een specifiek profiel. In alle inkomensklassen vinden we personen die beslisten om deze vraag open te laten. Het is dus niet zo dat het enkel respondenten zijn die leven in een huishouden met een laag bruto belastbaar gezinsinkomen die een vraag onbeantwoord laten. We zien, bijvoorbeeld, dat meer dan de helft (52,2%) van de basisrespondenten die de vraag niet invulden, een inkomen hebben dat hoger is dan 60.000 euro per jaar. Bij de respondenten van de extra steekproef gaat het eerder over gezinnen met een lager inkomen die de vraag openlaten. Daarop aansluitend zien we eenzelfde patroon bij de werkintensiteit en het opleidingsniveau. Het blijkt ook hier voornamelijk de hooggeschoolde respondenten (43,3%) in de basissteekproef met een (zeer) hoge werkintensiteit (76,7%) en een hoog gezinsinkomen (52,5%) die deze vraag niet invulden. Hoewel de verdeling van de 'item niet-respondenten' overeenkomt met de respondenten, zien we echter een significant verschil wat betreft het opleidingsniveau: in de basissteekproef zijn 22,2% van de mensen die de vraag openlaten, laagopgeleid. Dit aandeel ligt significant lager bij de groep die de vraag wel invulden (12,5%). Diezelfde significante afwijking vinden we daaropvolgend eveneens terug bij de totale steekproef.

Tabel 17 Profielen van item non respondenten voor materiële deprivatie (2: het kunnen veroorloven van een jaarlijkse vakantie)(Bron: ongewogen cijfers via DWH AM&SB en gezinsenquête 2016).

	INR basis (n= 63)	IR basis (n= 2.126)	INR extra (n= 18)	IR extra (n=475)	INR totaal (n= 81)	IR totaal (n= 2.601)
Gezinsinkomen (incl. uitkeringen)						
<30.000	17,5%	17,0%	33,3%	39,4%	21,0%	21,1%
30.000-60.000	41,3%	36,3%	50,0%	34,1%	43,2%	35,9%
>60.000	39,7%	44,9%	11,1%	20,4%	33,3%	40,4%
Geen informatie	1,5%	1,8%	5,6%	6,1%	2,5%	2,6%
Werkintensiteit (WI)						
Zeer (lage) WI	14,3%	13,0%	22,2%	24,8%	16,0%	15,2%
Medium WI	11,1%	11,7%	38,9%	28,0%	17,3%	14,7%
Zeer (hoge) WI	73,0%	74,6%	38,9%	44,4%	65,4%	69,1%
Geen informatie	1,6%	0,7%	0,0%	2,7%	1,2%	1,0%
Opleidingsniveau*^{1,2}						
Laag	23,8%	12,6%	27,8%	28,6%	24,7%	15,5%
Midden	27,0%	28,5%	22,2%	31,2%	25,9%	29,0%
Hoog	33,3%	55,0%	50,0%	37,7%	37,0%	51,9%
Geen informatie	15,9%	3,9%	0,0%	2,5%	12,3%	3,6%

*Opleidingsniveau is gebaseerd op enquêtegegevens (Bron: Gezinsenquête 2016). ¹Significant verschil (Pearson $\chi^2=31,9$, $p=0,000$) tussen item non respons en item respons basis gerealiseerde steekproef. ²Significant verschil (Pearson $\chi^2=22,9$, $p=0,000$) tussen item non respons en item respons totaal gerealiseerde steekproef.

Daaropvolgend gaan we na wie de non respondenten zijn voor de vraag rond het kunnen veroorloven van een jaarlijkse vakantie (Tabel 17). Opnieuw is ook hier een gelijkaardig patroon te vinden, namelijk dat de item non respons voor beide steekproeven relatief willekeurig plaatsvindt. De non respondenten

van de basissteekproef zijn verdeeld over de inkomensverdeling, leven in een gezin met een hoge tot zeer hoge werkintensiteit en kennen verscheidene opleidingsniveaus. De extra steekproef wordt gekenmerkt door lage tot midden inkomensgezinnen, die hoogopgeleid zijn, en die in gezinnen leven met verschillende werkintensiteiten. Opnieuw moeten we deze cijfers, voornamelijk wat betreft de extra steekproef, nuanceren daar het gaat om een laag absoluut aantal ($n=18$). Zoals in Tabel 16 is ook hier een lichte oververtegenwoordiging van laagopgeleiden te vinden bij de item non respondenten.

Algeheel zien we voor beide indicatoren van materiële deprivatie dat in feite de non respons relatief willekeurig plaatsvond, behalve voor laagopgeleiden: hier zijn het meer laagopgeleiden die de vraag niet wensen te beantwoorden, dan dat er laagopgeleiden zijn die dat wel doen. Daarnaast valt op dat het voornamelijk 'sterke' profielen zijn die de vragen niet beantwoorden, namelijk personen die in een werkrijk gezin leven. Let wel, van de non respondenten van de voorgenoemde items (110 van kunnen rondkomen, 81 van jaarlijkse vakantie) vinden we 61 personen terug die geen enkele van beide vragen beantwoordden.

Inkomens

In de vragenlijst van de gezinsenquête luidt de inkomensvraag: "Had u en/of iemand anders in uw gezin inkomsten uit één van de volgende inkomensbronnen vorige maand?". De item non respons die we hier beschrijven moet met enige nuance geïnterpreteerd worden. De gezinsenquête peilt namelijk naar het huishoudinkomen van de *maand* voordien. Dit verschilt van administratieve gegevens omdat we op *jaarbasis* de verkregen bruto belastbare gezinsinkomen bekijken. Het is dus perfect mogelijk dat een gezin aangeeft geen inkomsten uit arbeid te ontvangen omdat dit de maand voordien ook niet het geval was, terwijl administratieve gegevens wel aangeven dat het huishouden op één jaar tijd een inkomen uit arbeid verkreeg.

Van de 2.682 respondenten van de gezinsenquête, vulden 2.528 of 94,2% één van de inkomensbronnen in. 155 personen of 5,8% weigerden deze vraag in te vullen. De reden hiervan is niet te achterhalen: het gaat namelijk om een postenquête waarin de respondent niet moet aangeven wat de reden is van het niet invullen van een item. Wat we wel kunnen achterhalen, is *wie* de personen zijn die deze vraag niet hebben ingevuld. Dit doen we aan de hand van beschikbare administratieve gegevens.

Tabel 18 Profiel van item niet-respondenten voor de inkomensvraag (Bron: ongewogen cijfers via DWH AM&SB en gezinsenquête 2016)

	INR basis (n= 119)	IR basis (n= 2.070)	INR extra (n= 36)	IR extra (n= 457)	INR totaal (n= 155)	IR totaal (n= 2.527)
Gezinsinkomen¹ (incl. uitkeringen)						
<30.000	24,4%	16,6%	50,0%	38,3%	30,3%	20,5%
30.000-60.000	29,4%	36,8%	30,6%	35,0%	29,7%	36,5%
>60.000	45,4%	44,7%	8,3%	21,0%	36,8%	40,4%
Geen informatie	2,5%	1,9%	11,1%	5,7%	3,2%	2,6%
Werkintensiteit (WI)						
(Zeer) lage WI	15,1%	12,9%	38,9%	23,6%	20,6%	14,9%
Medium WI	9,2%	11,8%	22,2%	28,9%	12,2%	14,9%
(Zeer) Hoge WI	74,8%	74,5%	36,1%	44,8%	65,8%	69,2%
Geen informatie	0,9%	0,8%	2,8%	2,6%	1,4%	1,0%
Opleidingsniveau*^{2,3}						
Laag	31,1%	11,9%	41,7%	27,6%	33,6%	14,7%
Midden	28,6%	28,4%	22,2%	31,5%	27,1%	29,0%
Hoog	24,4%	56,1%	30,6%	38,7%	25,8% ³	53,0%
Andere	3,4%	1,6%	2,8%	0,4%	3,2%	1,4%
Geen informatie	2,5%	2,0%	2,8%	1,8%	10,3%	1,9%

*Opleidingsniveau is gebaseerd op enquêtegegevens (Bron: Gezinsenquête 2016).¹Significant verschil (Pearson $\chi^2=9,2$, $p=0,027$) tussen item non respons en item respons totaal gerealiseerde steekproef. ²Significant verschil (Pearson $\chi^2=92,9$, $p=0,000$) tussen item non respons en item respons totaal gerealiseerde steekproef; ³Significant verschil (Pearson $\chi^2=93,6$, $p=0,000$) tussen item non respons en item respons basis gerealiseerde steekproef.

Van de basissteekproef zijn het voornamelijk personen uit hogere inkomensgezinnen (45,4%) en werkrijke gezinnen (74,8%) die de inkomensvraag niet beantwoorden. In de extra steekproef zijn het eerder laagopgeleide personen en personen uit inkomensarme gezinnen die de vraag niet invulden. Algemeen geldt ook hier, zoals eerder vermeld, dat het voornamelijk laagopgeleiden zijn die de vraag niet beantwoorden (wijkt significant af met het profiel van de respondenten). Opmerkelijk in Tabel 18 is daarnaast ook het significant verschil in de verdeling van de item respons en item non respons wat betreft de gezinsinkomens in de totaal gerealiseerde steekproef. Van de respondenten die niet reageren op de inkomensvraag, zijn meer dan één op drie (36,8%) personen uit inkomensrijke gezinnen, 30,3% zijn inkomensarme gezinnen. De item non respons blijkt, met andere woorden, deels vertekend te zijn: 1) de laagopgeleiden blijken meer vertegenwoordigd te zijn bij de item non respondenten wanneer het gaat om vragen zoals het kunnen rondkomen met het beschikbare inkomen, het kunnen veroorloven van een jaarlijkse vakantie en de inkomensvraag; 2) significant meer lage inkomensgezinnen lieten de inkomensvraag onbeantwoord, in vergelijking met diegene die de vraag wel beantwoorden.

De conclusie van de analyse van de item non respons kunnen we koppelen aan de keuze in data-verzamelmethode (Hoofdstuk 3). Waarom de respondenten beslissen om een vraag onbeantwoord te laten, is onbekend. Maar omwille van een opvallende afwezigheid van antwoorden van laagopgeleide respondenten, kan dit een belangrijk signaal zijn om de redenen wel te achterhalen. Door de gebruikte methode, namelijk postenquôte, aan te vullen met een meer persoonlijke benadering, kan men de motivering hierachter mogelijks wel onderzoeken. Indien, bijvoorbeeld, een vraag onduidelijk werd gesteld, kan de interviewer deze ter plaatse verduidelijken enerzijds, maar anderzijds ook signalen opvangen om vragenlijsten van volgende edities meer helder te omschrijven. Hoewel er geen kant-en-klare oplossing bestaat over welke methodes men kan hanteren om specifieke doelpopulaties te bereiken, en om item non respons te reduceren, kan een *mixed method* design in volgende edities bijgevolg een uitweg bieden.

Opmerkelijk is dat we van slechts zeven personen van de 135 die deze inkomensvraag niet invulden, effectief geen informatie over een inkomen uit arbeid of zelfstandige activiteit hebben. Eén van deze zeven personen heeft geen enkele huishoudelijke informatie over andere inkomenscomponenten, de overige zes personen hebben wel een uitkering ontvangen in 2015. Hoewel we gemeenschappelijke non respondenten vonden voor de vragen gerelateerd aan de materiële deprivatie, vinden we geen respondenten terug die én de inkomensvraag niet beantwoordden én, bijvoorbeeld, de diplomavraag openlieten.

5 Strategieën om de respons te verhogen

Hoe kan men de non respons of vertekening tegengaan? In de literatuur rond surveyonderzoek wordt er veel aandacht besteed aan het design van een enquête en welke impact dit heeft op de participatie- of responsgraad. De respons varieert van survey tot survey en is sterk afhankelijk van de dataverzamelingsmode, de aard van de steekproefeenheden en de veldwerkprocedures (Lynn, 2008). Het biedt dan ook een meerwaarde om te achterhalen welke achterliggende motieven er zijn om niet te willen deelnemen. Onderzoek vanuit een sociaalpsychologische invalshoek verklaart waarom individuen wel of niet deelnemen aan een enquête. Volgens Dillman (1978) ligt de sociale ruiltheorie hier aan de grondslag: een persoon zal enkel participeren in geval de intrinsieke voordelen groter zijn dan de (sociale)kosten (Harrison, 2010). Groves and Singer (2000) introduceerden hiervoor de *leverage-salience* theorie, waarin de *leverage* een bepaald kenmerk van de enquête naar voren schuift dat aantrekkelijk is voor de respondent om deel te nemen (bijvoorbeeld de timing) en *salience* datgene is dat men extra benadrukt om iemand te doen overtuigen tot deelname (bijvoorbeeld door te benadrukken dat de vragenlijst niet te veel tijd vraagt). Bij wijze van voorbeeld kan men, met deze theorie in het achterhoofd, bijvoorbeeld voltijds tewerkgestelde personen contacteren na de werkuren en hen een online enquête aanbieden die ze op het werk, al pendelend, of thuis kunnen invullen. In de literatuur vinden we verscheidene experimenten die bestuderen in welke mate men de respons kan verhogen, zowel in eenmalige, cross-sectionele als longitudinale studies. In deze sectie omschrijven we enkele strategieën die de literatuur naar voren schuift om de responsgraad te verhogen.

Een eerste strategie is **het continue opvolgen van de respons**. De follow-up of opvolgen van de respons wordt door Dillman et al. (2014) als één van de meest doorslaggevende strategieën omschreven om de respons te verhogen. Voor de gezinsenquêtes werkte men een verzendschema uit, naar de principes van de *Total Design Method* van Dillman et al. (2014), om de respons op te volgen. Het verzenden van de vragenboekjes ging gepaard met een introductiebrief, een antwoordkaartje (waarop men kon

aangeven niet te willen deelnemen) en een retourenvelop. In week 1 stuurde men de steekproefpersonen deze bundel op (zending 1), na een week gevolgd door een herinnerings- en bedankkaartje (zending 2). Vier weken na de eerste zending, ging men opnieuw over tot het versturen van de vragenboekjes (zending 3) naar wie nog niet had deelgenomen en niet expliciet had geweigerd. Twee weken daarna stuurde men een herinneringskaartje (zending 4) (Audenaert & Stuyck, 2018a). Voor de Stadsmonitor 2017 hanteerde men eenzelfde communicatieplan van vier zendingen, met daar nog bij geteld een bedankkaartje en cadeaubon (geschenkbond van 6 euro) als vijfde zending (Agentschap Binnenlands Bestuur, 2018a). Schockaert et al. (2012) suggereert na te gaan wat de oorzaken van weigering en niet-bereiken zijn, en deze reduceren. Dit kan men doen door meer contactpogingen te ondernemen, contacteren op andere tijdstippen, en dergelijke. Schmeets and Van der Bie (2006) baseerden zich op twee face-to-face bevestigingen en zetten zo een strategie op bij een moeilijk te bereiken doelgroep, namelijk bij personen met een vreemde herkomst waarbij ze de poging tot contact van minimaal drie naar zes pogingen verhoogden. Deze maatregel in deze Nederlandse studie bleek een succes: de respons steeg met 10 procentpunt bij personen met een niet-Westerse herkomst en bij personen die in het buitenland geboren waren, en met 2 procentpunt voor de 'autochtone' respondenten. De meeste schriftelijke bevestigingen werken met herinneringsbrieven om weigeraars te overtuigen. In een paper van Heerwegh, Abts, and Loosveldt (2007) analyseerden ze de impact van extra contactpogingen, en bleek dat een hogere respons niet noodzakelijk resulteert in minder vertekening. Dit omdat degene die men nog steeds niet kon contacteren erg verschillende achtergrondkenmerken hebben in vergelijking met de personen die wel deelnemen. Het gaat dus niet enkel om het opvolgen van de respons om de algehele responsgraad te kunnen optrekken, maar ook met expliciet het oog op het versmallen van de kloof van profielen van respondenten en niet-respondenten.

Een tweede strategie en bekende methode om aangeschreven personen aan te zetten tot het participeren in de bevestiging, is het inzetten van **beloningen of incentives**. Het kan gaan over *cash*, een waardebon, 'het winnen van tickets eens men participeert', en dergelijke om personen aan te moedigen om deel te nemen. Deze beloning kan vóór de deelname beloofd en overhandigd worden aan de respondent, of kan na de deelname gedeeld worden. Het financieel stimuleren van respondenten om deel te nemen aan de enquête vinden we ook terug bij de Gezondheidsenquête in 2013. Gezien de budgettaire beperkingen, ging het om een zeer beperkte beloning, namelijk een waardebon van €10 per huishouden. In de introductiebrief werd hiervan melding gemaakt. In het methodologische rapport nuanceert men deze methodiek, doordat men niet kon inschatten welk effect deze beloning heeft op de respons van de aangeschreven personen (Demarest, Van der Heyden, Sarr, De Waleffe, & Meersseman, 2015). Ook in de Survey Samenleven in Diversiteit 2017 gaf men de respondenten na een succesvolle deelname een beloning ter waarde van €15, met daarbij een bedankbrief met een verwijzing naar de website waar de resultaten op gepubliceerd werden (Stuyck et al., 2018). Ten slotte maakt ook de EU-SILC-enquête gebruik van een vergoeding van €30. De onderzoekers beargumenteren dit expliciet te doen om (verdere) deelname van arme(re) huishoudens blijvend aan te moedigen.

In de literatuur vinden we verscheidene effecten van het inzetten van *incentives* op de responsgraad. Een meta-analytische studie van Mercer, Caporaso, Cantor, and Townsend (2015) bestudeerden de samenhang tussen financiële beloningen en de responsgraad in huishoudelijke enquêtes voor post-, telefoon- en face-to-face- interviews. Uit hun bevindingen blijkt dat er een sterke relatie is tussen de methode of *mode* en de timing wanneer men de beloning overhandigt. Dit werd bevestigd door enquêteurs van de EU-SILC, die in het onderzoek van Schockaert et al. (2012) waarschuwen dat het tijdsinterval tussen het afnemen van de bevestiging en het ontvangen van de vergoeding niet te groot

mag zijn en dat de overhandiging van de vergoeding niet al te bureaucratisch mag verlopen. Mercer en collega's beargumenteren dat de meest succesvolle manier in een postenquête de overhandiging van de financiële beloning is vooraleer men deelneemt aan de enquête. Eerder onderzoek, specifiek gericht naar het verhogen van de responsgraad in een postenquête, beargumenteert eveneens dat de *prepaid incentives* resulteren in een hogere responsgraad. Aanvullend stelt men ook dat hogere beloningen meer respons in de hand werkt (Church, 1993; Singer & Ye, 2012). Men vond geen significante resultaten over het beloven voor en overhandigen na deelname, noch bewijs over hoe hoog de beloning dan wel zou moeten zijn (Mercer et al., 2015).

Een derde strategie wijst erop dat het voorzien van financiële beloningen niet de enige mogelijkheid is om respondenten aan te moedigen tot deelname. Onderzoekers kunnen ook inzetten op het verhogen van de **intrinsieke motivatie** van de steekproefpersonen om te participeren aan de survey (Singer & Ye, 2012). Het belang van extra aandacht naar deze niet-financiële incentives werd door Couper (2011) als volgt verwoord: *“Survey participation has long been viewed as an act of social reciprocation, not purely an act of selfless civic duty on the one hand, or of economic exchange on the other. Focusing solely on the benefits that may accrue to us as survey researchers has meant that we have increasingly ignored the things that may be important to respondents. The very act of taking the time and effort to convince potential respondents to participate in surveys helps communicate to them the value of such surveys, in ways that automated methods do not.”* In geval van de gezinsenquête verwerkte men dit in de introductiebrief, door te vermelden *“Wij willen onderzoeken hoe Vlaamse gezinnen er vandaag uit zien, hoe ouders hun kinderen opvoeden en hoe ze de zorg voor de kinderen combineren met hun werk. We willen ook graag weten wat gezinnen daarbij moeilijk vinden. Alleen zo kan Vlaanderen een sterk gezinsbeleid ontwikkelen dat goed aansluit bij wat Vlaamse gezinnen écht nodig hebben.”* Men benadrukt hierbij dat de gezinnen op die manier een steentje bijdragen in het gezinsbeleid, en het belang van informatie om het gezinsbeleid op maat af te stemmen. Voor weigeraars kan een extra herinnering aan deze meerwaarde hen toch over de brug halen om deel te nemen.

Een vierde strategie is **pre- en/of post- evaluatie** of testen van de gehanteerde vragenlijst. Het pretesten van een vragenlijst is essentieel om mogelijke drempels te identificeren die de dataverzameling tegengaan. Zoals eerder vermeld in Hoofdstuk 3.1 legde men een eerste draft van de vragenlijst voor aan 40 vrijwillige proefpersonen. Het testen van de vragenlijst wordt in de survey literatuur eveneens expliciet naar voren geschoven als een strategie om de respons te verhogen. In geval men een representatieve groep neemt die zowel het inhoudelijke als het visuele aspect beoordeelt, zorgt men op voorhand –op goedkope wijze- ervoor dat potentieel onzichtbare barrières worden weggewerkt. Het betrekken van kwetsbare gezinnen in deze evaluatie zorgt voor een implementatie van hun beoordeling in de finale versie. Dit pretesten van de vragenlijst gebeurt volgens Harkness et al. (2014) zelden op vertaalde vragenlijsten. Het opzetten van de vragenlijsten in meerdere talen zou volgens Harkness en collega's gebeuren door een externe actor, waarna het verkennen van mogelijke taalbarrières of taalfouten overkeken kan worden. Het testen van meertalige vragenlijsten (zoals deze in de extra steekproef van de gezinsenquête) kan bijgevolg nuttig zijn wanneer men een representatieve staal van de kwetsbare gezinnen wil bereiken.

Naast het evalueren van de vragenlijst via de pilootstudie, kan men naar het voorbeeld van de Survey Samenleven in Diversiteit inzetten op evaluatie na deelname aan de finale versie. In deze enquête gaf men extra aandacht aan de evaluatie door elke respondent de mogelijkheid te geven zijn of haar mening te formuleren over het onderzoek via een online vragenlijst. De respondent werd op de hoogte gebracht van deze optie via de bedankbrief die werd overhandigd na de deelname. De respondenten kregen ook

de optie om het telefoonnummer achter te laten waardoor een telefonisch evaluatiegesprek kon plaatsvinden. Bijna de helft (48%) van de respondenten die de bevraging aflegde via een *computer-assisted personal interview* (CAPI) maakte tijd voor een evaluatiegesprek. De online versie van het evaluatiegesprek was minder succesvol (Stuyck et al., 2018), wat impliceert dat men ook hier moet durven experimenteren in (de combinatie van) verschillende methodieken om de respons van de steekproefpersonen aan te moedigen.

Ten vierde kan men in het design meer de nadruk leggen op *tailoring* of *targeting*. Het afnemen van een enquête om data te verzamelen is in se een uitdaging: er zijn altijd specifieke groepen of personen met een gekende, lage participatiegraad. De vraag die we in dit geval als startpunt nemen, is: 'Slagen we erin om kwetsbare gezinnen te bereiken met deze steekproeftrekking?'. In de literatuur vinden we twee manieren om het steekproefdesign aan te passen: *tailoring* of *targeting*. Terwijl *tailoring* betekent dat men zich aanpast aan het optimaliseren van het bereiken van een *individu*, is *targeting* het aangepast bereiken van een *groep* (Haan & Ongena, 2014; Kreuter, 2003). Wij zijn in dit rapport voornamelijk geïnteresseerd in het *targeten* of doelgericht bereiken van kwetsbare gezinnen.

In het surveyonderzoek zien we de laatste jaren een shift van een volledig gestandaardiseerde procedure om een steekproef op te zetten, naar meer aangepaste versies. In de literatuur spreekt men van *adaptive survey designs* (Schouten, Calinescu, & Luiten, 2013). Een bekend voorbeeld is het aanpassen van de uitnodiging tot deelname aan de enquête door de lay-out, inhoud, taalgebruik, het toevoegen van een beloning, of dergelijke aan te passen aan de groep waar men zich naar richt. Verschillende interviewers worden vandaag getraind om deze eerste contactopname op maat te brengen in telefonische of persoonlijke interviews (Lynn, 2017). In het opzetten van een enquête neigen onderzoekers ernaar om alle steekproefeenheden gelijk te behandelen. Zo gaat men bijvoorbeeld reflecteren over een design van een uitnodiging die een breed publiek aanspreekt, waarna men ditzelfde exemplaar van uitnodiging doorstuurt naar alle geselecteerde personen. Toch reageren sommige personen of groepen verschillend op dit specifieke design (dit kan de inhoud zijn van de uitnodiging, maar ook de lay-out, het taalgebruik, etc). De beste strategie om personen te bereiken is in dat geval om over te gaan tot het *targeten* of aanpassen gericht op bepaalde bevolkingsgroepen. Een alternatief voor gestandaardiseerde procedures is het verschillend behandelen van deze segmenten, afhankelijk van wat we weten hoe we de participatie in deze groep kunnen verhogen (Lynn, 2013). Een concreet Vlaams voorbeeld wordt toegepast in het MISSION-project in Kortrijk, waar men kansarme gezinnen tracht te bereiken via verschillende manieren: sms, telefoon, Whatsapp, deurbel, en dergelijke (Van Lancker et al., 2020).

Maar het gaat niet enkel om het doelbewust aanpassen van de manier waarop men personen bereikt of overtuigt. Men kan ook de vragenlijst aanpassen: afstappen van een gestandaardiseerde vragenlijst (i.e. een vragenlijst die aan elk gezin dezelfde vragen stelt) en overstappen naar meer aangepaste alternatieven waarbij respondenten als individu of als groep een verschillende vragenlijst verkrijgen. Het aanpassen van de vragenlijst naar de steekproefpersonen blijft zeer *niche* in de literatuur. De meerderheid van de aspecten van het design van een enquête en de uitvoering blijven grotendeels gestandaardiseerd. Een bulk aan literatuur toont echter aan dat er wel degelijk positieve respons-effecten zijn wanneer men het design aanpast. Groves, Cialdini, and Couper (1992) waarschuwde begin jaren '90 al voor het blindelings toepassen van universele designs. De onderzoekers stelden voor om aangepaste technieken toe te passen, afhankelijk van je doelpopulatie. Een voorbeeld van een studie die een aangepaste vragenlijst hanteerde, is de extra steekproef van kansarme gezinnen die men opnam in het VLAS-onderzoek. Terwijl de basissteekproef van het JOnG!-onderzoek meer vragen kreeg met

betrekking tot bijvoorbeeld zwangerschap, werden enkele vragen voor de extra steekproef geschrapt om plaats te maken voor meer vragen omtrent de sociale context en kwetsbaarheid van het gezin. Ook extreem gevoelige vragen werden voor deze doelgroep geschrapt (Lemmens & Groenez, 2015a). In het onderzoek naar een 'verborgen groep armen' in de EU-SILC, concludeerden Schockaert et al. (2012) dat non respons bij deze kwetsbare doelgroep op preventieve wijze voorkomen kan worden door de vragenlijsten te vereenvoudigen. Een respondent de mogelijkheid geven om een vraag over te slaan wanneer hij of zij moeilijkheden ondervindt om specifieke vragen te beantwoorden, zou unit non respons tegen werken. In Schockaert et al. (2012) zette men interviewers in om deelname aan te moedigen bij de steekproefeenheden. Het verder verfijnen van de opleiding van deze interviewers, voornamelijk door verder in te gaan in inkomensgerelateerde vragen, zou de respons verhogen. Ten slotte bleek het ook een meerwaarde om interviewers met ervaring in te zetten om zijn of haar collega's op te leiden. Hun ervaring met het anticiperen op non respons, hun stijl van zichzelf te presenteren en manier van communiceren, kan een adequate strategie zijn om kwetsbare doelgroepen te bereiken. Een alternatieve manier om de respons te verhogen bij kwetsbare bevolkingssegmenten is door samenwerkingsverbanden aan te sluiten met middenveldorganisaties die in contact staan met deze doelgroep. Onderzoek toont aan dat het werken met een tussenpersoon die gekend is in de betrokken gemeenschap, het vertrouwen in het onderzoek en dus ook de respons doet verhogen (Bonevski et al., 2014).

Let wel, de keuze van het steekproefdesign wordt volgens Kalton (2014) mede bepaald door de zeldzaamheid van de doelpopulatie. Wanneer een doelgroep verwaarloosbaar klein is, kan men bij een wijziging in design eerder fouten in de hand werken, in plaats van deze op te lossen (Schepers et al., 2017). Een *targeting* strategie gaat dan niet noodzakelijk over het volledig omgooien van het gebruikte design, maar over het maken van methodologische keuzes op basis van inhoudelijke argumenten. Zo nam men in de gezinsenquête de beslissing om, op vraag van het Agentschap Binnenlands Bestuur, een extra steekproef te trekken bij gezinnen met een niet-EU-herkomst. De laatste jaren nam het aandeel van het gezinnen met een buitenlandse herkomst gestaag toe, waardoor een extra steekproef ervoor kan zorgen dat deze gezinnen voldoende vertegenwoordigd worden in de gezinsenquête. Deze groep werd anders benaderd als de basissteekproef: de vragenlijst werd ingekort en werden, naast het Nederlands, ook in het Engels en Frans aangeboden.

Een vijfde strategie om de respons te verhogen, is aandacht te schenken aan de **taal** van de doelpopulatie. Om de vragenlijst volledig toegankelijk te maken voor alle gezinnen, is het belangrijk om genoeg aandacht te geven aan de aanwezige talen in de doelpopulatie. Het beschikbaar maken van boekjes in allerlei talen kan een manier zijn om een breder bereik te verkrijgen in je steekproef. Hierbij moet wel steeds de afweging worden gemaakt welke taal of talen een meerwaarde kunnen bieden wat betreft de respons. Conform de regelgeving, werden de vragenlijsten van de gezinsenquête uitsluitend in het Nederlands opgesteld en verstuurd. Doordat personen met een herkomst buiten de EU de taal niet steeds voldoende machtig zijn, werd bij de uitbreidende steekproef gebruik gemaakt van een Frans- en Engelstalige versie. Toch kan men een andere uitbreiding in talen overwegen, door in te spelen op de realiteit van de diversiteit in talen in de Belgische samenleving. Het gaat niet enkel om de taal zelf, maar eveneens om het (vaak complex) taalgebruik. Het VLAS- onderzoek dat de ontwikkeling en opvoeding bij kwetsbare gezinnen met jonge kinderen op longitudinale wijze bestudeert, zag er expliciet op toe dat er geen abstract taalgebruik werd gehanteerd. De vragen werden op een meer informele en duidelijke manier genoteerd (Lemmens & Groenez, 2015a).

Een alternatieve strategie om de respons te verhogen is door preventief **reservehuishoudens** selecteren. Een concrete uitwerking van deze tactiek vindt men in de Belgische Gezondheidsenquête,

waar men het probleem van weigerende of niet te contacteren huishoudens opvangt door reserve-huishoudens op te nemen in het design. Voor elk initieel geselecteerd huishouden werden drie reservehuishoudens geselecteerd, die dezelfde huishoudgrootte, dezelfde leeftijd van de referentiepersoon en een gelijke statistische sector kennen. Op deze manier vormt men een 'cluster' van vier huishoudens. Wanneer het eerste huishouden weigert om deel te nemen, of het kan niet gecontacteerd worden, gaat men deze vervangen door een volgend huishouden in de cluster. Wanneer al de huishoudens in de cluster zijn uitgeput, gaat men een nieuwe (vervang-)cluster activeren, die niet meer dezelfde karakteristieken bevat. Voor deze nieuwe cluster wordt een identieke procedure toegepast in geval van weigerende of niet contacteerbare huishoudens (Demarest et al., 2015). Dit is een interessante benadering wanneer men inzet op het reduceren van de variatie in achtergrondkenmerken tussen respondenten en niet-respondenten. Hier houdt men zich niet enkel bezig met het reduceren van het aantal weigerende huishoudens, maar ook met het bereiken van een voldoende hoog aantal personen of huishoudens met specifieke kenmerken.

Een laatste strategie die we eerder hebben besproken in Hoofdstuk 2 is het **koppelen van data**. Wanneer men enquêtegegevens koppelt met administratieve gegevens kan men bijvoorbeeld de keuze maken om de vragenlijst in te korten. Alle antwoordmogelijkheden die men kan achterhalen uit administratieve gegevens laat men dan weg, tenzij men een extra controle wil op enkele items (zoals bijvoorbeeld de feitelijke woonsituatie). Dit kan er ook voor zorgen dat je enkele meer complexe vragen niet moet stellen die als te complex of gevoelig worden ervaren door de respondenten, denk maar aan de vraag wat betreft de inkomens.

Om af te sluiten, benadrukken we nog even **een kritische kanttekening** wanneer we praten over respons. Instinctief lijkt een hoge respons de meest cruciale factor om de selectiviteit in de respons terug te brengen. Toch blijkt uit onderzoek dat het generiek ophogen van de responsgraad dit niet noodzakelijk als gevolg heeft. Door enkel te focussen op het verhogen van de responsgraad voor de gehele steekproef, zou men niet noodzakelijk meer informatie verkrijgen over deze doelgroep. Indien men een selectieve respons heeft, wordt daarom aanbevolen om intensief in te zetten op het benaderen van de groepen die structureel ondervertegenwoordigd zijn (Beukenhorst, 1999, 2001; Te Riele, 2002). Beullens and Loosveldt (2012) waarschuwen eveneens voor de eenzijdige focus op het verhogen van de responsgraad. Wanneer het behalen van een hoge respons het doel wordt, gaat men voorbij aan de onderliggende prioriteit om *bias* te reduceren. Er zijn verschillende redenen waarom onderzoekers zich eenzijdig toespitsen op het maximaliseren van deze respons. Volgens Beullens & Loosveldt zou men zich liever toespitsen op de respons, omdat deze gemakkelijker te rapporteren en berekenen is. Het risico op vertekening daarentegen is een zeer complex gegeven. Ook zij stellen voor om de eigen data aan te vullen met andere data (zoals administratieve gegevens) om het verschil tussen respondenten en niet-respondenten te bekijken. Een laatste suggestie die de auteurs meegeven is het doelgericht focussen op subgroepen waarvan men de kans op weigering op voorhand al hoog inschatte, om dan meerdere contactpogingen te ondernemen, te wisselen of combineren van dataverzamelmethode, enzovoort (Beullens & Loosveldt, 2012).

6 Kritische kanttekening bij administratieve gegevens

In dit rapport maken we gebruik van een koppeling tussen enquêtegegevens en administratieve gegevens (DWH AM&SB). Dat levert heel wat inzicht in de bruikbaarheid van dergelijke koppeling om de volgende golven van de gezinsenquête mee te ondersteunen (zoals we beschrijven in scenario's 3 tot en met 5 in sectie 2.4). Algemeen stellen we vast dat het hanteren van administratieve gegevens om de Gezinsenquête te verrijken op dit moment nog heel wat problemen kent.

Ten eerste was het niet mogelijk om in de eerste golf van de gezinsenquête een **imputatie** uit te voeren om enkele ontbrekende items in te vullen met beschikbare administratieve gegevens. Item non respons zorgt namelijk voor incomplete gegevens, waardoor analyses van enkel de complete observaties leidt tot verlies van informatie. Deze gegevens kunnen aangevuld worden door middel van imputatie, een complexe techniek die ontbrekende gegevens in de dataset opvult. Doordat we administratieve gegevens verkregen hebben vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming zou het in theorie mogelijk kunnen zijn om gegevens die men niet verkreeg in de enquête aan te vullen met administratieve gegevens. Dit blijkt echter niet mogelijk omdat de administratieve gegevens vaak een administratieve logica en dito operationalisering kennen die niet noodzakelijk gelijkloopt met de gegevens die bevestigd werden in de gezinsenquête. We illustreren dit met een voorbeeld. De gezinsenquête bevestigt haar respondenten naar de netto beschikbare gezinsinkomens op maandbasis, voor een brede waaier aan bronnen (wedden of lonen; netto-bedrijfsinkomen voor zelfstandigen; uitkering voor werkloosheid of activering; uitkering voor ziekte, invaliditeit of ongeval; leefloon; kinderbijslag; uitkering voor loopbaanonderbreking of tijdskrediet; pensioenen; alimentatiegelden; financiële steun van externen; huuropbrengsten; interesten; andere). Dit inkomen is het inkomen na afhouding van sociale bijdragen en belastingen en is het eigenlijk het *besteedbaar* inkomen. In enquêteonderzoek (en zeker in postenquêtes) maakt men vaak gebruik van dit concept, omdat het gemakkelijker is voor gezinnen om na te gaan welk bedrag ze op het eind van de maand krijgen (eerder dan dat men bruto (belastbare) inkomens bevestigt). Het Datawarehouse Arbeidsmarkt en Sociale Bescherming omvat (althans op het moment van schrijven) enkel bruto belastbare gezinsinkomen op jaarbasis. Het is, met andere woorden, niet mogelijk om de onbeantwoorde inkomensvragen in de enquête aan te vullen met administratieve inkomensgegevens die we kunnen verkrijgen via het Datawarehouse, omdat de inkomensconcepten sterk verschillen.

Ten tweede maakt dit verschil in operationalisering en conceptualisering tussen de administratieve gegevens en gezinsenquête gegevens **vergelijking** moeilijk tot onmogelijk. Het was, bijvoorbeeld, niet mogelijk om de vraag naar de inkomens en de oorsprong hiervan in de gezinsenquête te vergelijken met gekoppelde inkomensgegevens vanuit het Datawarehouse. Het is perfect mogelijk voor een respondent om de maand voordien geen leefloon te ontvangen, en dit dus niet aan te geven in de gezinsenquête, maar op jaarbasis wel een leefloon te hebben ontvangen. Wat de respondent dan aangeeft in de Gezinsenquête is dan niet fout, maar de interpretatie van het inkomensstatuut of de inkomenssituatie van deze respondent zal anders zijn naarmate men de administratieve of enquêtegegevens gebruikt. Een ander voorbeeld, zoals eerder aangehaald in 4.2, is de vraag naar werk. In de gezinsenquête peilt men naar het huidige beroepsstatuut, met daarbij zes antwoordopties: arbeider/arbeidster (private sector), bediende (private sector), contractueel of statutair ambtenaar (openbare sector), zelfstandige/vrij beroep zonder personeel, zelfstandige/vrij beroep met personeel, helper of meewerkend familielid. Indien de respondent meerdere jobs combineert, wordt in de gezinsenquête gestipuleerd om 'de voornaamste job' te kiezen. Een variabele uit de administratieve gegevens die het dichtst bij dit concept aanleunt is de nomenclatuur van socio-economische positie. Hierbij worden er

meerdere opties met allerlei specificaties toegekend, gaande van werkend in één job in loondienst, werkend in meerdere jobs in loondienst, werkend als zelfstandige in hoofdberoep, werkend als zelfstandige in bijberoep, etc. Of men al dan niet personeel ter beschikking heeft als zelfstandige, wordt niet vermeld in Datawarehouse gegevens. Maar zelfs al men alle zelfstandigen samenneemt, blijft een vergelijking moeilijk. De nomenclatuur wordt toegekend op de laatste dag van het kwartaal op basis van de informatie uit vier instellingen (RSZ, RSZPPO, RSVZ en RVA). Bovendien krijgt elk individu slechts één nomenclatuurpositie toegekend: wanneer een persoon meerdere posities invult, wordt deze toegewezen aan de nomenclatuurpositie die het dichtst aansluit bij de arbeidsmarkt²⁶. Concreet betekent dit dat gepensioneerden met een kleine deeltijdse job bij de ‘werkenden’ wordt ingedeeld, en niet bij de gepensioneerden. Een laatste hiaat in deze administratieve gegevens is dat er nog steeds beperkte informatie beschikbaar is, bijvoorbeeld wat betreft de nomenclatuur van de socio-economische positie van de respondenten. Zoals we eerder zagen in 4.2.2 behoren 8,5% van de basis geselecteerde basissteekproef en 22,8% van de extra geselecteerde steekproef tot een restcategorie. De beperkte informatie over arbeidsmarkt gerelateerde kenmerken wordt eveneens duidelijk wanneer we terugblikken naar het contracttype, waarbij 32,7% van de geselecteerde basissteekproef en 50,2% van de extra geselecteerde steekproef geen contract of geen informatie over een contract hebben. Een mogelijke verklaring van deze incomplete dataset kan zijn dat dit kenmerk zeer gefaseerd wordt ingevuld vanuit vier verschillende instellingen, waardoor dit vertraging kan oplopen (Datawarehouse Arbeidsmarkt en Sociale Bescherming, z.j.-a).

Kortom, zoals eerder theoretisch besproken in 2.4 blijken de scenario’s waarin administratieve gegevens de enquêtegegevens verrijken in de praktijk op verscheidene beperkingen te botsen. Volgend op de verschillende valkuilen die we in dat hoofdstuk hebben besproken, stellen we vast dat het invullen van item non respons in de dataset met administratieve gegevens vandaag zeer moeilijk is. Zowel de gezinsenquête als het Datawarehouse hanteert een eigen begrippenkader en operationaliseren de variabelen op een andere manier, wat een vergelijking en imputatie moeilijk maakt.

In toekomstige golven van de gezinsenquête kan er natuurlijk wel overwogen worden om bepaalde kenmerken van gezinnen uit de administratieve data te halen zodat die niet meer bevestigd moeten worden in de enquête zelf. Het probleem is dan dat een vergelijking doorheen de tijd niet meer mogelijk is, omdat de concepten niet vergelijkbaar zijn. Bovendien zit er steeds een zekere vertraging op de dataverzameling van het Datawarehouse omdat de gegevens vanuit verscheidene instellingen afkomstig zijn. De winst die men zo kan boeken door een koppeling met administratieve data te maken moet dus worden afgewogen tegenover het verlies aan informatie wanneer opeenvolgende gezinsenquêtes niet vergelijkbaar zijn over de tijd heen.

Samenvattend

In dit hoofdstuk analyseren we de respons van de gezinsenquête, de representativiteit van de gerealiseerde steekproef en de item non respons. Op beschrijvende wijze rapporteren we de gegevens van verschillende individuele- en gezinskenmerken die we verkregen uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming. We vergeleken deze kenmerken door volgende steekproeven naast elkaar te leggen: (1) de gerealiseerde *basissteekproef*, (2) de gerealiseerde *extra*

²⁶ Meer informatie over de nomenclatuur van de socio-economische positie uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming via www.ksz-bcss.fgov.be

steekproef, (3) de geselecteerde *basissteekproef*, (4) de geselecteerde *extra steekproef*, (5) de totaal gerealiseerde steekproef (of 'De gezinsenquête') en waar mogelijk (6) de populatiegegevens.

Een belangrijke vaststelling is het verschil in de verdeling van de achtergrondkenmerken van respondenten in de basis en in de extra steekproef. De steekproefpersonen in de extra steekproef zijn jonger, vrouwelijk, laag- of midden geschoold en kennen een zeer diverse herkomst. Bovendien zijn het voornamelijk grote gezinnen die gehuwd zijn, met lage bruto belastbare gezinsinkomens en een lage tot zeer lage werkintensiteit. De basissteekproef ziet er geheel anders uit: we zien hier voornamelijk mannen én vrouwen, die ouder zijn dan de extra steekproef, midden- of hoogopgeleid, met een minder diverse herkomst. Daaropvolgend gaat het over kleinere gezinnen met een hoger bruto belastbaar gezinsinkomen in vergelijking met deze van de extra steekproef, en daarbij aansluitend een hogere werkintensiteit. Toch blijkt deze gerealiseerde *basissteekproef* amper af te wijken van de *basisselectie*. Daar het om een *aselecte* steekproeftrekking ging, concluderen we dat de lage responsgraad van de gezinsenquête de representativiteit eigenlijk niet heeft geschaad. Toch zien we dat gezinnen met bepaalde socio-economische kenmerken ondervertegenwoordigd zijn, zoals gezinnen die een (equivalent) leefloon ontvangen of personen met een lage scholingsgraad. Voor deze gezinnen was de extra steekproef nuttig om voldoende vertegenwoordigd te worden in de gezinsenquête. Zoals eerder vermeld in Hoofdstuk 3.4 maakt ook dit hoofdstuk echter duidelijk dat, hoewel de extra steekproef zorgt voor een betere vertegenwoordiging van gezinnen met een aantal specifieke kenmerken, deze tegelijkertijd zorgt voor een vertekening van de totaal gerealiseerde steekproef, waardoor deze geen representatieve staal meer vormt voor de geselecteerde basissteekproef.

Daarbovenop stellen we vast dat de item non respons van de gezinsenquête relatief laag ligt voor enkele geselecteerde items (jaarlijkse vakantie, het kunnen rondkomen, hoogst behaalde diploma en beschikbare inkomen). Opvallend is het aandeel van de 'sterkere' profielen bij het niet beantwoorden van deze vragen: vooral personen uit een werkrijk gezin beslisten om de vraag open te laten. Desalniettemin zien we een vertekening in item non respons als we het opleidingsniveau en het inkomen onder de loep nemen: het zijn vooral meer laagopgeleiden die de vragen niet beantwoorden enerzijds, en lage inkomensgezinnen die de inkomensvraag onbeantwoord laten anderzijds. De respons voor de gezinsenquête bleef evenwel relatief laag. Enkele strategieën die hier worden benoemd om dit cijfer omhoog te halen, zijn: (1) het continue opvolgen van de respons, (2) het inzetten van *incentives*, (3) het testen voor en evalueren na deelname, (4) het meer doelgericht bereiken of *targeten* van bepaalde doelgroepen door de communicatie of de vragenlijst zelf meer op maat te maken, (5) te reflecteren over het taalaspect (niet enkel met betrekking tot de taal zelf, maar ook over het taalgebruik), (6) het preventief selecteren van reservehuishoudens en (7) het koppelen van (administratieve) data.

We leren ook een belangrijke les voor de organisatie van volgende golven van de gezinsenquête: een koppeling met administratieve gegevens is niet zonder problemen, omdat de administratieve concepten en operationalisering van de variabelen niet noodzakelijk strookt met de concepten en operationalisering van de variabelen in de gezinsenquête. Als vergelijkbaarheid over de tijd heen belangrijk is, dan lijkt een koppeling met administratieve gegevens om een aantal gegevens van respondenten vooraf te laten invullen voorlopig weinig realistisch.

Hoofdstuk 5

Gewichten

De gezinsenquête beoogt uitspraken te kunnen doen over de doelpopulatie, namelijk de Nederlandstalige gezinnen in Vlaanderen en het Brussels Hoofdstedelijke Gewest. Doordat men slechts een deel van de populatie bevraagt, treden er enkele afwijkingen op waar men rekening mee moet houden. Enerzijds kan steekproefvariabiliteit optreden, wat betekent dat men met dezelfde steekproeftrekking een andere steekproef aan respondenten kan bekomen die tot andere resultaten leidt. Deze steekproefvariabiliteit neemt af bij het optrekken van de steekproefgrootte. Anderzijds kan bias of vertekening optreden, als gevolg van fouten in de steekproeftrekking of onderzoeksopzet. Dit betekent dan weer dat men steeds dezelfde fouten in schattingen verkrijgt bij het trekken van een nieuwe steekproef. De vertekening is dan weer niet zozeer een gegeven dat afneemt naarmate de steekproef groter wordt (Pickery, 2010a). Met het construeren en gebruiken van gewichten willen we in feite systematische fouten in de gezinsenquête zo klein mogelijk houden. Met andere woorden, we willen de vertekening reduceren die het gevolg is van een dekkingsfout, non-respons of van ongelijke selectiekansen (Pickery, 2010a).

1 Construeren van gewichten

Het construeren van deze gewichten vindt plaats na het verzamelen en verwerken van de data, maar voor de data-analyse. Op aanvraag van het Departement WVG construeerde de Studiedienst van de Vlaamse Regering twee basisgewichten: weeg 1 en weeg 2 op de verdeling geslacht en leeftijd. Herkomst wordt apart meegenomen in de berekening. Het eerste gewicht (*weeg 1*) gebruikt men bij analyse van de gehele groep respondenten ($n = 2.683$), terwijl het tweede gewicht (*weeg 2*) enkel wordt gebruikt bij vragen die niet aan de uitbreidende steekproef bij niet-EU-burgers werd gesteld ($n = 2.189$). Dit impliceert dat de tweede weegcoëfficiënt respondenten met een herkomst buiten de Europese Unie een hoger gewicht geeft dan de eerste weegcoëfficiënt. De gewichten werden berekend op basis van de gekende populatieverdelingen. De vuistregel die hier werd gehanteerd om stabiele weegcoëfficiënten te construeren, is dat het aantal respondenten binnen elke categorie niet kleiner mag zijn dan tien. Diezelfde vuistregel pasten we eveneens toe op onze weegcoëfficiënten. In wat volgt bespreken we de poststratificatiegewichten die wij construeerden op basis van verkregen populatiegegevens.

Hoe construeren we dit gewicht? Concreet betekent dit dat we een nieuwe variabele, w_i , creëren voor elke respondent (i) in de steekproef. Het gewicht kan men interpreteren als het aantal individuen in de populatie die worden vertegenwoordigd door de respondent in de steekproef. In geval men een weegcoëfficiënt of gewicht van 100 bekomt, betekent dit dat de respondent zichzelf en 99 andere personen van de doelpopulatie vertegenwoordigt (Biemer & Christ, 2008). Om dit te bewerkstelligen, deden we een aanvraag bij het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM & SB) in de vorm van twee kruistabellen (zie tabel 19 en 20). Deze twee kruistabellen geven de populatieproporties weer voor een aantal kenmerken. Een eerste kruistabel (Tabel 19) vraagt een combinatie van het aantal gezinnen met kinderen jonger dan 25 jaar naar het type huishouden, het aantal kinderen in dit gezin per leeftijdsklasse en de herkomst.

Tabel 19 Kruistabel 1: Aantal gezinnen met kinderen jonger dan 25 jaar naar huishoudtype, aantal kinderen per leeftijdsklasse en herkomst in Vlaanderen.

Combinatie-variabelen	Type huishouden	Aantal kinderen in het gezin	Herkomst
Waarden	1: (On)gehuwd paar met kinderen	Per leeftijdsklasse: 0-11 jaar 12-24 jaar	1: Belg 2: EU, geen Belg 3: Niet-EU

In een tweede kruistabel (Tabel 20) vroegen we populatiegegevens voor diezelfde huishoudtypes, maar dit keer in combinatie met mogelijke beroepsstatuten.

Tabel 20 Kruistabel 2: Aantal gezinnen met kinderen jonger dan 25 jaar naar huishoudtype en beroepsstatuut in Vlaanderen

Combinatie-variabelen	Type huishouden	Beroepsstatuut
Waarden	1: (On)gehuwd paar met kinderen 2: Eenoudergezinnen	1: Andere 2: Arbeidsongeschikt 3: Werkloos 4: Pensioen 5: Zelfstandige 6: Werknemer 7: Onbekend

We kozen voor deze variabelen omdat we –naast leeftijd, geslacht en herkomst- deze variabelen aanschouwen als mede doorslaggevende kenmerken voor specifiek kwetsbare gezinnen. Zoals uitgebreid wordt besproken in het inhoudelijk rapport, wijst een brede waaier aan literatuur ons op het gegeven dat eenoudergezinnen en voornamelijk alleenstaande moeders een kwetsbare doelgroep is. Hun positie is vaak precair, mede vanwege achterstellingen die blijven cumuleren bij deze gezinnen. Verwacht wordt bovendien dat de omvang van deze groep zal toenemen, waardoor deze de nodige onderzoeks- en beleidsaandacht verdient. In 2013 bedroeg de werkzaamheid bij alleenstaande moeders op actieve leeftijd slechts 62% (Van Peer, 2016). De combinatie met de variabele ‘beroepsstatuut’ is dus eveneens belangrijk om te betrekken in onze gewichten.

Wat een poststratificatiegewicht dan doet is de proportie van een specifieke variabele of combinatie van variabelen in de populatie afzetten tegenover de proportie die men uit de steekproef haalt. Een voorbeeld hiervan is de verdeling van geslacht in de populatie: stel dat 50% van de populatie bestaat uit mannen en 50% uit vrouwen. In de steekproef bestaat deze nochtans uit 60% van het mannelijk geslacht en 40% van het vrouwelijke geslacht. Wat dit gewicht dan doet, is de populatieproportie delen door deze van de steekproef. Voor mannen doen we dan $0.5/0.6$, wat een gewicht oplevert van 0.833 . Vrouwen krijgen dan een gewicht van $0.5/0.4$, wat 1.25 betekent. Omdat vrouwen minder vertegenwoordigd zijn in de steekproef, krijgen ze dus een groter gewicht om tegemoet te komen aan de realiteit (Johnson, 2008). Ditzelfde doen we door in dit hoofdstuk enkele variabelen te combineren. Zoals eerder vermeld, vroegen we de populatiegegevens op voor enkele combinaties van kenmerken. Let wel, het aantal gevallen per cel in de steekproef mag niet te klein worden (Johnson, 2008). Gelet op de grootte

van onze steekproef en de verschillende combinaties die we opvroegen, hebben we hier enkele combinaties moeten samenvoegen. Ook in onze gewichten pasten we dus de vuistregel toe (i.e. het aantal respondenten binnen elke categorie mag niet kleiner zijn dan tien).

Tabel 21 Voorbeeld combinatie bij het construeren van gewichten.

Gezins- type	Aantal kind- eren 0-11	Aantal kind- eren 12-24	Her- komst	Totaal gezinnen in GE	Coëf- ficiënt GE	Totaal gezinnen in de populatie (KSZ)	Coëfficiënt populatie (KSZ)	Gewicht
Alleen- staand	0	1	Belg	123	0,047	48120	0,053	1,13

Allereerst geven we een hypothetisch voorbeeld van hoe dit gewicht wordt geconstrueerd (Tabel 21). Hier bekijken we bijvoorbeeld het aantal gezinnen die alleenstaande ouders zijn, geen kind hebben tussen 0 en 11 jaar, één kind hebben tussen 12 en 24 jaar en van Belgische herkomst zijn. In de gezinsenquête zijn er 123 gezinnen die deze specifieke kenmerken hebben, wat een coëfficiënt van 0,047 oplevert eens we deze in verhouding plaatsen met het totale aantal gezinnen. Ditzelfde doen we voor de populatiegegevens. Hieruit blijkt dat we 48.120 eenoudergezinnen met Belgische herkomst en één kind tussen 12 en 24 jaar kennen in de populatie, wat neerkomt op een coëfficiënt van 0,053. Vervolgens delen we de coëfficiënt van de populatie met deze van de steekproef, en bekomen we het gewicht. In deze specifieke combinatie is dat 1,13. Concreet betekent dit dat de eenoudergezinnen die aan deze voorwaarden voldoen in de gezinsenquête licht ondervertegenwoordigd zijn in de gezinsenquête, en daardoor een hoger gewicht krijgen toegeschreven opdat we correcte uitspraken kunnen doen over deze gezinnen in Vlaanderen.

Hoewel weegcoëfficiënten ingezet kunnen worden om systematische fouten tegen te gaan, botsen we ook op enkele beperkingen inherent aan gewichten. Gewichten zijn voornamelijk te hanteren bij descriptieve analyses, maar kunnen een avers effect hebben op standaardfouten en inferentiële gegevens. Weegcoëfficiënten zullen nagenoeg altijd de standaardfouten van de schattingen doen verhogen en hierdoor instabiliteit teweegbrengen in de betrokken data. Om deze instabiliteit tegen te gaan, worden de gewichten soms *getrimd*, zodat de extreem hoge gewichten ingeperkt worden. Dit 'trimmen' zorgt dan weer voor het afnemen van de representativiteit van deze gewogen data. Het construeren van gewichten is hierdoor steeds een *trade off* tussen het representatief maken van de data en het reduceren van de instabiliteit (Chromy & Abeyasekera, 2005; Johnson, 2008).

Een andere kanttekening specifiek voor deze gewichten is het feit dat we enkel populatiegegevens verkregen voor de Vlaamse gezinnen en de gezinnen in het Brussels Hoofdstedelijk Gewest. Van deze laatste regio kon men niet de Nederlandstalige gezinnen selecteren, waardoor we enkel verder gingen met populatiegegevens van Vlaanderen (Zie ook Hoofdstuk 1.3). Dit maakt dat we geen Brusselse populatiegegevens konden integreren waardoor de populatiegegevens niet 100% accuraat zijn om af te zetten tegenover gezinsenquête-gegevens.

2 Weegcoëfficiënten op basis van de eerste kruistabel

De eerste kruistabel die we opvraagden bij het Datawarehouse Arbeidsmarkt en Sociale Bescherming toont het aantal gezinnen in het Vlaams Gewest met minstens één kind jonger dan 25 jaar naar vier kenmerken: herkomst, gezinstype, aantal kinderen van 0 t.e.m. 11 jaar en aantal kinderen van 12 t.e.m. 24 jaar.

Een eerste variabele die wordt opgenomen in de kruistabel, is **herkomst**. In de gezinsenquête achterhaalt men de herkomst van de respondent via vijf vragen: (1) 'In welk land bent u geboren?', (2) 'Welke nationaliteit(en) heeft u op dit moment?', (3) 'Had u een andere nationaliteit bij uw geboorte?', (4) 'Welke was de nationaliteit van uw vader bij zijn geboorte?' en (5) 'Welke was de nationaliteit van uw moeder bij haar geboorte?'. Uit deze vragen werd de herkomst van de respondent afgeleid, zoals deze bepaald wordt conform het Vlaams Integratiedecreet. Dit gebeurde middels een cascadesysteem: ten eerste deelt men in naargelang de nationaliteit van de vader bij geboorte in enerzijds een Belgische nationaliteit of anderzijds een buitenlandse nationaliteit. In geval deze Belgisch is of onbekend, bestudeert men in een tweede stap de geboortenationaliteit van de moeder. Opnieuw wordt de respondent dan ingedeeld in een Belgische of buitenlandse nationaliteit. In geval de moeder een Belgische of onbekende geboortenationaliteit kent, bekijkt men deze van de respondent zelf in een derde stap. Wanneer men opnieuw vaststelt dat deze Belgisch of onbekend is, bestudeert men de huidige nationaliteit van de respondent. Zijn alle nationaliteiten Belgisch? Dan wordt de respondent onderverdeeld in de groep van personen met een Belgische herkomst. Men heeft een niet-Belgische herkomst in geval minstens één van de ouders een niet-Belgische geboortenationaliteit heeft, of ingeval de respondent zelf beschikt over een buitenlandse (geboorte- of huidige) nationaliteit. Vanuit deze indeling, construeerde men een 'herkomst'-variabele met drie mogelijke waarden: Belg, Europees of niet-Europees. Vanuit het Datawarehouse Arbeidsmarkt & Sociale Bescherming kregen we het aantal gezinnen in het Vlaams Gewest met minstens één kind jonger dan 25 jaar met een Belgische, een Europese of een niet-Europese herkomst.

Een tweede kenmerk is **gezinstype**. Afleidend uit de eerste enquêtevraag, namelijk 'Wie behoort volgens u tot uw gezin?', kon men de subjectieve gezinssamenstelling in kaart brengen. Deze definiëring verschilt van de economische invalshoek die een 'huishouden' omschrijft als de personen die op eenzelfde wooneenheid verblijven. Terwijl deze economische definiëring een strikte afbakening hanteert, houdt de vraag in de gezinsenquête rekening met een moderne notie van het gezin waarbij de respondent op subjectieve wijze omschrijft wat 'het gezin' voor hem of haar betekent (Audenaert, 2018). Tal van antwoordmogelijkheden werden gereduceerd tot drie gezinstypes: gehuwd koppel, ongehuwd koppel en eenoudergezin. Via het Datawarehouse Arbeidsmarkt & Sociale Bescherming konden we de variabele 'type huishouden' opvragen, aangeleverd vanuit het Rijks- en KSZ- register. Het type huishouden werd geconstrueerd in samenwerking met het Centrum voor Bevolkings- en Gezinsstudie (CBGS) en gedefinieerd op basis van de LIPRO- typologie. Een huishouden wordt hier gedefinieerd als de eenheid waarin alle personen dezelfde referentiepersonen hebben (Datawarehouse Arbeidsmarkt en Sociale Bescherming, z.j.-b; Lodewijckx, 2001). We vroegen hier de (on)gehuwde koppels met kinderen en de eenpersoonshuishoudens op. Doordat de eenheden van de steekproef werden geselecteerd op basis van private huishoudens, filterden we de collectieve huishoudens uit de populatiegegevens. Hierdoor bleven we, zowel in de gezinsenquête- als in de populatiedata, over met twee categorieën om te vergelijken: (on)gehuwde koppels met kinderen en eenoudergezinnen.

De laatste kenmerken omschrijven **het aantal kinderen** dat een gezin heeft in de **leeftijdsgroep** van 0 tot en met 11 jaar en het aantal kinderen van 12 tot en met 24 jaar. In de gezinsenquête kan men via de eerste vraag deze informatie afleiden. Opnieuw gaat het hier over een subjectieve perceptie van de respondent op de vraag ‘Wie behoort volgens u tot uw gezin?’. We bekijken dus de kinderen die, volgens de respondent, op het ogenblik van de bevraging tot het gezin behoren. Dit geeft een brede waaier aan antwoorden: terwijl 96,7% de eigen kinderen vernoemt, doet 3,3% dit niet. Daarnaast noemt 6,1% van de respondenten ook stiefkinderen, enkelen noemen daarnaast ook schoonkinderen of een kleinkind (Audenaert, 2018).

De **eerste stap** die we ondernemen, is het bekijken van de ‘originale combinaties’ van de gezinsenquête (zie Tabel 22 voor een voorbeeld, en Bijlage 3 voor het gehele overzicht).

Tabel 22 Originele combinaties gezinsenquête (volledig overzicht in Bijlage 3).

Aantal combinaties	Herkomst	Gezinstype	Aantal kinderen 0-11	Aantal kinderen 12-24	Totale aantal kinderen	Aantal gezinnen in Gezins-enquête
1	Belg	(on)gehuwd koppel met kinderen	0	1	1	205
2	Belg	(on)gehuwd koppel met kinderen	0	2	2	256
3	Belg	(on)gehuwd koppel met kinderen	0	3	3	93

De **tweede stap** verkent de vuistregel. Om consequent te werk te gaan, passen we dezelfde vuistregel toe die ook voor de *design* gewichten eerder werd gehanteerd. Deze vuistregel stelt dat het aantal respondenten binnen elke categorie minstens tien moet bedragen om stabiele weegcoëfficiënten te kunnen construeren. In bovenstaande Tabel (zie volledige uitwerking van deze stap in Bijlage 3) markeren we de categorieën waar we de vuistregel op toepassen.

De vuistregel stelt dat er niet meer dan tien observaties in één categorie mogen bevinden. We maken echter de beslissing om de combinaties die niet voldoen aan de vuistregel niet uit de analyse te halen, maar integreren hen in een andere categorie. De algemene regel die hier geldt, om de integratie consequent te doen verlopen, is het indelen van de observaties bij de categorie die het meest hierop gelijkt. In een **derde stap** passen we de vuistregel bijgevolg toe. We kijken dan eerst naar het totale aantal kinderen, om dan te kijken naar een gelijkaardige indeling. In geval we een gezin hebben met vijf kinderen, zullen we deze indelen bij andere gezinnen met vijf kinderen. Let op, we bekijken daaropvolgend ook steeds de verdeling van de kinderen in de leeftijdsgroepen. In geval het gaat om een groot gezin met voornamelijk jonge kinderen, delen we deze in bij andere grote gezinnen met voornamelijk jonge kinderen.

Tabel 23 Voorbeeld toepassing vuistregel (1)

Herkomst	Gezinstype	Aantal kinderen in leeftijdsgroep 0-11	Aantal kinderen in leeftijdsgroep 12-24	Totaal aantal kinderen	Aantal gezinnen in de gezinsenquête
Belg	(on)gehuwd paar met kinderen	3	0	3	10
Belg	(on)gehuwd paar met kinderen	3	1	4	1
Belg	(on)gehuwd paar met kinderen	3	2	5	10

Bovenstaande tabel illustreert dit voorbeeld. Volgens de vuistregel moet het ene Belgische koppel met kinderen met 3 kinderen tussen 0 en 11 jaar en 1 kind tussen 12 en 24 jaar elders ingedeeld worden. Volgens de algemene regel, delen we dit gezin in bij het gezin 'hierboven', dus bij de tien gezinnen met drie kinderen tussen 0 en 11 jaar en 0 kinderen tussen 12 en 24 jaar. Let op, de uitzondering op deze algemene regel stelt dat we de grote gezinnen steeds bij elkaar proberen te plaatsen. In geval, bijvoorbeeld, dit ene gezin met 3 kinderen tussen 0 en 11 jaar en 1 kind tussen 12 en 24 jaar in feite 5 kinderen had (de categorieën zijn in de begrensd tot maximaal 3 kinderen), dan delen we deze in bij de gezinnen daaronder, waar men in totaal 5 kinderen heeft.

Vervolgens bekijken we in een **vierde stap** de originele combinaties van de populatiegegevens (zie Bijlage 3 voor het overzicht). Net zoals in de eerste stap bekijken we voor de populatiegegevens het aantal combinaties naar herkomst, gezinstype en aantal kinderen per leeftijdscategorie. Vervolgens bekijken we of de combinaties van deze populatiegegevens *matchen* met die van de gezinsenquête (na het toepassen van de vuistregel in stap 3). In een **vijfde stap** worden daaropvolgend het aantal combinaties van de populatiegegevens gereduceerd naar deze die aanwezig zijn in de gezinsenquête. In een laatste, **zesde, stap** construeren we effectief de weegcoëfficiënten (zie Tabel 21 voor een voorbeeld, en bijlage 3 voor het gehele overzicht).

3 Weegcoëfficiënten op basis van de tweede kruistabel

Een tweede kruistabel (zie Tabel 20) toont het aantal gezinnen in het Vlaams Gewest met minstens één kind jonger dan 25 jaar naar drie verschillende kenmerken. Het gaat om herkomst (vier waarden: Belg, EU, niet- EU, onbekend), gezinstype (twee waarden: (on)gehuwd paar met kinderen en eenoudergezinnen) en beroepsstatuut (zeven waarden: andere, arbeidsongeschikt, werkloos (leefloon, tijds-krediet, werkzoekende), pensioen, zelfstandige, werknemer of onbekend). Alle drie deze kenmerken werden identiek gecodeerd in beide datasets.

Een eerste kenmerk dat werd opgenomen in de kruistabel, is de herkomst. Deze is identiek aan de herkomstvariabele in de eerste kruistabel. Ook het gezinstype heeft diezelfde definiëring als in kruistabel 1. Een derde en laatste kenmerk is het beroepsstatuut. Via de Kruispuntbank Sociale Zekerheid achterhaalden we het beroepsstatuut middels de beschikbare variabele 'nomenclatuur'. Dit is de arbeidsmarktpositie van het individu op de laatste dag van het kwartaal (in dit geval:

2015)(Datawarehouse Arbeidsmarkt en Sociale Bescherming, z.j.-a). Deze werden opgevraagd volgens negen categorieën: werknemer, zelfstandige, helper, werkzoekend, loopbaanonderbreking of tijdskrediet, leefloon, pensioen, arbeidsongeschikt, andere (Zie bijlage 1). In de gezinsenquête stelt men enkele vragen aan de respondent om te peilen naar zijn of haar werksituatie. Zo wordt nagegaan of deze betaald werk heeft, het huidige beroepsstatuut, het contracttype, een voltijds of deeltijdse tewerkstelling, en dergelijke. Om de gegevens vanuit de gezinsenquête vergelijkbaar te maken met deze van de KSZ, combineerden we enkele antwoorden van de respondenten om tot eenzelfde 'beroepsstatuut'-definiëring te komen zoals in de KSZ wordt gehanteerd. Betreffende het beroepsstatuut werd de vraag gesteld 'Wat is uw huidige beroepsstatuut? U bent...'. De respondenten kregen zes antwoordmogelijkheden aangeboden: arbeider (private sector), bediende (private sector), contractueel of statutair ambtenaar (openbare sector), zelfstandige/vrij beroep zonder personeel, zelfstandige/ vrij beroep met personeel, helper/ meewerkend familielid. Een andere vraag polst naar het al dan niet betaald werk hebben van de respondent. Het zijn deze twee vragen die we combineerden om een nieuwe variabele 'Beroepsstatuut' aan te maken die vergelijkbaar is met KSZ- gegevens.

Om de gewichten te construeren op basis van deze tweede kruistabel, volgden we hetzelfde stappenplan als voor de eerste kruistabel (zie Hoofdstuk 5.3). In bijlage 4 worden de originele combinaties van in de gezinsenquête en deze van de populatie weergegeven.

4 Toepassen van de gewichten

In deze sectie bekijken we wat de impact is van de nieuwe gewichten op de gegevens van de gezinsenquête. De meerwaarde van deze weegcoëfficiënten worden duidelijk in geval men het verschil bekijkt tussen het integreren of negeren van gewichten. Wanneer men beslist de gewichten te negeren, wordt de steekproef als een simpele, random steekproef behandeld zonder afwijkingen in de dekking of non respons. Concreet stelt men dan dat $w_i = 1$ voor elke respondent. Dit betekent dat men geen rekening houdt met potentiële vertekening als gevolg van de steekproeftrekking, en dit leidt tot fouten in schattingen en vertekende resultaten (Biemer & Christ, 2008). Let wel, in wetenschappelijke analyses kan men echter enkel gebruik maken van één gewicht per case. Dit betekent dat, wanneer men weegt voor verschillende kenmerken, deze gewichten gecombineerd moeten worden in één gewicht (Johnson, 2008). In wat volgt bestuderen we de verschillen in resultaten met en zonder de opgestelde gewichten (Tabel 24).

Tabel 24 Voorbeeld toepassing nieuwe gewichten (Bron: DWH AM&SB en gezinsenquête 2016).

	Moeilijk tot zeer moeilijk om de eindjes aan elkaar te knopen			Betaald werk		
	<i>n</i>	%	<i>n</i>	<i>n</i>	%	<i>n</i>
Ongewogen	310	12%	n= 2.514	2.189	84%	n=2.608
Met basisgewicht (weeg1)	264	10%	n= 2.514	2.272	87%	n=2.608
Met nieuw gewicht (W1)	213	8%	n= 2.383	2.153	87%	n=2.468
Met nieuw gewicht (W2)	221	10%	n=2.249	2.128	91%	n=2.330
Met combinatiegewicht (weeg1*W1)	198	8%	n=2.383	1.439	87%	n=1.653
Met combinatiegewicht (weeg1*W2)	204	9%	n=2.249	2.146	92%	n=2.330

Het ongewogen aandeel van personen die het moeilijk hebben om de eindjes aan elkaar te knopen bedraagt 12%. Gewogen met het basisgewicht daalt het aandeel naar 10%. Wanneer we dan het nieuwe gewicht W1 (of de eerste poststratificatieweegcoëfficiënt op basis van sociodemografische variabelen) hanteren, dan daalt het aandeel verder naar 8% van personen die het moeilijk hebben om rond te komen. Toepassing van W2 (of het tweede poststratificatieweegcoëfficiënt op basis van beperkte socio-demografische variabelen en arbeidsmarktpositie) leidt tot een aandeel van 10%. Wanneer we het eerste basisgewicht combineren met het eerste poststratificatiegewicht, bekomen we een aandeel van 8% die moeite hebben de eindjes aan elkaar te knopen. Het basis gewicht vermenigvuldigd met het tweede poststratificatiegewicht geeft 9% die die moeite hebben om rond te komen.

Een tweede variabele waarop we de nieuwe weegcoëfficiënten toepassen, is of men al dan niet beschikt over een betaalde job. Ongewogen hebben 84% van de respondenten van de gezinsenquête betaald werk. Toepassing van het poststratificatiegewicht W1 dat gebaseerd is op de verhouding van louter sociodemografische variabelen in de populatie verandert dit aandeel niet. Toepassing van poststratificatiegewicht W2 dat ook weegt op arbeidsmarktstatuut heeft een grote invloed: het aandeel personen met betaald werk stijgt naar 91%, en in combinatie tot 92%.

Voor de vervolgenquêtes van de Gezinsenquête lijkt het aangewezen om poststratificatiegewichten ter beschikking te stellen. Zeker voor beschrijvende analyses is het aangewezen om zowel design als poststratificatiegewichten toe te passen omdat het een meer precieze veralgemening naar de bevolking toelaat. De keuze voor de achterliggende variabelen is echter cruciaal, en is bij voorkeur niet alleen gebaseerd op louter sociodemografische gegevens. Er bestaat echter niet zoals als de 'ideale variabelen' om poststratificatiegewichten te ontwikkelen. De keuze zal gebaseerd zijn op beschikbaarheid van populatiegegevens, en de inhoudelijke verschillen waarin onderzoekers geïnteresseerd zijn (Pickery, 2010a).

Hoofdstuk 6

Conclusie

De gezinsenquête 2016 werd ontwikkeld met het oog op het bestuderen van ‘het gezin’ in Vlaanderen en Nederlandstalig Brussel. Vanuit een bezorgdheid dat niet alle bevolkingsgroepen werden bereikt om een totaalbeeld van ‘het gezinsleven’ te kunnen schetsen, en meer specifiek de financieel kwetsbare gezinnen, werd deze studie opgezet. Het opzet van dit methodologisch rapport is om de representativiteit van de gezinsenquête onder de loep te nemen, en daarnaast ook aandacht te schenken aan andere componenten van het design van de gezinsenquête in functie van toekomstige edities. Deze analyse vond plaats aan de hand van een uitgebreide literatuurstudie en statistische descriptieve analyses via verkregen administratieve gegevens vanuit het Datawarehouse Arbeidsmarkt en Sociale Bescherming.

Ten eerste stellen we vast dat de gerealiseerde basissteekproef een adequate weerspiegeling biedt van de basis geselecteerde gezinnen. Met andere woorden, de respondenten die men in deze initieel opgezette steekproef heeft bereikt, zijn algemeen een representatief staal van de doelpopulatie (i.e. alle gezinnen met minstens één kind jonger dan 25 jaar in Vlaanderen of Nederlandstalig Brussel). Er zijn geen significante verschillen te vinden tussen de geselecteerde basissteekproef en de gerealiseerde basissteekproef op vlak van geslacht, leeftijd, gezinstype, provincie en herkomst, noch op vlak van de nomenclatuur van de socio-economische positie, het contracttype en de bruto belastbare gezinsinkomens.

Ten tweede concluderen we dat de extra steekproef bij gezinnen met een niet-EU-achtergrond voor- en nadelen heeft met betrekking tot de representativiteit van (kwetsbare) gezinnen. Vanuit een bezorgdheid dat personen met een herkomst van buiten de Europese Unie vaak ondervertegenwoordigd zijn in bevragingen, werd een samenwerkingsverband opgezet met het Agentschap Binnenlands Bestuur (ABB) om een uitbreidende steekproef te trekken die specifiek op het criterium ‘herkomst’ selecteert. Uit de analyses blijken personen en gezinnen met een niet-EU-herkomst echter voldoende vertegenwoordigd te zijn, zowel in de geselecteerde als in de gerealiseerde basissteekproef. Met andere woorden, de gerealiseerde basis steekproef weerspiegelt de facto een divers publiek op vlak van herkomst, waardoor de extra steekproef, specifiek met dit vooropgestelde criterium ‘herkomst’, overbodig bleek. Sterker, deze uitbreiding zorgde voor een neveneffect, namelijk de oververtegenwoordiging van specifieke gezinnen waardoor de totaal gerealiseerde steekproef in feite vertekend was. Desalniettemin zagen we dat deze inhoudelijk wel een waardevolle bijdrage leverde om voldoende specifieke profielen te bereiken, zoals bijvoorbeeld leefloongerechtigden, lage inkomensgezinnen, grote gezinnen en laaggeschoolden. Voor de organisatie van de gezinsenquête in de toekomst is het belangrijk om op voorhand af te wegen wat het doel is van een eventuele extra steekproef, en wie men daar mee wil bereiken.

Gekoppeld aan de literatuurstudie met betrekking tot de steekproeftrekking is daarbij de vraag in welke mate mogelijke alternatieve criteria kunnen vooropgesteld worden bij volgende golven van de gezinsenquête. Dit behoeft een zekere reflectie: wat is het onderzoeksopzet? Wat wil men bereiken met een extra steekproef? En dient deze uitbreiding voornamelijk om inhoudelijke verdiepingen te verkrijgen en/of eerder om methodologisch specifieke groepen beter te kunnen bereiken? In dat eerste geval kan het interessant zijn om in een uitbreidende steekproef andere vragen te stellen, in het tweede geval

wordt aangeraden om steeds een voldoende grote steekproef te trekken. In geval van deze uitbreidende steekproef zagen we namelijk dat, doordat men 'slechts' 493 respondenten bereikte, deze groep voor bepaalde kenmerken (zoals herkomst, inkomens en werkintensiteit) geen goede weerspiegeling bood van de extra geselecteerde steekproef. Kortom, de centrale vraag die hier naar voren wordt geschoven, is of én het steekproefkader én de steekproeftrekking die men kiest, het beste alternatief zijn om álle gezinnen te kunnen bereiken om een representatief beeld van gezinnen met kinderen in Vlaanderen te kunnen verkrijgen. Een extra steekproef van niet-EU-respondenten bleek interessant om specifieke profielen te bereiken, zoals leefloongerechtigden. Maar met het oog op volgende edities kan het interessanter zijn om specifiek een extra steekproef te trekken van financieel kwetsbare gezinnen waar ook een duidelijke inhoudelijke (beleids?-)component mee samenhangt. Men zou er bijvoorbeeld voor kunnen opteren om een extra steekproef te trekken op basis van de kansarmoede-index van Kind en Gezin. Dit is niet alleen een goed steekproefkader om een aselechte steekproef uit te trekken, maar biedt ook inhoudelijk heel wat kansen omdat de kansarmoede-index van Kind en Gezin ook beleidsmatig een belangrijke rol speelt.

Niet enkel de steekproeftrekking behoeft een reflectie: ook andere componenten van de gezinsenquête kunnen in functie van volgende golven in vraag gesteld worden. Het cross-sectionele design van de gezinsenquête is interessant om algemeen trends te kunnen volgen in het gezinsleven, maar wanneer men dynamieken en transitieën waar gezinnen zich in bevinden of in kunnen terechtkomen nauwgezetter wil opvolgen, kan een longitudinaal design meer effectief werken om gezinnen op dynamische wijze te benaderen. Het bespreken van enkele scenario's maakt duidelijk dat er verscheidene mogelijkheden zijn voor komende golven van de gezinsenquête, maar elk scenario kent voor- en nadelen. Doorheen het rapport konden we de proef op de som nemen om scenario's waarin administratieve gegevens gekoppeld werden met surveygegevens te testen. De administratieve gegevens kennen echter tal van beperkingen die deze koppeling bemoeilijken. Zo zijn deze gegevens steeds met enige vertraging te verkrijgen, terwijl er met enquêtegegevens veel sneller aan de slag kan worden gegaan. Bepaalde bevroegde concepten in de gezinsenquête kwamen niet overeen met hoe administratieve gegevens zijn opgebouwd, wat imputatie onmogelijk maakt. Denk bijvoorbeeld aan de inkomensvraag, waarbij in de gezinsenquête de netto beschikbare gezinsinkomens werden bevroegd op maandbasis, terwijl in de administratieve gegevens enkel bruto (belastbare) gezinsinkomens beschikbaar zijn op jaarbasis. Daarnaast zagen we dat administratieve gegevens vaak ook tekortschoten voor enkele cruciale kenmerken, zoals het opleidingsniveau. Het centraal stellen van enquêtegegevens is, op basis van deze eerste analyse van de eerste golf, de beste optie, hoewel men steeds moet bewaken dat vergelijking (of imputatie) met alternatieve databronnen mogelijk blijven.

Daarbovenop stelden we vast dat de item non respons in de gezinsenquête relatief laag ligt voor vier geselecteerde items (het kunnen rondkomen, jaarlijkse vakantie, hoogst behaalde diploma en beschikbare inkomen). Degene die de vragen niet beantwoordden, waren opvallend de 'sterkere' profielen: vooral personen uit een jobrijk gezin beslisten om de vraag open te laten. Desalniettemin observeerden we een vertekening in de item non respons wat betreft het opleidingsniveau en het inkomen. Het waren namelijk significant meer laaggeschoolden die de vragen onbeantwoord lieten enerzijds, en meer inkomensarme gezinnen die specifiek de vraag naar inkomens openlieten anderzijds. Terugkoppelend naar de dataverzamelmethode van de gezinsenquête (Hoofdstuk 3), kan men de item non respons zien als een nadeel van de gehanteerde postenquête. De reden waarom specifiek laagopgeleiden de vragen openlieten, is daarenboven onbekend. Door de postenquête aan te vullen met persoonlijke interviews, kan men deze reden mogelijks achterhalen. Indien, bijvoorbeeld, een vraag niet

duidelijk is, kan de interviewer deze verduidelijken en signalen opvangen om de vragenlijst van volgende edities meer helder te omschrijven. Hoewel er geen kant-en-klare oplossing bestaat over welke methode(s) men best kan hanteren om specifieke doelpopulaties te bereiken, en om item non respons te reduceren, kan een *mixed method* design in volgende edities een uitweg bieden.

Referenties

- Agentschap Binnenlands Bestuur. (2018a). Gemeente- en stadsmonitor Vlaanderen: Surveys. *Gemeente- en stadsmonitor*. Retrieved from <https://gemeente-en-stadsmonitor.vlaanderen.be/surveys>
- Agentschap Binnenlands Bestuur. (2018b). Stadsmonitor 2017: een monitor voor leefbare en duurzame Vlaamse steden. Retrieved from https://gemeente-en-stadsmonitor.vlaanderen.be/sites/gemeente-en-stadsmonitor/files/stadsmonitor_pub.pdf
- Audenaert, V. (2018). *Gezinsenquête 2016: De samenstelling van het gezin. Gezinnen in Vlaanderen over hoe het gezin er uit ziet, wie bij het gezin hoort, waar gezinsleden wonen, over hun kindwens en wat men liever anders had gezien in de samenstelling van het gezin*. Retrieved from www.gezinsenquête.be
- Audenaert, V., & Stuyck, K. (2018a). *Gezinsenquête 2016: Methodologisch Rapport*. Retrieved from Brussel: www.gezinsenquête.be
- Audenaert, V., & Stuyck, K. (2018b). *Methodologisch rapport. Bijlage 2: Het steekproefplan*. Retrieved from Brussel: <https://gezinsenquête.be/methodologisch-rapport>
- Babbie, E. R. (1973). *Survey research methods*. Belmont, CA: Wadsworth.
- Bakker, B., Arts, K., & Toor, L. (2005). Arbeidsmarktstatistieken in Nederland. De opzet van het sociaal statistisch bestand. *OVER-WERK Tijdschrift van het Steunpunt WAV*, 1(15), 50-54.
- Bennett, F., & Roberts, M. (2004a). *From input to influence Participatory approaches to research and inquiry into poverty*. Oxford: Joseph Rowntree Foundation.
- Bennett, F., & Roberts, M. (2004b). Participatory approaches to research on poverty. An overview of research approaches which give people with experience of poverty more involvement and influence. Retrieved from <https://www.jrf.org.uk/report/participatory-approaches-research-poverty>
- Beukenhorst, D. J. (1999). *Vertekening door Non-Respons in het WBO*. Retrieved from Heerlen:
- Beukenhorst, D. J. (2001). *Rendement van het mixed-mode design van POLS en vertekening*. Retrieved from Heerlen:
- Beullens, K., & Loosveldt, G. (2012). Should high response rates really be a primary objective? *Survey Practice*, 5(3), 1-5.
- Beullens, K., & Loosveldt, G. (2016). Interviewer Effects in the European Social Survey. *Survey Research Methods*, 10(2), 103-118.
- Biemer, P. P., & Christ, S. L. (2008). Weighting Survey Data. In J. J. Hox, E. De Leeuw, & D. A. Dillman (Eds.), *The International Handbook Of Survey Methodology* (pp. 317-341). New York: Lawrence Erlbaum Associates.
- Bircan, T., Nicaise, I., Gábos, A., & Tóth, G. (2017). *Priorities in data collection for European poverty research (Deliverable 20.4)*. Retrieved from Leuven: www.hiva.kuleuven.be
- BISA. (2016). *Fiscale statistiek van de inkomens, een geschikte levensbron om de levensstandaard van de Brusselaars te meten?* Retrieved from Brussel: http://bisa.brussels/bestanden/publicaties/focus-van-het-bisa/focus_14_februari_2016
- Blasius, J., & Thiessen, V. (2012). *Assessing the quality of survey data*. London: SAGE Publications Ltd.
- Blumberg, S., Luke, J., Ganesh, N., Davern, M., Boudreaux, M. H., & Soderberg, K. (2011). *Wireless substitution: State-level estimates from the National Health Interview Survey*. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/21568134>
- Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., . . . Hughes, C. (2014). Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical research Methodology*, 14(42), 1-29.
- Boyd, H., & Westfall, R. (1985). Optimal design of interviewer variance experiments in complex surveys. *Journal of the American Statistical Association*, 80, 158-166.
- Braes, S., & Herremans, W. (2012). *Inhoudelijke analyse Datawarehouse AM&SB: Low Work Intensity at Household Level*. Retrieved from Brussel: https://www.ksz-bcss.fgov.be/nl/dwh/dwh_page/content/websites/datawarehouse/others/lwi-indicator.html
- Bronselaer, J., Vandezande, V., Vanden Boer, L., & Boons, I. (2015). *Duurzame mantelzorg in Vlaanderen: methodologisch rapport*. Retrieved from Brussel: https://www.departementwvg.be/sites/default/files/media/Departement_WVG_MethodologischRapport_Mantelzorg.pdf

- Chaudhuri, S., Jalan, J., & Suryahadi, A. (2002). *Assessing household vulnerability to poverty from cross-sectional data: a methodology from estimates from Indonesia*. Retrieved from New York: <https://academiccommons.columbia.edu/doi/10.7916/D85149GF>
- Chromy, J. R., & Abeyasekera, S. (2005). Chapter 19: Statistical analysis of survey data. In United Nations Statistical Division (Ed.), *Household surveys in developing and transition countries: design, implementation and analysis*. United States.
- Church, A. H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, 57(1), 62-79.
- Cohen, L., & Manion, L. (1980). *Research Methods in Education*. London: Croon Helm.
- Corluy, V., & Verbist, G. (2010). *Inkomen en diversiteit: onderzoek naar de inkomenspositie van migranten in België*. Retrieved from Antwerpen: <http://www.centrumvoorsociaalbeleid.be/docs/20100507104613RYRB.pdf>
- Couper, M. P. (2011). The future of modes of data collection. *Public Opinion Quarterly*, 75, 889-908.
- Datawarehouse Arbeidsmarkt en Sociale Bescherming. (z.j.-a). Variabele: Nomenclatuur van socio-economische positie. Details van de variabele. Retrieved from <https://www.ksz-bcss.fgov.be/nl/dwh/variabledetail/kruispuntbank-van-de-sociale-zekerheid/Variables/nomenclatuur-van-socio-economische-positie-1.html?filter=institution&institution=Kruispuntbank%20van%20de%20Sociale%20Zekerheid&sources=&themes=>
- Datawarehouse Arbeidsmarkt en Sociale Bescherming. (z.j.-b). Variabele: Type huishouden. Details van de variabele. Retrieved from <https://www.ksz-bcss.fgov.be/nl/dwh/variabledetail/rijks-en-ksz-register/Variables/type-huishouden.html?filter=institution&institution=Rijks-%20en%20KSZ-register&sources=&themes=>
- De Keulenaer, F. (2002). *Characteristics of second-wave non-respondents in the panel study of Belgian households*.
- De Keulenaer, F., & Levecque, K. (2004). *Mood and socio-economic status bias in survey non-response: results from an 11-wave panel*. Paper presented at the 6th International Conference on Social Science Methodology: Recent Developments and Applications in Social Research Methodology, Amsterdam.
- De Lathouwer, L., Bogaerts, K., & Dillen, H. (2006). *Transities op de arbeidsmarkt. Een oefening met administratieve panelgegevens*. Gent: Academia Press.
- De Leeuw, E., Hox, J. J., & Dillman, D. A. (2008). *International Handbook of Survey Methodology*. New York: Taylor & Francis Group.
- Dekkers, G., Tarantchenko, E., & Van den Bosch, K. (2019). *Medium-term projection for Belgium of the at-risk-of-poverty and social exclusion indicators based on EU-SILC*. Retrieved from Brussels: https://www.plan.be/admin/uploaded/201903110928260.WP_1903_11838.pdf
- Demarest, S., Van der Heyden, J., Sarr, K., De Waleffe, S., & Meersseman, E. (2015). *Gezondheidsenquête 2013. Rapport 5: preventie*. Retrieved from Brussel:
- Dillman, D. A. (1978). *Mail and telephone surveys: The Total Design Method*. New York: John Wiley & Sons.
- Dillman, D. A. (2012). Introduction to Special Issue of Survey Practice on item nonresponse. *Survey Practice*, 5, 1-3.
- Dillman, D. A., Smyth, J., & Christian, L. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method*. Hoboken, NJ: John Wiley.
- Dillman, D. A., Smyth, J., & Christian, L. (2014). *Internet, phone, mail and mixed-mode surveys. The tailor design method: 4th Editon*. New Jersey: John Wiley & Sons.
- Driessens, K. (2003). Mensen in armoede als participanten in onderzoek. Over de meerwaarde van kwalitatief onderzoek in het welzijnswerk. In A. Claeys, A. De Cock, & J. Vanhee (Eds.), *Armoede en participatie* (pp. 115-136). Gent: Academia Press.
- Eurostat. (2018). Europe 2020 indicators - poverty and social exclusion. Retrieved from http://ec.europa.eu/eurostat/statistics-explained/index.php/Europe_2020_indicators_-_poverty_and_social_exclusion
- Eurostat. (n.d.). Glossary: EU statistics on income and living conditions (EU-SILC). *Statistics explained*. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_statistics_on_income_and_living_conditions_%28EU-SILC%29
- FOD Binnenlandse Zaken. (2019). Rijksregister. Retrieved from <http://www.ibz.rn.fgov.be/nl/rijksregister/>
- Frans, D., Mortelmans, D., Sels, L., & Forrier, A. (2009). *Haalbaarheidsstudie loopbaanpanel: Technisch rapport*. Retrieved from Leuven: www.steunpuntwse.be

- Fulton, J., & Medway, R. (2012). *When more gets you less: A meta-analysis of the effect of concurrent web options on mail survey response rates*. Paper presented at the 67th Annual Conference of American Association for Public Opinion research, May 17-20, Orlando.
- Ghysels, J., & Debacker, M. (2007a). *Zorgen voor kinderen in Vlaanderen: een dagelijkse evenwichtsoefening?* Leuven/Voorburg: Acco.
- Ghysels, J., & Debacker, M. (2007b). *Zorgen voor kinderen in Vlaanderen: Een dagelijkse evenwichtsoefening? OVER-WERK Tijdschrift van het Steunpunt WSE*, 3(4/2007), 141-145.
- Groves, R. M. (2006). Non-response rates and non-response bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.
- Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56(4), 475-495.
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.
- Groves, R. M., & Peytcheva, E. (2006). *The impact of nonresponse rates on nonresponse bias: A meta-analysis*. Paper presented at the 17th International Workshop on Household Survey Nonresponse, Omaha, NE, USA.
- Groves, R. M., & Singer, E. C., Amy. (2000). Leverage-Saliency Theory of Survey Participation: Description and an Illustration *Public Opinion Quarterly*, 64(3), 299-308.
- Haan, M., & Ongena, Y. (2014). Tailored and targeted designs for hard-to-survey populations. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 555-574). Cambridge: Cambridge University Press.
- Haan, M., Ongena, Y., & Aarts, K. (2014). Reaching Hard-to-Survey Populations: Mode Choice and Mode Preference. *Journal of Official Statistics*, 30(2), 355-379.
- Hantrais, L. (2004). *Family policy matters. Responding to family change in Europe*. Bristol: The Policy Press.
- Harkness, J., Stange, M., Cibelli, K. L., Mohler, P., & Pennell, B.-E. (2014). In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 245-269). Cambridge: Cambridge University Press.
- Harris, S. R. (2008). What is family diversity? *Journal of Family Issues*, 29(11), 1407-1425.
- Harrison, C. H. (2010). Mail Surveys and Paper Questionnaires. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research: 2nd edition* (pp. 499-526). United Kingdom: Emerald Group Publishing Limited.
- Haskey, J. (2010). Measuring Family and Household Variables In M. Bulmer, J. Gibbs, & L. Hyman (Eds.), *Social Measurement through Social Surveys: An Applied Approach* (pp. 9-28). Surrey, England: Ashgate Publishing Limited.
- Heerwegh, D., Abts, K., & Loosveldt, G. (2007). Minimizing survey refusal and noncontact rates: Do our efforts pay off? *Survey Research Methods*, 1(1), 3-10.
- Huisman, M. (2007). Multipiele imputatie van ontbrekende scores. In A. E. Bronner, P. Dekker, E. De Leeuw, L. J. Paas, K. De Ruyter, A. Smidts, & J. E. Wieringa (Eds.), *Ontwikkelingen in het marktonderzoek. Jaarboek 2007* (pp. 171-188). Haarlem: Spaar en Hout.
- Janssens, R. (2001). *Taalgebruik in Brussel. Taalverhoudingen, taalverschuivingen en taalidentiteit in een meertalige stad*. Brussel: VUB Press.
- Jäntti, M., Törmälehto, V.-M., & Marlier, E. (2013). *The use of registers in the context of EU-SILC: challenges and opportunities*. Retrieved from Brussels: <https://ec.europa.eu/eurostat/documents/3888793/5856365/KS-TC-13-004-EN.PDF>
- Jenkins, C., & Dillman, D. A. (1997). Towards a theory of self-administered questionnaire design. In R. M. Groves, D. A. Dillman, J. Eltinge, & R. J. Little (Eds.), *Survey nonresponse* (pp. 165-196). New York: Wiley.
- Johnson, D. R. (2008). *Using weights in the analysis of survey data*. Retrieved from Pennsylvania: <http://www.nyu.edu/classes/jackson/design.of.social.research/Readings/Johnson%20-%20Introduction%20to%20survey%20weights%20%28PRI%20version%29.pdf>
- Juchtmans, G., & Groenez, S. (2016). *Opgroeien in kansarmoede: een profielschets van de VLAS-responderen*. Retrieved from Antwerpen: <https://www.uantwerpen.be/nl/steunpunten/vlaams-armoedesteunpunt/publicaties/>
- Kalton, G. (2014). Probability sampling methods for hard-to-sample populations. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-survey populations* (pp. 401-423). Cambridge: Cambridge University Press.

- Kind en Gezin. (2018). *Het kind in Vlaanderen 2018*. Retrieved from Brussel: <https://www.kindengezin.be/img/KIV2018-HFST4.pdf>
- Kind en Gezin. (z.d.). Kinderopvang Brussel. Retrieved from www.kindengezin.be/kinderopvang/starter/babys-en-peuters/brussel
- Kreuter, M. W. (2003). Tailored and targeted health communication: strategies for enhancing information relevance. *American Journal of Health Behavior*, 27(3), 227-232.
- Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.
- Künn, S. (2015). *The challenges of linking survey and administrative data*. Retrieved from Maastricht, the Netherlands: <https://wol.iza.org/uploads/articles/214/pdfs/challenges-of-linking-survey-and-administrative-data.pdf>
- Kuppens, S., Lemmens, E., Juchtmans, G., & Groenez, S. (2016). *VLAS-Studies 36. Opgroeien in kansarmoede: Verdere uitdieping van de VLAS-cohorte*. Retrieved from Antwerpen: <https://steunpuntwvg.be/jong>
- Lauwereys, G., Neels, K., & De Winter, T. (2011). *GGs Wave 1 Belgium: Final disposition codes and standardised response rates*. Retrieved from www.ggps.be
- Lemmens, E., & Groenez, S. (2015a). *Een longitudinaal onderzoek naar ontwikkeling en opvoeding bij kwetsbare gezinnen met jonge kinderen: een overzicht van onderzoeksopzet en instrumentarium*. Retrieved from Antwerpen: https://www.expoo.be/sites/default/files/atoms/files/20151215_VLAS%20Studies%2025_Overzicht%20van%20onderzoeksopzet%20en%20instrumentarium%20JONG%20databank_LemmensGroenez.pdf
- Lemmens, E., & Groenez, S. (2015b). *Stilstaan bij de implicaties van armoedeonderzoek: Kritische reflecties over een VLAS- onderzoek met kansarme gezinnen*. Retrieved from Antwerpen: <https://www.uantwerpen.be/nl/steunpunten/vlaams-armoedesteunpunt/publicaties/>
- Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data (2nd edition)*. New York: John Wiley.
- Lodewijckx, E. (2001). *Huishoudens in België. Een ontsluiting van rijksregistergegevens*. Retrieved from <https://lib.ugent.be/nl/catalog/rug01:000696833>
- Lodewijckx, E., & Deboosere, P. (2011). *Huishoudens en families: Stabiliteit en snelle veranderingen gaan hand in hand*. Retrieved from Brussel: http://www.ggps.be/doc/GGP_Belgium_Paper_Series_6-NL.pdf
- Luyten, D. (2017). *Presentatie op Dag van het gezin 2017: Het gezin in Vlaanderen 2.0. Uitdagingen voor het gezinsbeleid van vandaag en morgen*. Retrieved from Brussel:
- Lynn, P. (2008). The problem of nonresponse. In E. De Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology* (pp. 35-55). New York/London: Taylor & Francis Group.
- Lynn, P. (2013). *Targeted response inducement strategies on longitudinal surveys*. Retrieved from Essex: <https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2013-02.pdf>
- Lynn, P. (2017). From standardised to targeted survey procedures for tackling non-response and attrition. *Survey Research Methods*, 11(1), 93-103.
- Maniquet, F., Schokkaert, E., Cantillon, B., & De Rock, B. (2018). *MEQIN. Measuring Equivalent Incomes: End report*. Retrieved from Brussels: http://www.belspo.be/belspo/brain-be/projects/FinalReports/MEQUIN_en.pdf
- Mercer, A., Caporaso, A., Cantor, D., & Townsend, R. (2015). How much gets you how much? Monetary incentives and response rates in household surveys. *Public Opinion Quarterly*, 79(1), 105-129.
- O'Hara, A., Shattuck, R. M., & Goerge, R. M. (2017). Linking Federal Surveys With Administrative Data to Improve Research on Families. *ANNALS of the AAPSS*, 669, 63-74. doi:10.1177/0002716216678391
- Pasteels, I., Lodewijckx, E., & Mortelmans, D. (2013). Gezinstransities in de levensloop van mannen en vrouwen. In M. Corijn & C. Van Peer (Eds.), *Gezinstransities in Vlaanderen* (pp. 37-72). Brussel: Studiedienst van de Vlaamse Regering.
- Pickery, J. (2010a). *Aanmaak en gebruik van gewichten voor surveydata*. Retrieved from Brussel: www.vlaanderen.be
- Pickery, J. (2010b). *Kunnen telefonische surveys nog representatief zijn nu het aantal vaste telefoonlijnen bij de huishoudens blijft dalen?* . Retrieved from Brussel: <https://www.vlaanderen.be/publicaties/kunnen-telefonische-surveys-nog-representatief-zijn-nu-het-aantal-vaste-telefoonlijnen-bij-de-huishoudens-blijft-dalen>
- Rookey, B. D., Hanway, S., & Dillman, D. A. (2008). Does a probability-based household panel benefit from assignment to postal response as an alternative to internet-only? *Public Opinion Quarterly*, 72(5), 1-23.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.

- Ruspini, E. (2002). *Introduction to longitudinal research*. London: Routledge.
- Schepers, W., Juchtmans, G., & Nicaise, I. (2017). *Reaching out to hard-to-survey groups among the poor. Survey protocols, statistical issues and research design*. Retrieved from Leuven: <https://inclusivegrowth.be/downloads/output/ms87-m20-5-hard-to-survey-groups.pdf>
- Schmeets, H., & Van der Bie, R. (2006). *Enquêteonderzoek onder Allochtonen: problemen, oplossingen en uitdagingen*. Retrieved from <https://www.cbs.nl/-/media/imported/documents/2006/48/2006-k1-b15-p067-art.pdf>
- Schockaert, I., Morissens, A., Cincinnato, S., & Nicaise, I. (2012). *Armoede tussen de plooiën. Aanvullingen en correcties op EU-Silc voor verborgen groepen armen*. Retrieved from Leuven: https://lirias.kuleuven.be/bitstream/123456789/372186/2/R1473_armoede_tussen_de_plooiën.pdf
- Schouten, B. (2004). *Adjustment for Bias in the Integrated Survey on Household Living Conditions (POLS)*. Retrieved from Voorburg/Heerlen:
- Schouten, B., Calinescu, M., & Luiten, A. (2013). Optimizing quality of response through adaptive survey designs. *Survey methodology*, 39(1), 29-58.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Orlando, FL: Academic Press.
- Singer, E., & Ye, C. (2012). The use and effects of incentives in surveys. *ANNALS of the AAPSS*, 645(1), 112-141.
- Skutnabb-Kangas, T., & Phillipson, R. J. (1989). 'Mother Tongue': The theoretical and sociopolitical construction of a concept. In U. Ammon (Ed.), *Status and function of languages and language varieties* (pp. 450-477). Berlin/New York: De Gruyter.
- Sociaal en Cultureel Planbureau. (2007). *Permanent Onderzoek Leefsituatie: Documentatie*. Retrieved from Nederland: https://www.scp.nl/Onderzoek/Bronnen/Beknopte_onderzoeksbeschrijvingen/Permanent_onderzoek_naar_de_leefsituatie_POLS
- Statbel. (2017). Meer dan helft Belgische internetgebruikers koopt online. *Actueel*. Retrieved from <https://statbel.fgov.be/nl/nieuws/meer-dan-helft-belgische-internetgebruikers-koopt-online>
- Statbel. (2019a). Enquête naar de arbeidskrachten (EAK): Beschrijving. Retrieved from <https://statbel.fgov.be/nl/enquete/enquete-naar-de-arbeidskrachten-eak>
- Statbel. (2019b). Structuur van de bevolking: metadata. *Verwerking SV-datawarehouse*.
- Stedman, R. C., Connelly, N. A., Heberlein, T. A., Decker, D. J., & Allred, S. B. (2018). The End of the (Research) World As We Know It? Understanding and Coping With Declining Response Rates to Mail Surveys. *Society & Natural Resources*, 32(10), 1139-1154.
- Stern, M. J., Adams, A., & Elsasser, S. (2009). Digital inequality and place: The effects of technological diffusion on Internet proficiency and usage across rural, suburban and urban counties. *Sociological Inquiry*, 79, 391-417.
- Stern, M. J., Bilgen, I., & Dillman, D. A. (2014). The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. *Field Methods*, 26(3), 284-301.
- Stern, M. J., & Dillman, D. A. (2006). Community participation, social ties, and use of the Internet. *City and Community*, 5, 409-424.
- Stoop, I. (2014). Representing the populations: what general social surveys can learn from surveys among specific groups. In R. Tourangeau, B. Edwards, T. P. Johnson, K. M. Wolter, & N. Bates (Eds.), *Hard-to-Survey Populations* (pp. 225-244). Cambridge: Cambridge University Press.
- Studiedienst van de Vlaamse Regering. (2016). *Alleenstaande ouders in de schijnwerper*. Retrieved from Brussel: <http://www.statistiekvlaanderen.be/sites/default/files/atoms/files/svr-stats-2016-10-alleenstaande-ouders.pdf>
- Stuyck, K., Doyen, G., Feys, Y., Noppe, J., Jacques, A., & Buysschaert, P. (2018). *Survey Samenleven in Diversiteit 2017*. Retrieved from Brussel:
- Te Riele, S. (2002). *Vertekening door non-respons. Hoe nauwkeurig zijn de uitkomsten van persoonsenquêtes?*. Retrieved from Den Haag: <https://www.cbs.nl/NR/rdonlyres/C4F72666-8C9D-463D-89E1-768FD57B0555/0/2002m04v4p020art.pdf>
- Termote, A., & Depickere, A. (2018). *De hervormde enquête naar de arbeidskrachten in 2017*. Retrieved from Brussel: https://statbel.fgov.be/sites/default/files/Over_Statbel_FR/Analyse_eak_2017_nl_20181220.pdf
- Van Crombrugge, H. (2013). *Ouders in soorten*. Antwerpen: Garant.
- Van Lancker, W., Cappelle, H., Cox, N., Decorte, A., Defosse, E., Denoo, H., . . . Willems, A. (2020). *MISSION: Mobile Integrated Social Services Increasing employment Outcomes for people in Need: Final report*

- Retrieved from Kortrijk: http://www.wimvanlancker.be/wp-content/uploads/2020/02/MISSION_FINAL-REPORT_31-01-20.pdf
- Van Peer, C. (2016). *Alleenstaande ouders op de arbeidsmarkt in Vlaanderen en de EU*. Retrieved from Brussel: https://www.statistiekvlaanderen.be/sites/default/files/atoms/files/svr-verkenning-2016-01-alleenstaande-ouders-arbeidsmarkt_0.pdf
- Vlaamse Overheid. (2019). Woononderzoek en statistieken: Demografische cijfers. *Onderzoek en Statistieken*. Retrieved from <https://www.wonenvlaanderen.be/woononderzoek-en-statistieken/demografische-cijfers>
- Weisberg, H. F. (2009). *The total survey error approach: a guide to the new science of survey research*. Chicago: The University of Chicago Press.
- Yorburg, B. (1975). The nuclear and the extended family: An area of conceptual confusion. *Journal of Comparative Family Studies*, 6(1), 5-14.

Bijlagen

- 1 Overzicht gevraagde gegevens per instelling uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB)**

Tabel 25 Overzicht aanvraag bij het Datawarehouse Arbeidsmarkt en Sociale Bescherming met de gegevens per instelling

<i>Leverende instantie</i>	<i>Beschrijving gegeven</i>	<i>Benaming gegeven en bron</i>	<i>Mogelijke waarden</i>
Agentschap voor Hoger Onderwijs, Volwassenenonderwijs, Kwalificaties en Studietoelagen	Taal van het onderwijs	Taal Bron: DWH_AHOVOKS_LED	1. Nederlands 2. Frans
	Aard van het onderwijs gevolgd in het secundair onderwijs	Onderwijsvorm Bron: DWH_AHOVOKS_LED	ASO: Algemeen Secundair Onderwijs BSO: Beroepssecundair Onderwijs KSO: Kunstsecundair Onderwijs TSO: Technisch Secundair Onderwijs
	Classificatie van onderwijsprogramma's en studierichtingen	ISCED_niveau Bron: DWH_AHOVOKS_LED	1 Lager onderwijs – Basiseducatie 2 Eerste graad secundair onderwijs 3A ASO, TSO en KSO secundair onderwijs 3C BSO Secundair onderwijs 4 Post-secundair niet-hoger onderwijs 5 Hoger onderwijs 5A Hoger academisch onderwijs 5B Hoger professioneel onderwijs 6 Doctoraat NBS Niet beschikbaar
	Classificatie van de bewijzen	Categorie Bron: DWH_AHOVOKS_LED	Basiseducatie Secundair onderwijs Hoger beroepsonderwijs Hoger onderwijs Ervaringsbewijzen Leertijd Ondernemersopleidingen

FAMIFED	Hoedanigheid van persoon m.b.t. de kinderbijslag	Hoed	Andere opleidingen
		Bron: DWH_FAMIFED	101 Rechthebbende 102 Bijslagtrekkende type 1 103 Bijslagtrekkende type 2
KSZ	Nomenclatuur van socio-economische positie	Nomenc	Nomenclatuur tot op drie posities nauwkeurig voor elk gezinslid
		Bron: DWH_BCSS_NomenclatureVarDer	Per kwartaal van 2015Q1 – 2015Q4 Per jaar op 31-12-2010, ook voor 2011, 2012, 2013, 2014, 2015.
	Arbeidsvolume huishoudniveau	op Voltijds_equivalent_def_1 Voltijds_equivalent_def_2 Bron: DWH_BCSS_IntTravail	
	Werkintensiteit huishoudniveau	op WI_def_1 WI_def_2 Bron: DWH_BCSS_IntTravail	Van 0 tot en met 1, tot op twee decimalen nauwkeurig
	Inkomen	brutoloon_rsz loon_BB_rsz Bron: DWH_BCSS_Revenu Brutoloon_rszppo loon_BB_rszppo Bron: DWH_BCSS_Revenu inkomen_rsvz Bron: DWH_BCSS_Revenu	Individueel (op jaarbasis/12) in inkomensklassen (van 50 euro)

Inkomenscomponenten	Bruto belastbare uitkeringen		
	Uitkering_BB_fao		
	Uitkering_BB_fbz		
	Uitkering_BB_fod_sz		
	Uitkering_BB_nic		
	Uitkering_BB_pod_mi		
	Uitkering_BB_riziv		
	Uitkering_BB_rkw		
	Uitkering_BB_rsvz_kb		
	Uitkering_BB_rva		
	Uitkering_BB_rvp		
	Bron: DWH_BCSS_Revenu		
	Bruto uitkeringen		
	Uitkering_fao		
	Uitkering_foz_sz		
	Uitkering_rvp		
	Bron: DWH_BCSS_Revenu		
POD Maatschappelijke Integratie (POD MI)	Categorie van de leefloongerechtigde	CAT_CNT_ID Bron: DWH_SPPIS_Paiements	1. Alleenstaande 2. Persoon die met andere al dan niet verwante personen samenwoont 3. Persoon die uitsluitend met kinderen ten laste samenwoont
Rijks- en KSZ-register (RR-KSZ)	Geboortedatum	D_geboor Bron: DWH_RN_BCSS_BisTer	Leeftijdsklassen per vijf.
	Geslacht	Bron: DWH_RN_BCSS_BisTer	1. Man 2. Vrouw

Geslacht van de referentiepersoon	Gesl_ref Bron: DWH_RN_BCSS_BisTer	1. Man 2. Vrouw
Herkomst op basis van nationaliteit	Nation Bron: DWH_RN_BCSS_BisTer Eerste_natio Bron: DWH_RN_Origine Huidige_natio Bron: DWH_RN_Origine	
Herkomst van de ouders	Eerste_natio_ouder1 Huidige_natio_ouder1 Bron: DWH_RN_Origine Eerste_natio_ouder2 Huidige_natio_ouder2 Bron: DWH_RN_Origine	België; Buurlanden van België; West-Europa; Zuid-Europa; EU12; Centraal- en Oost-Europa; Turkije, Noord-Afrika; Centraal- en Zuid-Afrika; Noord-Amerika; Latijns-Amerika; Azië; Overige ²⁷
Positie in het gezin volgens de LIPRO-typologie	LIPRO Bron: DWH_RN_BCSS_BisTer	SING: alleenstaande MAR0: gehuwde zonder kind MAR+: gehuwde met kind CMAR: kind bij gehuwd paar UNM0: ongehuwd samenwonende zonder kind UNM+: ongehuwd samenwonende met kind(eren)

²⁷ Een gedetailleerde beschrijving van de clusters is te vinden in bijlage 9.3.

			CUNM: kind bij ongehuwd paar H1PA: Hoofd eenoudergezin C1PA: Kind in een eenoudergezin NFRP: Andere inwonende OTHR: Overige personen COLL: Wonend in een collectief huishouden
Verwantschapsrelatie met de referentiepersoon	met	Relatie Bron: DWH_RN_BCSS_BisTer	00 onbekend 01 referentiepersoon 02 echtgenoot, echtgenote 03 zoon, dochter 04 schoonzoon, schoondochter 05 kleinzoon, kleindochter 06 vader, moeder 07 schoonvader, schoonmoeder 08 grootvader, grootmoeder 09 broer, zus 10 schoonbroer, schoonzus 11 verwant(e) 12 zonder familieverband 13 stiefzoon, stiefdochter 14 achterkleinzoon of dochter 15 tante, oom 16 nicht, neef (3 ^e graad verwantschap) 17 nicht, neef (4 ^e graad verwantschap) 20 gemeenschappen, tehuizen
Aantal gezinsleden		Aantal Bron: DWH_RN_BCSS_BisTer	Aantal kinderen Aantal kinderen <25 jaar Aantal volwassenen in het gezin
Type huishouden		Type_huishouden	1. Gehuwd

	Bron: DWH_RN_BCSS_BisTer	2. Ongehuwd 3. Eenoudergezin 4. Andere
Regio	Woonplaats gemeente (Codnis) Bron: DWH_RN_BCSS_BisTer	Eenzijds twee NUTS-niveaus (Nomenclature of Territorial Units for Statistics): <ul style="list-style-type: none"> - Regionaal (NUTS 1): Vlaanderen en Brussel - Provinciaal (NUTS 2): Antwerpen, Limburg, Oost- en West-Vlaanderen, Vlaams-Brabant <p>Anderzijds de regionale VRIND-indeling: grootsteden, centrumsteden, grootstedelijke rand, stedelijke rand rond Brussel, regionaal stedelijke rand, structuurondersteunende steden, kleinstedelijk provinciaal, overgangsgebied, platteland.</p> <p>Zie 9.4</p>
Burgerlijke staat	Burgerlijke_staat Bron: DWH_RN_EtatCivil	10 ongehuwd 20 gehuwd 25 nietigverklaring van huwelijk 26 putatief huwelijk 30 weduwnaar/weduwe 40 echtgescheiden 41 echtgescheiden vanaf 1/10/1994 50 gescheiden van tafel en bed 51 gescheiden van tafel en bed vanaf 1/10/1994 60 verstoting 80 partnerschap 81 beëindiging partnerschap 90 onbepaald

Bijlagen

Rijksdienst voor de Sociale Zekerheid (RSZ)	Arbeidsprestatie	T_prest Bron: DWH_ONSS_StatbaseLATG	1. Voltijds 2. Deeltijds 3. Andere 4. Onbepaald
	Indicator arbeidsprestatie	R_exclus Bron: DWH_ONSSAPL_Statbase	Per kwartaal van 2015Q1-2015Q4
	Percentage deeltijds	Tauxpt Bron: DWH_ONSS_StatbaseDMFA	Percentage van een voltijdse job dat de deeltijdse werknemer presteert
Sigedis	Gepresteerde dagen	Aantal gepresteerde dagen Bron: DWH_SIGEDIS_Inscription	
VDAB – FOREM – Actiris - ADG	Studies	Niveau_etudes Bron: DWH_VDAB_FOREM_ACTIRIS_ADG	1. Lager onderwijs/geen diploma; 2. Secundair onderwijs; 3. Hoger onderwijs; 4. Andere

2 Afbakening herkomst

In het Datawarehouse Arbeidsmarkt en Sociale Bescherming (DWH AM&SB) vroegen we volgende landenclusters aan: België, buurlanden, West-Europa, Zuid-Europa, EU-12, Centraal- en Oost-Europa, Turkije, Noord-Afrika, Centraal- en Zuid-Afrika, Noord-Amerika, Latijns-Amerika, Azië, Nabije Oosten en overige. Tabel 27 geeft een overzicht van desbetreffende landenclusters, en hoe we de landen indelen in vier categorieën: personen met een Belgische, EU- of niet-EU-herkomst, of een restcategorie.

Tabel 26 Overzicht afbakening herkomstlanden (Bron: DWH AM&SB)

BE/EU/N-EU	Landencluster	Landen
BE	België	België
EU	Buurlanden	Duitsland, Frankrijk, Luxemburg, Nederland
	West-Europa	Denemarken, Finland, Ierland, IJsland, Liechtenstein, Noorwegen, Oostenrijk, Verenigd Koninkrijk, Zweden, Zwitserland
	Zuid-Europa	Griekenland, Italië, Portugal, Spanje
	EU12	Bulgarije, Cyprus, Estland, Hongarije, Letland, Litouwen, Malta, Polen, Roemenië, Slovenië, Slowakije, Tsjechië, Tsjecho-Slowakije
Niet-EU	Centraal- en Oost-Europa	Albanië, Bosnië-Herzegovina, Joegoslavië (ex), Joegoslavië (fed rep), Kroatië*, Macedonië, Moldavië, Oekraïne, Rusland, Sovjetunie (ex), USSR, Wit-Rusland
	Turkije	Turkije
	Noord-Afrika	Algerije, Egypte, Libië, Marokko, Tunesië
	Centraal- en Zuid-Afrika	Afars en Issas (FR), Angola, Angola (PT), Archipel van de Comoren, Azoren (PT), Belgisch Kongo, Benin, Bophuthatswana (ZA), Botswana, Burkina Faso, Burundi, Cabinda (AO), Canarische Eilanden, Centraal-Afrikaanse Republiek, Comoren, Congo, Congo (Democratische Republiek), Congo (Republiek), Djibouti, Eiland van Santhome (PT) Equatoriaal Guinea, Ethiopië, Eritrea, Fernando Poo (GQ), Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Ivoorkust, Kaapverdië, Kaapverdische Eilanden, Kameroen, Kenia, Lesotho, Liberia, Madagaskar, Madeira (PT), Malawi, Mali, Mauritanië, Mauritius, Mayotte (FR), Mozambique, Mozambique (PT), Namibië, Ngwame (Koninkrijk Swaziland) (SZ), Niger, Nigeria, Opper-Volta, Portugees Guinea, Réunion (FR), Rhodesië, Rwanda, Sahara (MR), Sao Tomé en Príncipe, Senegal, Senegambia,

	Seychellen, Sierra Leone, Sint Helena, Soedan, Somalië, Swaziland, Tanzania, Togo, Tsjaad, Transkei (ZA), Uganda, Urundi (BE), Zaire (Republiek), Zambia, Zimbabwe, Zuid-Afrika (Republiek)
Noord-Amerika	Bermuda, Canada, Groenland, Saint Pierre en Miquelon, Verenigde Staten
Latijns-Amerika	Anguilla (GB), Anguilla en Barbuda, Antillen (FR), Antillen (GB), Antillen (NL), Antillen (US), Argentinië, Bahama's, Bahama's (GB), Barbados, Belize, Belize (GB), Bolivia, Brazilië, Brits Honduras, Chili, Colombia, Cuba, Curaçao (NL), Dominicaanse Republiek, Dominica, Dominica (GB), Ecuador, El Salvador, Falklandeilanden (GB), Frans-Guyana, Grenada, Guadeloupe (FR), Guatemala, Guyana, Haïti, Honduras, Jamaica, Kaaiman Eilanden (GB), Kitts en Nevis (GB), Maagdeneilanden (GB), Martinique (FR), Mexico, Montserrat (GB), Nederlands-Guyana, Nicaragua, Panama, Paraguay, Peru, Porto Rico (US), Sint Kitts en Nevis, Sint Lucia, Sint Vincent, Suriname, Trinidad en Tobago, Turks- en Caicoseilanden (GB), Uruguay, Venezuela
Azië	Bangladesh, Bhutan, Brunei, Cambodja, China, Filipijnen, Hong Kong, Hong Kong (GB), Indië, Indonesië, Japan, Kazachstan, Khmerische Republiek Cambodja, Laos, Macao (PT), Maldiven, Maleisië, Mongolië, Myanmar, Nepal, Noord-Korea, Singapore, Sri Lanka, Taiwan, Thailand, Timor (ID), Vietnam, Zuid-Korea, Zuid-Vietnam
Nabije oosten	Kirgizië, Oezbekistan, Tadzjikistan, Turkmenistan, Armenië, Azerbeidzjan, Afghanistan, Saoedi Arabië, Georgië, Irak, Iran, Israël, Jordanië, Libanon, Pakistan, Verenigde Emiraten, Syrië, Koeweit, Yemen, Oman, Qatar, Abu Dhabi, Bahrein, Palestina
Overige of niet beschikbaar	Overige Amerikaanse onderhorigheden, Amerikaans Samoa, Australië, Australische onderhorigheden, Carolinen (US), Cook (NZ), Fiji, Frans-Polynesië, Guam (US), Hawaii (US), Kiribati, Nauru, Nieuw-Caledonië (FR), Nieuwe Hebriden (GB), Nieuw-Zeeland, Nieuw-Zeelandse onderhorigheden, Niue (NZ), Noord-Marianen, Palau (US), Papoea-Nieuw-Guinea, Pitcairn (GB), Salomon, Salomon (GB), Stille Oceaan (eilanden), Tahiti (FR), Tokelau (NZ), Tonga, Tuvalu, Vanuatu, Wallis en Futuna (FR), West-Samoa Politiek vluchteling, Vaderlandloos, Onbekend

*Let wel, de afbakening die men in het Methodologisch rapport van Audenaert and Stuyck (2018a) wijkt licht af van onze afbakening. Zij konden Kroatië namelijk toevoegen aan personen met een EU-herkomst. Wij kunnen geen individueel land (Kroatië) uit de landencluster (Centraal- en Oost-Europa) halen, waardoor in dit geval personen afkomstig van Kroatië behoren tot personen met een niet-EU-herkomst.

3 Overzicht construeren van gewichten (Kruistabel 1)

Stap 1: Originale combinaties gezinsenquête

Aantal combinaties	Herkomst	Gezinstype	Aantal kinderen 0-11	Aantal kinderen 12-24	Totale aantal kinderen	Aantal gezinnen in Gezins-enquête
1	Belg	(on)gehuwd koppel met kinderen	0	1	1	205
2	Belg	(on)gehuwd koppel met kinderen	0	2	2	256
3	Belg	(on)gehuwd koppel met kinderen	0	3	3	93
4	Belg	(on)gehuwd koppel met kinderen	1	0	1	114
5	Belg	(on)gehuwd koppel met kinderen	1	1	2	67
6	Belg	(on)gehuwd koppel met kinderen	1	2	3	37
7	Belg	(on)gehuwd koppel met kinderen	1	3	4	10
8	Belg	(on)gehuwd koppel met kinderen	2	0	2	205
9	Belg	(on)gehuwd koppel met kinderen	2	1	3	29
10	Belg	(on)gehuwd koppel met kinderen	2	2	4	13
11	Belg	(on)gehuwd koppel met kinderen	2	3	5	1
12	Belg	(on)gehuwd koppel met kinderen	3	0	3	75
13	Belg	(on)gehuwd koppel met kinderen	3	1	4	3
14	Belg	(on)gehuwd koppel met kinderen	3	3	6	1
15	Belg	eenoudergezinnen	0	1	1	152
16	Belg	eenoudergezinnen	0	2	2	115
17	Belg	eenoudergezinnen	0	3	3	23
18	Belg	eenoudergezinnen	1	0	1	140
19	Belg	eenoudergezinnen	1	1	2	36
20	Belg	eenoudergezinnen	1	2	3	11
21	Belg	eenoudergezinnen	1	3	4	10
22	Belg	eenoudergezinnen	2	0	2	146

23	Belg	eenoudergezinnen	2	1	3	22
24	Belg	eenoudergezinnen	2	2	4	2
25	Belg	eenoudergezinnen	2	3	5	2
26	Belg	eenoudergezinnen	3	0	3	22
27	Belg	eenoudergezinnen	3	1	4	3
28	Belg	eenoudergezinnen	3	3	6	1
29	EU	(on)gehuwd koppel met kinderen	0	1	1	16
30	EU	(on)gehuwd koppel met kinderen	0	2	2	14
31	EU	(on)gehuwd koppel met kinderen	0	3	3	7
32	EU	(on)gehuwd koppel met kinderen	1	0	1	8
33	EU	(on)gehuwd koppel met kinderen	1	1	2	7
34	EU	(on)gehuwd koppel met kinderen	1	2	3	1
35	EU	(on)gehuwd koppel met kinderen	1	3	4	1
36	EU	(on)gehuwd koppel met kinderen	2	0	2	10
37	EU	(on)gehuwd koppel met kinderen	2	1	3	1
38	EU	(on)gehuwd koppel met kinderen	3	0	3	4
39	EU	(on)gehuwd koppel met kinderen	3	2	5	1
40	EU	eenoudergezinnen	0	1	1	12
41	EU	eenoudergezinnen	0	2	2	9
42	EU	eenoudergezinnen	1	0	1	14
43	EU	eenoudergezinnen	1	1	2	6
44	EU	eenoudergezinnen	1	2	3	3
45	EU	eenoudergezinnen	1	3	4	2
46	EU	eenoudergezinnen	2	0	2	16
47	EU	eenoudergezinnen	2	1	3	1
48	EU	eenoudergezinnen	3	0	3	2
49	EU	eenoudergezinnen	3	1	4	1
50	Niet-EU	(on)gehuwd koppel met kinderen	0	1	1	42
51	Niet-EU	(on)gehuwd koppel met kinderen	0	2	2	34
52	Niet-EU	(on)gehuwd koppel met kinderen	0	3	3	25

53	Niet-EU	(on)gehuwd koppel met kinderen	1	0	1	94
54	Niet-EU	(on)gehuwd koppel met kinderen	1	1	2	33
55	Niet-EU	(on)gehuwd koppel met kinderen	1	2	3	16
56	Niet-EU	(on)gehuwd koppel met kinderen	1	3	4	12
57	Niet-EU	(on)gehuwd koppel met kinderen	2	0	2	94
58	Niet-EU	(on)gehuwd koppel met kinderen	2	1	3	12
59	Niet-EU	(on)gehuwd koppel met kinderen	2	2	4	8
60	Niet-EU	(on)gehuwd koppel met kinderen	2	3	5	7
61	Niet-EU	(on)gehuwd koppel met kinderen	3	0	3	60
62	Niet-EU	(on)gehuwd koppel met kinderen	3	1	4	11
63	Niet-EU	(on)gehuwd koppel met kinderen	3	2	5	4
64	Niet-EU	(on)gehuwd koppel met kinderen	3	3	6	2
65	Niet-EU	eenoudergezinnen	0	1	1	27
66	Niet-EU	eenoudergezinnen	0	2	2	15
67	Niet-EU	eenoudergezinnen	0	3	3	7
68	Niet-EU	eenoudergezinnen	1	0	1	48
69	Niet-EU	eenoudergezinnen	1	1	2	22
70	Niet-EU	eenoudergezinnen	1	2	3	5
71	Niet-EU	eenoudergezinnen	1	3	4	2
72	Niet-EU	eenoudergezinnen	2	0	2	40
73	Niet-EU	eenoudergezinnen	2	1	3	7
74	Niet-EU	eenoudergezinnen	2	2	4	4
75	Niet-EU	eenoudergezinnen	2	3	5	1
76	Niet-EU	eenoudergezinnen	3	0	3	11
77	Niet-EU	eenoudergezinnen	3	1	4	1
78	Niet-EU	eenoudergezinnen	3	3	6	2
79	Missing	(on)gehuwd koppel met kinderen	0	1	1	5
80	Missing	(on)gehuwd koppel met kinderen	0	2	2	3
81	Missing	(on)gehuwd koppel met kinderen	0	3	3	6

82	Missing	(on)gehuwd koppel met kinderen	1	0	1	3
83	Missing	(on)gehuwd koppel met kinderen	1	1	2	3
84	Missing	(on)gehuwd koppel met kinderen	1	2	3	1
85	Missing	(on)gehuwd koppel met kinderen	2	0	2	2
86	Missing	(on)gehuwd koppel met kinderen	2	1	3	1
87	Missing	(on)gehuwd koppel met kinderen	3	0	3	1
88	Missing	(on)gehuwd koppel met kinderen	3	3	6	1
89	Missing	eenoudergezinnen	0	1	1	9
90	Missing	eenoudergezinnen	0	2	2	1
91	Missing	eenoudergezinnen	1	0	1	1
92	Missing	eenoudergezinnen	1	2	3	1
93	Missing	eenoudergezinnen	2	0	2	2

Stap 4: Originale combinaties populatie

Afkomst	Gezinstype	0-11	12-24	Aantal kinderen	Aantal gezinnen
Belg	(on)gehuwd paar met kinderen	0	1	1	119843
Belg	(on)gehuwd paar met kinderen	0	2	2	104098
Belg	(on)gehuwd paar met kinderen	0	3	3	28275
Belg	(on)gehuwd paar met kinderen	1	0	1	101974
Belg	(on)gehuwd paar met kinderen	1	1	2	35955
Belg	(on)gehuwd paar met kinderen	1	2	3	14246
Belg	(on)gehuwd paar met kinderen	1	3	6	3339
Belg	(on)gehuwd paar met kinderen	2	0	2	120007
Belg	(on)gehuwd paar met kinderen	2	1	3	10473
Belg	(on)gehuwd paar met kinderen	2	2	4	2863
Belg	(on)gehuwd paar met kinderen	2	3	6	754
Belg	(on)gehuwd paar met kinderen	3	0	5	27493
Belg	(on)gehuwd paar met kinderen	3	1	7	2286
Belg	(on)gehuwd paar met kinderen	3	2	6	629
Belg	(on)gehuwd paar met kinderen	3	3	8	357
Belg	eenoudergezinnen	0	1	1	48120
Belg	eenoudergezinnen	0	2	2	24869
Belg	eenoudergezinnen	0	3	5	5642
Belg	eenoudergezinnen	1	0	1	22424

Belg	eenoudergezinnen	1	1	2	7077
Belg	eenoudergezinnen	1	2	3	2529
Belg	eenoudergezinnen	1	3	5	626
Belg	eenoudergezinnen	2	0	2	11126
Belg	eenoudergezinnen	2	1	3	1527
Belg	eenoudergezinnen	2	2	4	473
Belg	eenoudergezinnen	2	3	6	113
Belg	eenoudergezinnen	3	0	4	2038
Belg	eenoudergezinnen	3	1	6	340
Belg	eenoudergezinnen	3	2	6	109
Belg	eenoudergezinnen	3	3	8	53
EU	(on)gehuwd paar met kinderen	0	1	1	10973
EU	(on)gehuwd paar met kinderen	0	2	2	9380
EU	(on)gehuwd paar met kinderen	0	3	4	2621
EU	(on)gehuwd paar met kinderen	1	0	1	16171
EU	(on)gehuwd paar met kinderen	1	1	2	5431
EU	(on)gehuwd paar met kinderen	1	2	3	2134
EU	(on)gehuwd paar met kinderen	1	3	5	700
EU	(on)gehuwd paar met kinderen	2	0	2	15706
EU	(on)gehuwd paar met kinderen	2	1	3	1895
EU	(on)gehuwd paar met kinderen	2	2	4	742
EU	(on)gehuwd paar met kinderen	2	3	5	291
EU	(on)gehuwd paar met kinderen	3	0	3	4874
EU	(on)gehuwd paar met kinderen	3	1	4	742
EU	(on)gehuwd paar met kinderen	3	2	6	281
EU	(on)gehuwd paar met kinderen	3	3	8	223
EU	eenoudergezinnen	0	1	1	5256
EU	eenoudergezinnen	0	2	2	2587
EU	eenoudergezinnen	0	3	4	615
EU	eenoudergezinnen	1	0	1	3872
EU	eenoudergezinnen	1	1	2	1210
EU	eenoudergezinnen	1	2	3	405
EU	eenoudergezinnen	1	3	5	129
EU	eenoudergezinnen	2	0	2	1765
EU	eenoudergezinnen	2	1	3	349
EU	eenoudergezinnen	2	2	4	106
EU	eenoudergezinnen	2	3	6	50
EU	eenoudergezinnen	3	0	3	438
EU	eenoudergezinnen	3	1	7	121
EU	eenoudergezinnen	3	2	5	52
EU	eenoudergezinnen	3	3	6	18
Niet-EU	(on)gehuwd paar met kinderen	0	1	1	7436
Niet-EU	(on)gehuwd paar met kinderen	0	2	2	7270

Niet-EU	(on)gehuwd paar met kinderen	0	3	3	4839
Niet-EU	(on)gehuwd paar met kinderen	1	0	1	14490
Niet-EU	(on)gehuwd paar met kinderen	1	1	2	4859
Niet-EU	(on)gehuwd paar met kinderen	1	2	3	4528
Niet-EU	(on)gehuwd paar met kinderen	1	3	7	2820
Niet-EU	(on)gehuwd paar met kinderen	2	0	2	17705
Niet-EU	(on)gehuwd paar met kinderen	2	1	3	4199
Niet-EU	(on)gehuwd paar met kinderen	2	2	4	2464
Niet-EU	(on)gehuwd paar met kinderen	2	3	7	1282
Niet-EU	(on)gehuwd paar met kinderen	3	0	7	12290
Niet-EU	(on)gehuwd paar met kinderen	3	1	5	2587
Niet-EU	(on)gehuwd paar met kinderen	3	2	5	1027
Niet-EU	(on)gehuwd paar met kinderen	3	3	9	708
Niet-EU	eenoudergezinnen	0	1	1	5510
Niet-EU	eenoudergezinnen	0	2	2	3015
Niet-EU	eenoudergezinnen	0	3	7	1398
Niet-EU	eenoudergezinnen	1	0	1	5838
Niet-EU	eenoudergezinnen	1	1	2	1844
Niet-EU	eenoudergezinnen	1	2	3	1060
Niet-EU	eenoudergezinnen	1	3	4	497
Niet-EU	eenoudergezinnen	2	0	2	3343
Niet-EU	eenoudergezinnen	2	1	3	893
Niet-EU	eenoudergezinnen	2	2	4	408
Niet-EU	eenoudergezinnen	2	3	5	156
Niet-EU	eenoudergezinnen	3	0	4	1610
Niet-EU	eenoudergezinnen	3	1	6	377
Niet-EU	eenoudergezinnen	3	2	8	152
Niet-EU	eenoudergezinnen	3	3	9	79
Missing	(on)gehuwd paar met kinderen	0	1	1	101
Missing	(on)gehuwd paar met kinderen	0	2	2	93
Missing	(on)gehuwd paar met kinderen	0	3	5	34
Missing	(on)gehuwd paar met kinderen	1	0	1	232
Missing	(on)gehuwd paar met kinderen	1	1	2	70
Missing	(on)gehuwd paar met kinderen	1	2	3	46
Missing	(on)gehuwd paar met kinderen	1	3	4	24
Missing	(on)gehuwd paar met kinderen	2	0	2	254
Missing	(on)gehuwd paar met kinderen	2	1	3	33
Missing	(on)gehuwd paar met kinderen	2	2	4	23
Missing	(on)gehuwd paar met kinderen	2	3	8	21
Missing	(on)gehuwd paar met kinderen	3	0	4	139
Missing	(on)gehuwd paar met kinderen	3	1	4	26
Missing	(on)gehuwd paar met kinderen	3	2	7	15
Missing	(on)gehuwd paar met kinderen	3	3	12	19

Missing	eenoudergezinnen	0	1	1	81
Missing	eenoudergezinnen	0	2	2	46
Missing	eenoudergezinnen	0	3	3	9
Missing	eenoudergezinnen	1	0	1	81
Missing	eenoudergezinnen	1	1	2	23
Missing	eenoudergezinnen	1	2	3	10
Missing	eenoudergezinnen	1	3	5	4
Missing	eenoudergezinnen	2	0	2	52
Missing	eenoudergezinnen	2	1	3	5
Missing	eenoudergezinnen	2	2	4	6
Missing	eenoudergezinnen	2	3	5	4
Missing	eenoudergezinnen	3	0	3	17
Missing	eenoudergezinnen	3	1	4	4
Missing	eenoudergezinnen	3	2	5	1
Missing	eenoudergezinnen	3	3	7	2

Stap 6: Construeren van de gewichten

Aantal combi-naties	afkomst	Gezinstype	0-11	12-24	Aantal kinderen	Aantal gezinnen GE	Aantal gezinnen populatie	Coëfficiënt Gezinsenquête	Coëfficiënt populatie	Gewichten
1	Belg	(on)gehuwd paar met kinderen	0	1	1	205	119843	0,0784	0,13243	1,68994
2	Belg	(on)gehuwd paar met kinderen	0	2	2	256	104098	0,0979	0,11503	1,17548
3	Belg	(on)gehuwd paar met kinderen	0	3	3	93	28275	0,0356	0,03124	0,87888
4	Belg	(on)gehuwd paar met kinderen	1	0	1	114	101974	0,0436	0,11268	2,58581
5	Belg	(on)gehuwd paar met kinderen	1	1	2	67	35955	0,0256	0,03973	1,55130
6	Belg	(on)gehuwd paar met kinderen	1	2	3	37	14246	0,0141	0,01574	1,11302
7	Belg	(on)gehuwd paar met kinderen	1	3	6	10	3339	0,0038	0,00369	0,96522
8	Belg	(on)gehuwd paar met kinderen	2	0	2	205	120007	0,0784	0,13261	1,69225
9	Belg	(on)gehuwd paar met kinderen	2	1	3	29	10473	0,0111	0,01157	1,04396
10	Belg	(on)gehuwd paar met kinderen	2	2	4	15	3974	0,0057	0,00439	0,76586
11	Belg	(on)gehuwd paar met kinderen	3	0	5	78	30408	0,0298	0,03360	1,12695
12	Belg	eenoudergezinnen	0	1	1	152	48120	0,0581	0,05317	0,91515
13	Belg	eenoudergezinnen	0	2	2	115	24869	0,0440	0,02748	0,62513
14	Belg	eenoudergezinnen	0	3	5	23	5642	0,0088	0,00623	0,70911
15	Belg	eenoudergezinnen	1	0	1	140	22424	0,0535	0,02478	0,46302
16	Belg	eenoudergezinnen	1	1	2	36	7077	0,0138	0,00782	0,56827
17	Belg	eenoudergezinnen	1	2	3	11	2529	0,0042	0,00279	0,66461
18	Belg	eenoudergezinnen	1	3	5	13	626	0,0050	0,00069	0,13920
19	Belg	eenoudergezinnen	2	0	2	146	11126	0,0558	0,01229	0,22029
20	Belg	eenoudergezinnen	2	1	3	24	2166	0,0092	0,00239	0,26089
21	Belg	eenoudergezinnen	3	0	4	25	2487	0,0096	0,00275	0,28757
22	EU	(on)gehuwd paar met kinderen	0	1	1	24	27144	0,0092	0,02999	3,26944

Bijlagen

23	EU	(on)gehuwd paar met kinderen	0	2	2	30	20266	0,0115	0,02239	1,95280
24	EU	(on)gehuwd paar met kinderen	2	0	2	16	24754	0,0061	0,02735	4,47236
25	EU	eenoudergezinnen	0	1	1	21	8458	0,0080	0,00935	1,16429
26	EU	eenoudergezinnen	1	0	1	25	5616	0,0096	0,00621	0,64938
27	EU	eenoudergezinnen	2	0	2	20	2899	0,0076	0,00320	0,41901
28	Niet-EU	(on)gehuwd paar met kinderen	0	1	1	42	7436	0,0161	0,00822	0,51180
29	Niet-EU	(on)gehuwd paar met kinderen	0	2	2	34	7270	0,0130	0,00803	0,61811
30	Niet-EU	(on)gehuwd paar met kinderen	0	3	3	25	4839	0,0096	0,00535	0,55953
31	Niet-EU	(on)gehuwd paar met kinderen	1	0	1	94	14490	0,0359	0,01601	0,44561
32	Niet-EU	(on)gehuwd paar met kinderen	1	1	2	33	4859	0,0126	0,00537	0,42564
33	Niet-EU	(on)gehuwd paar met kinderen	1	2	3	16	4528	0,0061	0,00500	0,81808
34	Niet-EU	(on)gehuwd paar met kinderen	1	3	7	19	4102	0,0073	0,00453	0,62410
35	Niet-EU	(on)gehuwd paar met kinderen	2	0	2	94	17705	0,0359	0,01956	0,54448
36	Niet-EU	(on)gehuwd paar met kinderen	2	1	3	20	6663	0,0076	0,00736	0,96305
37	Niet-EU	(on)gehuwd paar met kinderen	3	0	7	60	12290	0,0229	0,01358	0,59212
38	Niet-EU	(on)gehuwd paar met kinderen	3	1	5	17	4322	0,0065	0,00478	0,73493
39	Niet-EU	eenoudergezinnen	0	1	1	27	5510	0,0103	0,00609	0,58993
40	Niet-EU	eenoudergezinnen	0	2	2	22	4413	0,0084	0,00488	0,57986
41	Niet-EU	eenoudergezinnen	1	0	1	48	5838	0,0183	0,00645	0,35159
42	Niet-EU	eenoudergezinnen	1	1	2	29	3401	0,0111	0,00376	0,33902
43	Niet-EU	eenoudergezinnen	2	0	2	51	4800	0,0195	0,00530	0,27207
44	Niet-EU	eenoudergezinnen	3	0	4	15	2218	0,0057	0,00245	0,42745
45	.	(on)gehuwd paar met kinderen	0	1	1	26	1130	0,0099	0,00125	0,12564
46	.	eenoudergezinnen	0	1	1	14	345	0,0054	0,00038	0,07124
							Totaal:	Totaal:		
							2616	904.954		

4 Overzicht construeren van gewichten (Kruistabel 2)

Stap 1: Originale combinaties in de gezinsenquête

Aantal combinaties	Herkomst	Gezinstype	Beroepsstatuut	Aantal gezinnen gezinsenquête
1	Belg	(on)gehuwd paar met kinderen	Andere	38
2	Belg	(on)gehuwd paar met kinderen	Arbeidsongeschikt	30
3	Belg	(on)gehuwd paar met kinderen	Werkloos (leefloon, tijdkrediet, werkzoekende)	40
4	Belg	(on)gehuwd paar met kinderen	Pensioen	21
5	Belg	(on)gehuwd paar met kinderen	Zelfstandige	157
6	Belg	(on)gehuwd paar met kinderen	Werknemer	1186
7	Belg	(on)gehuwd paar met kinderen	.	2
8	Belg	eenoudergezinnen	Andere	2
9	Belg	eenoudergezinnen	Arbeidsongeschikt	20
10	Belg	eenoudergezinnen	Werkloos	13
11	Belg	eenoudergezinnen	Pensioen	17
12	Belg	eenoudergezinnen	Zelfstandige	20
13	Belg	eenoudergezinnen	Werknemer	248
14	EU	(on)gehuwd paar met kinderen	Andere	4
15	EU	(on)gehuwd paar met kinderen	Arbeidsongeschikt	4
16	EU	(on)gehuwd paar met kinderen	Werkloos	13
17	EU	(on)gehuwd paar met kinderen	Pensioen	1
18	EU	(on)gehuwd paar met kinderen	Zelfstandige	12
19	EU	(on)gehuwd paar met kinderen	Werknemer	65
20	EU	eenoudergezinnen	Arbeidsongeschikt	1
21	EU	eenoudergezinnen	Werkloos	3
22	EU	eenoudergezinnen	Pensioen	2
23	EU	eenoudergezinnen	Zelfstandige	3
24	EU	eenoudergezinnen	Werknemer	27
25	EU	eenoudergezinnen	.	1
26	Niet-EU	(on)gehuwd paar met kinderen	Andere	71
27	Niet-EU	(on)gehuwd paar met kinderen	Arbeidsongeschikt	27
28	Niet-EU	(on)gehuwd paar met kinderen	Werkloos	75
29	Niet-EU	(on)gehuwd paar met kinderen	Pensioen	8
30	Niet-EU	(on)gehuwd paar met kinderen	Zelfstandige	38
31	Niet-EU	(on)gehuwd paar met kinderen	Werknemer	301
32	Niet-EU	(on)gehuwd paar met kinderen	.	3
33	Niet-EU	eenoudergezinnen	Andere	7
34	Niet-EU	eenoudergezinnen	Arbeidsongeschikt	9
35	Niet-EU	eenoudergezinnen	Werkloos	23

36	Niet-EU	eenoudergezinnen	Pensioen	4
37	Niet-EU	eenoudergezinnen	Zelfstandige	5
38	Niet-EU	eenoudergezinnen	Werknemer	75
39	.	(on)gehuwd paar met kinderen	Andere	2
40	.	(on)gehuwd paar met kinderen	Arbeidsongeschikt	1
41	.	(on)gehuwd paar met kinderen	Werkloos	1
42	.	(on)gehuwd paar met kinderen	Zelfstandige	3
43	.	(on)gehuwd paar met kinderen	Werknemer	21
44	.	(on)gehuwd paar met kinderen	.	2
45	.	eenoudergezinnen	Andere	1
46	.	eenoudergezinnen	Arbeidsongeschikt	1
47	.	eenoudergezinnen	Werknemer	8

Stap 4: Originale combinaties in de populatiegegevens

Aantal combinaties	Herkomst	Gezinstype	Beroepsstatuut	Aantal gezinnen in de populatie
1	Belg	(on)gehuwd paar met kinderen	Andere	14842
2	Belg	(on)gehuwd paar met kinderen	Arbeidsongeschikt	8520
3	Belg	(on)gehuwd paar met kinderen	Werkloos (leefloon, tijdkrediet, werkzoekende)	3577
4	Belg	(on)gehuwd paar met kinderen	Pensioen	52935
5	Belg	(on)gehuwd paar met kinderen	Zelfstandige	135080
6	Belg	(on)gehuwd paar met kinderen	Werknemer	438677
7	Belg	eenoudergezinnen	Andere	4631
8	Belg	eenoudergezinnen	Arbeidsongeschikt	8567
9	Belg	eenoudergezinnen	Werkloos	3131
10	Belg	eenoudergezinnen	Pensioen	40914
11	Belg	eenoudergezinnen	Zelfstandige	21898
12	Belg	eenoudergezinnen	Werknemer	97533
13	EU	(on)gehuwd paar met kinderen	Andere	13901
14	EU	(on)gehuwd paar met kinderen	Arbeidsongeschikt	1847
15	EU	(on)gehuwd paar met kinderen	Werkloos	911
16	EU	(on)gehuwd paar met kinderen	Pensioen	3479
17	EU	(on)gehuwd paar met kinderen	Zelfstandige	16656
18	EU	(on)gehuwd paar met kinderen	Werknemer	41012
19	EU	eenoudergezinnen	Andere	2850
20	EU	eenoudergezinnen	Arbeidsongeschikt	1326
21	EU	eenoudergezinnen	Werkloos	885
22	EU	eenoudergezinnen	Pensioen	1909
23	EU	eenoudergezinnen	Zelfstandige	3223
24	EU	eenoudergezinnen	Werknemer	9623
25	Niet-EU	(on)gehuwd paar met kinderen	Andere	9177
26	Niet-EU	(on)gehuwd paar met kinderen	Arbeidsongeschikt	5113

27	Niet-EU	(on)gehuwd kinderen	paar	met	Werkloos	3356
28	Niet-EU	(on)gehuwd kinderen	paar	met	Pensioen	4818
29	Niet-EU	(on)gehuwd kinderen	paar	met	Zelfstandige	19674
30	Niet-EU	(on)gehuwd kinderen	paar	met	Werknemer	50263
31	Niet-EU	eenoudergezinnen			Andere	2578
32	Niet-EU	eenoudergezinnen			Arbeidsongeschikt	2420
33	Niet-EU	eenoudergezinnen			Werkloos	3984
34	Niet-EU	eenoudergezinnen			Pensioen	2269
35	Niet-EU	eenoudergezinnen			Zelfstandige	5135
36	Niet-EU	eenoudergezinnen			Werknemer	12037
37		(on)gehuwd kinderen	paar	met	Andere	157
38		(on)gehuwd kinderen	paar	met	Arbeidsongeschikt	32
39		(on)gehuwd kinderen	paar	met	Werkloos	71
40		(on)gehuwd kinderen	paar	met	Pensioen	44
41		(on)gehuwd kinderen	paar	met	Zelfstandige	275
42		(on)gehuwd kinderen	paar	met	Werknemer	616
43		eenoudergezinnen			Andere	48
44		eenoudergezinnen			Arbeidsongeschikt	21
45		eenoudergezinnen			Werkloos	46
46		eenoudergezinnen			Pensioen	29
47		eenoudergezinnen			Zelfstandige	72
48		eenoudergezinnen			Werknemer	171
Totaal:						1.050.333

Stap 6: Construeren van weegcoëfficiënten

	Herkomst	Gezinstype	Beroeps-statuu	Aantal gezin- nen gezins- enquête	Aantal gezin- nen populatie	Coëf- ficiënt gezin- senquête	Coëf- ficiënt popu- latie	Gewicht
1	Belg	(on)gehuwd paar met kinderen	Andere	40	14842	0,01529	0,01413	0,924
2	Belg	(on)gehuwd paar met kinderen	Arbeidsongeschikt	30	8520	0,01147	0,00811	0,707
3	Belg	(on)gehuwd paar met kinderen	Werkloos (leefloon, tijdkrediet, werkzoekende)	40	3577	0,01529	0,00341	0,223
4	Belg	(on)gehuwd paar met kinderen	Pensioen	21	52935	0,00803	0,05040	6,278
5	Belg	(on)gehuwd paar met kinderen	Zelfstandige	157	135080	0,06002	0,12861	2,143
6	Belg	(on)gehuwd paar met kinderen	Werknemer	1180	438677	0,45336	0,41766	0,921
7	Belg	eenoudergezinnen	Arbeidsongeschikt	20	8567	0,00765	0,00816	1,067
8	Belg	eenoudergezinnen	Werkloos	15	7762	0,00573	0,00739	1,289
9	Belg	eenoudergezinnen	Pensioen	17	40914	0,00650	0,03895	5,994
10	Belg	eenoudergezinnen	Zelfstandige	20	21898	0,00765	0,02085	2,727
11	Belg	eenoudergezinnen	Werknemer	248	97533	0,09480	0,09286	0,980
12	EU	(on)gehuwd paar met kinderen	Werkloos	22	20138	0,00841	0,01917	2,280
13	EU	(on)gehuwd paar met kinderen	Zelfstandige	12	16656	0,00459	0,01586	3,457
14	EU	(on)gehuwd paar met kinderen	Werknemer	65	41012	0,02485	0,03905	1,571
15	EU	eenoudergezinnen	Werkloos	10	6970	0,00268	0,00664	2,480
16	EU	eenoudergezinnen	Werknemer	30	12846	0,01147	0,01223	1,066
17	Niet-EU	(on)gehuwd paar met kinderen	Andere	74	9177	0,02829	0,00874	0,309
18	Niet-EU	(on)gehuwd paar met kinderen	Arbeidsongeschikt	27	5113	0,01032	0,00487	0,472
19	Niet-EU	(on)gehuwd paar met kinderen	Werkloos	83	8174	0,03173	0,00778	0,245
20	Niet-EU	(on)gehuwd paar met kinderen	Zelfstandige	38	19674	0,01453	0,01873	1,289
21	Niet-EU	(on)gehuwd paar met kinderen	Werknemer	301	50263	0,11506	0,04785	0,416
22	Niet-EU	eenoudergezinnen	Werkloos	43	11251	0,01644	0,01071	0,652

Bijlagen

23	Niet-EU	eenoudergezinnen	Werknemer	80	17172	0,03058	0,01635	0,535
24	.	(on)gehuwd paar met kinderen	Werkloos	10	304	0,00229	0,00029	0,126
25	.	(on)gehuwd paar met kinderen	Werknemer	24	891	0,00917	0,00085	0,092
26	.	eenoudergezinnen	Werknemer	10	387	0,00382	0,00037	0,096
				Totaal:	Totaal:			
				2616	1050333			

