

Haalbaarheidsstudie loopbaanpanel

Technisch rapport

Dorien Frans
Dimitri Mortelmans
Centrum voor Longitudinaal en Levenslooponderzoek (Cello)
Universiteit Antwerpen

Luc Sels
Steunpunt Werk en Sociale Economie
Faculteit Economie en Bedrijfswetenschappen, K.U.Leuven

Anneleen Forrier
Lessius Hogeschool

10-2009

WSE Report

Steunpunt Werk en Sociale Economie
Parkstraat 45 bus 5303 – 3000 Leuven
T:32(0)16 32 32 39 F:32(0)16 32 32 40
steunpuntwse@econ.kuleuven.be
www.steunpuntwse.be





Haalbaarheidsstudie loopbaanpanel. Technisch rapport

Dorien Frans
Dimitri Mortelmans
Luc Sels
Anneleen Forrier

Een onderzoek in opdracht van de Vlaamse minister van Werk, Onderwijs en Vorming, in het kader van het VIONA-onderzoeksprogramma

<p>Deze publicatie kwam tot stand met steun van het Europees Sociaal Fonds. Het ESF stelt middelen ter beschikking voor initiatieven die bijdragen tot meer en betere jobs voor meer mensen.</p> <p>Ontdek de werking in Vlaanderen via www.esf-agentschap.be.</p>	<p>Kernthema's ESF 2007-2013</p> <ul style="list-style-type: none"> Talenten activeren Arbeidskansen geven Ondernemen met mensen
	<p>ESF investeert in jouw toekomst.</p> 

Frans, Dorien, Mortelmans, Dimitri, Sels, Luc & Forrier, Anneleen.

Haalbaarheidsstudie loopbaanpanel. Technisch rapport.

Dorien Frans, Dimitri Mortelmans, Luc Sels & Anneleen Forrier – Leuven: Katholieke Universiteit Leuven. Steunpunt Werk en Sociale Economie / Antwerpen: CELLO, Universiteit Antwerpen, 2009, 55p.

ISBN-97 890-8873-031-3

Copyright (2009)

Steunpunt Werk en Sociale Economie
Parkstraat 45 bus 5303 – B-3000 Leuven
T:32(0)16 32 32 39 - F:32(0)16 32 32 40
steunpuntwse@econ.kuleuven.be
www.steunpuntwse.be
Universiteit Antwerpen
Centrum voor Longitudinaal en Levensloonderzoek (Cello)
Stadscampus - Sint Jacobstraat 2 - 2000 Antwerpen

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

No part of this report may be reproduced in any form, by mimeograph, film or any other means, without permission in writing from the publisher.

Inhoud

1	Een haalbaarheidsstudie voor toekomstig loopbaanonderzoek	1
2	Mogelijke scenario's voor een loopbaanpanel	2
2.1	Scenario 1 en 2: een zuiver loopbaanpanel.....	2
2.1.1	Voordelen.....	3
2.1.2	Nadelen.....	4
2.2	Scenario 3: opeenvolgende cross-secties	6
2.2.1	Voordelen.....	7
2.2.2	Nadelen.....	8
2.3	Scenario 4: focus op administratieve data	9
2.3.1	Voordelen.....	10
2.3.2	Nadelen.....	11
2.4	Scenario 5: administratieve data gekoppeld aan cross-sectie.....	12
2.4.1	Voordelen.....	13
2.4.2	Nadelen.....	13
2.5	Scenario 6: administratieve data gekoppeld aan panelstudie	15
2.5.1	Voordelen.....	15
2.5.2	Nadelen.....	16
2.6	Scenario 7. loopbaan-cohortenstudie (sonar-scenario).....	16
2.6.1	Voordelen.....	17
2.6.2	Nadelen.....	17
2.7	Samenvatting: overzicht van de mogelijke scenario's	18
3	Haalbaarheidsstudie voor Scenario 5, 6 en 7.....	20
3.1	Methodologische dimensie.....	21
3.1.1	Steekproefdesign	21
3.1.2	Steekproefomvang.....	23
3.1.3	Steekproefmethode.....	24
3.1.4	Initiële steekproef en opvolgregels	24
3.1.5	Panelduur, interval tussen en aantal golven.....	27
3.1.6	Onderzoekseenheid.....	28
3.1.7	Bevragingsmethode en non-responsstrategie	29
3.1.8	Een korte inleiding tot SONAR	31
3.2	Financiële dimensie	36
3.3	Utiliteit- en valorisatiedimensie.....	41
3.3.1	Databasemanagement	41

3.3.2	Vrijgeven van de dataset	41
3.3.3	Toegankelijkheid en verspreiding	42
4	Conclusie: SWOT- analyse.....	43
4.1	Administratieve data gekoppeld aan cross-sectionele surveys (scenario 5)	43
4.2	Administratieve data gekoppeld aan een panelsurvey (scenario 6)	45
4.3	Administratieve data gekoppeld aan de sonar cohort-studie (scenario 7).....	47
5	Bibliografie.....	49

1. Een haalbaarheidsstudie voor toekomstig loopbaanonderzoek

In welke richting evolueert de Vlaamse arbeidsmarkt? Wat beïnvloedt mensen in hun zoektocht naar een gezonde/evenwichtige werk-levenbalans? Welke factoren kunnen (succesvolle) loopbaantransities verklaren? Wat is de invloed van transitie of gebeurtenissen op de arbeidsmarkt op de verdere loopbaanontwikkeling en het einde van de loopbaan? Op welke manier kan flexibiliteit gekoppeld worden aan meer zekerheid? Hoe houden we oudere werknemers langer op de arbeidsmarkt?

Om deze prangende onderzoeksvragen te beantwoorden en een efficiënt beleid hiervoor op te stellen is er accurate informatie nodig betreffende de dynamiek van de Vlaamse arbeidsmarkt. Het beleid dient daarbij langetermijninzichten te krijgen in veranderingen in loopbaanpatronen, determinanten van loopbaantransities en effecten van beleidsmaatregelen.

In Vlaanderen zijn echter amper gegevens beschikbaar voor langetermijnonderzoek. De Panel Studie van Belgische Huishoudens (PSBH) die het dichtst aansluit bij een dergelijke opzet, biedt onvoldoende houvast en raakt verouderd. PSBH is een breed panel waar veel informatie beschikbaar is over verschillende levensdomeinen die bovendien op huishoudniveau beschikbaar is. Maar PSBH is niet specifiek gericht op arbeidsloopbanen en heeft bijgevolg ook voor arbeidsmarktonderzoek een beperkte relevantie. Bepaalde subgroepen op de arbeidsmarkt zijn bijvoorbeeld te beperkt vertegenwoordigd wat de analyses limiteert (zie bv. Frans en Mortelmans, 2008). Bovendien is de gegevensverzameling voor dit panel beëindigd in 2003. Voor diepgaande analyses van loopbaanpatronen en de gebeurtenissen die het loopbaanverloop bepalen, zijn recente data vereist.

Niettegenstaande het feit dat een loopbaanpanel zich primair zou moeten richten op loopbanen van individuen, onderkennen meer en meer onderzoekers ook het belang van een breder sociaal kader bij het bestuderen van loopbanen en loopbaantransities. Veranderingen in de loopbaan en de kans op transitie worden sterk medebepaald door de andere levensdomeinen van een individu. De loopbaan ontwikkelt zich immers in een sociale context die mogelijkheden kan creëren, maar ook beperkingen kan opleggen voor de uitbouw van een carrière. Onder een sociale context verstaan we zowel individuele kenmerken, het organisatiebeleid en het overheidsbeleid (Barnett, 1999; Van Aerscht, 2004). Om deze factoren te kunnen onderzoeken, is er een dynamisch, longitudinaal methodologisch perspectief nodig.

Het belang van longitudinale data wordt ook steeds meer benadrukt om loopbaantransities en de effecten van gebeurtenissen zoals werkloosheid, loopbaanonderbreking, onderwijsdeelname enzovoort te kunnen bestuderen. Deze studie tracht daarom tot een voorstel te komen voor het opzetten van een loopbaanonderzoek dat gericht is op het in kaart brengen van Vlaamse loopbanen en de interactie tussen arbeid en andere levenssferen. In een levensloopynisch perspectief op de arbeidsmarkt en individuen zijn longitudinale data en een bijhorende methodologische aanpak noodzakelijk. Omdat de keuze voor het type van dataverzameling nog open ligt, worden in deze studie verscheidene mogelijke scenario's besproken.

In deel 1 worden er zeven mogelijke scenario's kort besproken. Met het oog op toekomstig loopbaanonderzoek zullen we de voor- en nadelen van deze verschillende designs evalueren. De keuze voor het type design is daarbij afhankelijk van de onderzoeksvragen, de verwachte effecten en de (soort) gevolgen die men wil bestuderen. In het licht van deze laatste bemerking worden de laatste drie scenario's in deel 2 meer in detail uitgewerkt. Bij het ontwerpen van deze designs wordt onder meer rekening gehouden met volgende parameters: de inhoudelijke focus (de onderzoeksvragen), de methodologische mogelijkheden en implicaties (onderzoekseenheid, design en sampling techniek, vragenlijst en non-respons), de financiële implicaties (indicatie van kostprijs van het design), en de validiteit.

2. Mogelijke scenario's voor een loopbaanpanel

2.1 Scenario 1 en 2: Zuiver loopbaanpanel

Een longitudinaal paneldesign wordt in de literatuur omschreven als herhaalde metingen die op verschillende tijdstippen uitgevoerd worden met hetzelfde meetinstrument om ontwikkelingen en trends in een populatie in kaart te kunnen brengen. Deze definitie kan evenwel verschillende designs omvatten. In zijn restrictieve betekenis houdt een paneldesign in dat een representatieve groep individuen doorheen de tijd gevolgd wordt. In de opeenvolgende interviews of golven worden bij deze groep, op frequente basis, data verzameld. Omdat het telkens dezelfde individuen betreft, wordt dit design in de literatuur een *zuiver* paneldesign genoemd (De Keulenaer, 2007).

Aangezien zowel scenario 1 als scenario 2 vanuit een zuiver paneldesign vertrekken, worden ze in dit deel samen besproken (zie Figuur). De metingen worden doorgaans op verschillende, en vaak vaste tijdstippen genomen bij individuen en/of huishoudens (De Keulenaer, 2007). Het is enkel op gebied van dit tijdsinterval dat de twee scenario's verschillen. Waar het eerste scenario jaarlijks een bevraging afneemt, worden individuen in het tweede scenario pas om de twee jaren bevrraagd zoals in onderstaande figuur weergegeven wordt. De reden om een onderscheid in tijdsinterval te maken heeft met praktische haalbaarheid te maken. In een jaarlijks scenario moeten binnen het Steunpunt WSE zowel de dataverzameling van een volgende golf voorbereid en uitgevoerd worden alsook het analyseren van de voorgaande golf. Dat stelt zeer hoge eisen aan personeel en timing waardoor in het tweede scenario geopteerd wordt om deze taken in een driejaarlijkse cyclus van voorbereiding-uitvoering-analyse onder te brengen.

Figuur 2.1 Scenario 1 en 2. Een zuiver loopbaanpanel

Scenario 1/2 – een zuiver loopbaanpanel			
	Scenario 1	Scenario 2	
2010	Golf 1	Golf 1	Representatieve steekproef (Dis)proportionele stratificatie Panel + aanzuivering
2011	Golf 2		
2012	Golf 3	Golf 2	
2013	Golf 4		
2014	Golf 5	Golf 3	
2015	Golf 6		

2.1.1 Voordelen

De grootste meerwaarde van een longitudinaal panel is ongetwijfeld het *dynamische* karakter. We kunnen hier verschillende voordelen naar voor schuiven. Het biedt eerst en vooral de mogelijkheid om veranderingen uitvoerig te bestuderen. Een reeks momentopnames van steeds dezelfde mensen laat toe om een completer beeld te schetsen van de wijzigingen in de levensloop en in de arbeidsmarktgedragingen. Wanneer verandering het onderwerp van de studie is, zoals het beschrijven en verklaren van veranderingen in loopbanen, is een panel volgens De Vaus (2006) dan ook het meest gepaste design. Een panelsurvey heeft daarbij twee voornaamste doelen: de veranderingen binnen en tussen individuen blootstellen en deze veranderingen verklaren aan de hand van andere karakteristieken (Bijleveld en Van der Kamp, 1998). Dit design laat aldus analyses in de diepte toe naar determinanten van veranderingen in loopbaanontwikkelingen (Van Wissen en Meurs, 1989).

De verandering zelf, of de temporele ordening van gebeurtenissen, staat daarbij ook het dichtst bij een meting van *causaliteit*. In onderzoek naar loopbaanontwikkelingen en –veranderingen spelen longitudinale data dan ook een bijzondere rol. Panelsurveys bieden meer inzichten in de causale relaties tussen variabelen (op verschillende tijdstippen) dan cross-sectionele surveys, omdat ze de orde waarin gebeurtenissen elkaar opvolgen, kunnen onderscheiden. De studie van deze relaties kan gebruikt worden om gedetailleerde patronen van individuele verandering te beschrijven, of om de waarden op een later tijdstip te voorspellen aan de hand van metingen op een vorig tijdstip, of om een groter inzicht te verkrijgen in de onderliggende causale mechanismen. De schattingen van deze relaties kunnen overigens enkel door middel van longitudinale studies bekomen worden (Goldstein, 1979).

Cross-sectionele studies kunnen wel de correlaties tussen variabelen onderzoeken, maar zijn beperkt in hun mogelijkheid om de causale relaties vast te stellen omdat zij slechts één meetmoment hebben (Bijleveld en Van der Kamp, 1998). Cross-sectionele studies kunnen bijvoorbeeld concluderen dat loopbaanonderbrekers een grotere jobtevredenheid hebben of een betere algemene levenskwaliteit. Maar of deze link te wijten is aan het causale effect van het nemen van een loopbaanonderbreking dan wel aan selectieve factoren, namelijk dat mensen met een betere levenskwaliteit of jobtevredenheid loopbaanonderbrekingen nemen, kan niet beslist worden aan de hand van cross-sectionele studies. Enkel panelstudies kunnen de opeenvolgingen van gebeurtenissen opsporen door te kijken naar veranderingen in levenskwaliteit en jobtevredenheid als gevolg van een onderbreking van de carrière.

Cross-sectionele surveys hebben ook moeilijkheden om te zien hoe verschillende oorzaken van bepaalde arbeidsmarktgedragingen in elkaar verstrengeld zijn. Gezien er vaak vele factoren een invloed kunnen hebben op individueel arbeidsmarktgedrag is het niet altijd duidelijk wat de onderliggende causale structuur is. In longitudinale surveys is het mogelijk om zich enkel op de variabelen van interesse te richten en tegelijk rekening te houden met tijdsonafhankelijke geobserveerde en ongeobserveerde variabelen. Longitudinale surveys kunnen aldus een belangrijke bijdrage leveren op de vraag *waarom* bepaalde gedragingen zich afspelen (Van Wissen en Meurs, 1989).

Vanuit beleidsoptiek heeft een panel het voordeel om de effecten van het overheidsbeleid en de sturing door sociale partners te meten en te evalueren op een meer dynamische manier. Het *effect van het arbeidsmarktbeleid* dat meestal vanuit trends en veranderingen vertrekt, kan gemakkelijker gedetecteerd worden aan de hand van herhaalde observaties bij dezelfde individuen (Van Wissen en Meurs, 1989). Eenmalige cross-sectionele metingen slagen er niet in om beleidsmaatregelen te evalueren en gevolgen in kaart te brengen omdat zij geen longitudinale gegevens hebben van individuen om de exacte impact na te gaan. Om een efficiënt en effectief beleidsplan op te stellen is het niettemin belangrijk om de knelpunten en opportuniteiten op te sporen van het beleid en om

steeds op de hoogte te zijn van de behoeften van individuen op de arbeidsmarkt (De Winne, et al., 2003).

Een aantrekkelijk kenmerk van paneldesigns, is het feit dat een groot deel van de *informatie* over elk individu accumuleert over de tijd. Aangezien individueel gedrag zich afspeelt in een sociale en biografische context, kan de kennis over individuen over een bepaalde periode een beter inzicht geven in bepaalde beslissingen of ontwikkelingen dan een eenmalige momentopname zoals in een cross-sectie het geval is. De geschiedenis en vroegere ervaringen kunnen zeer belangrijk zijn om huidige gedragingen te verklaren. Een geschiedenis van vele werkloosheidsperiodes tijdens de loopbaan bijvoorbeeld kan verklaren waarom iemand de arbeidsmarkt sneller verlaat (Hayward en Grady, 1990). Het is aldus mogelijk dat de effecten pas zichtbaar zijn na verloop van tijd. Enkel longitudinale data zouden aldus in staat zijn om vertraagde effecten op te sporen. Een cruciaal aspect van paneldesigns is dan ook dat zij onderzoekers toelaten om te onderzoeken in hoeverre individuen veranderen op lange termijn (Van Wissen en Meurs, 1989; De Vaus, 2006).

Daarbovenop hebben panelsurveys vaak *rijke datasets* van variabelen waarin een schat van informatie over de individuen vervat zit. Dit biedt een ruime context waarin bepaalde gedragingen en veranderingen van individuen of groepen begrepen kunnen worden. Een voorbeeld van zo'n rijke dataset is die van OSA. De OSA organiseert een tweejaarlijks arbeidsaanbodpanel in Nederland en vormt daarmee een belangrijke bron van kennis over de Nederlandse arbeidsmarkt. Sinds de jaren tachtig verzamelen zij informatie over onder meer stromen, loonontwikkeling, werkbeleving, baanmobiliteit, scholing en gebruik van overheidsregelingen. Daarnaast worden er specifieke onderwerpen aangesneden die dieper ingaan op bepaalde zaken zoals kinderopvang (gebruik, kosten, en soort kinderopvang), soort werk (onbetaald, vrijwilligers), zwartwerk (mate van, inkomsten ervan, mening over), meningen en preferenties (over arbeid, vrije tijd, huidige baan) (OSA, 10/11/2008¹). Deze rijke dataset laat toe om een breed spectrum aan onderzoeksonderwerpen te bestuderen. Dit panel kan aldus als voorbeeld dienen voor het Vlaamse loopbaanpanel.

Een ander voordeel van panelsurveys voor dit onderzoek, is het feit dat bepaalde *subjectieve belevingen* bij de loopbaanontwikkeling (zoals motivaties, preferenties, intenties) opgevolgd kunnen worden in de tijd. Dit in tegenstelling tot cross-sectionele surveys waar dergelijke informatie vaak een bias (in geval van retrospectieve vragen, een herinneringsbias) met zich meebrengt. Preferenties zijn dynamisch en zijn gedurende de levensloop wellicht onderhevig aan continue aanpassingen. Hoe beslissingen op de arbeidsmarkt worden beïnvloed door dergelijke individuele belevingen en omgekeerd, bleef totnogtoe echter onderbelicht in loopbaanonderzoek, hoewel zij een belangrijke wederzijdse invloed kunnen uitoefenen (Hakim, 2000).

2.1.2 Nadelen

Panelsurveys hebben ook een aantal belangrijke nadelen. Hoewel de lijst kleiner lijkt, zijn de nadelen niet te onderschatten. Het grootste nadeel bij panelsurveys is de *non-respons bias* die de representativiteit in gedrang kan brengen. Er zijn drie situaties die hiertoe aanleiding kunnen geven: (a) individuen die niet meer opgespoord kunnen worden (bv. verhuis), (b) individuen die weigeren om mee te werken; (c) individuen die uitvallen (bv. overlijden). Het grote probleem hierbij is dat deze non-respons vaak niet *random* is daar in bijna alle surveys niet-responderende individuen verschillen van de meewerkende individuen in bepaalde achtergrondkenmerken. Hierdoor kan er sprake zijn van zelfselectie en over- of ondervertegenwoordiging van bepaalde groepen in de populatie. Dit probleem kan eveneens voorkomen bij cross-sectioneel onderzoek, maar bij panelon-

¹ De vragenlijsten kan men vinden op de website <http://www.uvt.nl/osa/data/>. Voor onderzoeksprojecten en publicatieoverzicht zie: <http://www.uvt.nl/osa/nieuws/jaarverslag2006.pdf>

derzoek kan dit nadeliger zijn, daar de non-respons bias op elke opeenvolgende golf kan voorkomen (Goldstein, 1979).

De non-respons na de eerste golf, waarbij individuen uitvallen tijdens de duur van het panel, duidt men aan met de term *attritie*. Attritie is de belangrijkste risicofactor bij panelsurveys. Er zullen respondenten na verloop van tijd ofwel geheel niet meer meewerken ofwel afwezig zijn voor bepaalde golven wat resulteert in ontbrekende records. De accumulatie van deze non-respons kan op langere termijn een aanzienlijke non-respons produceren. Als de groep individuen die uitvalt random zou zijn, zou het probleem eenvoudigweg opgelost kunnen worden door een voldoende grote steekproef te trekken in de eerste golf. Het grote probleem is echter dat ook bij attritie de uitval meestal selectief is, wat op termijn een serieuze bias kan veroorzaken (Goldstein, 1979). Individuen uit bepaalde categorieën vallen namelijk doorgaans disproportioneel uit het panel. Het zijn bijvoorbeeld vaker laaggeschoolden, alleenstaanden, werklozen die juist interessant kunnen zijn voor wetenschappelijk onderzoek en gericht arbeidsmarktbeleid. Dit kan leiden tot een verkeerde inschatting van de situatie van de arbeidsmarkt. Het is daarom noodzakelijk dat de uitval zoveel mogelijk binnen de perken gehouden wordt. Dit vergt vooral degelijk veldwerk, waarbij het contact met de respondenten behouden blijft doorheen het hele jaar (Magnusson en Bergman, 1990).

De *compositie* van bijna elke doelpopulatie, bestaande uit individuen of huishoudens, zal onderhevig zijn aan veranderingen over de tijd. Doorheen de jaren worden immers nieuwe individuen opgenomen in de bestudeerde populatie (bij geboorte, immigratie of verandering van status) terwijl andere individuen deze verlaten (bij sterfte, emigratie, of verandering van status). Zo kunnen er bij het bestuderen van huishoudens ook 'nieuwe' huishoudens gevormd worden (kinderen verhuizen, echtscheiding) en andere huishoudens verdwijnen uit de populatie (sterfte van alle huishoudleden, huwelijk van twee huishoudens). Panelsurveys dienen rekening te houden met deze veranderingen, om de representativiteit van het panel te bewaren (De Keulenaer, 2007). Hierdoor vereist een panel een redelijk complex design om veranderingen doorheen de tijd op een betrouwbare manier te meten en representatief te blijven doorheen de tijd:

"While panel surveys, in principle, allow for a more refined way of capturing change, the actual way of reliably measuring change over time are complex and continue to bedevil a great deal of analysis of longitudinal data" (De Vaus, 2006).

Er kunnen evenwel inspanningen gedaan worden om de representativiteit in het design te garanderen. Een goede initiële steekproef in de eerste golf is bijvoorbeeld essentieel voor het toekomstige succes en de kwaliteit van een panel. Hier kunnen dan bepaalde subgroepen oververtegenwoordigd (*oversampling*) worden als anticipatie op een te klein aantal respondenten in deze groep en/of op paneluitval in deze groepen. Het panel kan tevens aangevuld worden met een bijtrekking van nieuwe individuen, wat de representativiteit ook op elk meetmoment min of meer in stand kan houden. Een goed plan om de medewerking van het panel te behouden en goede opspoorstechnieken zijn daarbij eveneens van groot belang (De Keulenaer, 2007).

Een ander belangrijk nadeel van panelsurveys is dat zij behoorlijk *tijd- en geldconsumerend* zijn. Dit type design brengt een grotere kost met zich mee dan de andere designs. Dit heeft deels te maken met de relatief grote steekproef van de populatie in de eerste golf. Maar het heeft vooral te maken met het feit dat er voorzien moet worden in de opvolging van steeds dezelfde groep individuen doorheen de jaren en de technieken die ervoor moeten zorgen dat de doelgroep representatief blijft doorheen de tijd. Het ligt voor de hand dat de kostprijs in een dergelijk design hoger ligt dan in een steekproef van N personen op elk tijdstip. Bovendien kunnen de organisatorische kosten om de panelsurvey uit te voeren op lange termijn ook hoog oplopen (Bijleveld en Van der Kamp, 1998). Het is daarbij ook noodzakelijk om een lange-termijn planning op te stellen. Het po-

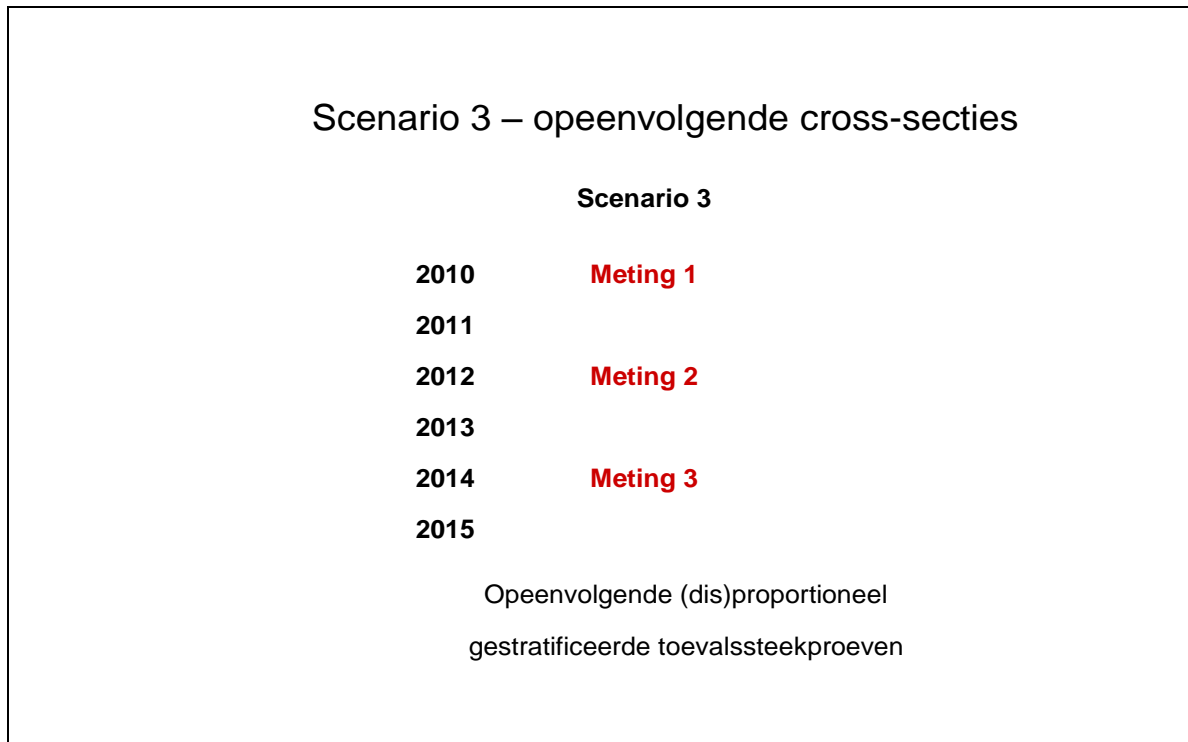
tentiële gevaar zit erin dat er vaak een beperkte financiering of een mogelijke beleidswissel het voortzetten van het project bemoeilijkt (Van Wissen en Meurs, 1989).

De ervaring leert dat de hoeveelheid tijd nodig voor de dataverzameling eveneens niet te onderschatten is. Het duurt namelijk een tijdje vooraleer de gegevens analyseklaar gemaakt zijn en de gepaste analyses daadwerkelijk uitgevoerd kunnen worden. De verzameling van de longitudinale gegevens vereist een enorme organisatie en veldwerk, waardoor resultaten pas na enige tijd openbaar gemaakt kunnen worden. Hierdoor dreigt de verhouding tussen dataverzameling enerzijds en data-interpretatie anderzijds onevenwichtig te worden. Scenario 2 kan hier de tijdsrapte enigszins verbeteren en een betere verhouding toelaten, gezien er meer tijd is tussen de opeenvolgende golven.

2.2 Scenario 3: Opeenvolgende cross-secties

Een cross-sectioneel design verzamelt data op één punt in de tijd. Uit de onderzoekspopulatie wordt een representatieve steekproef getrokken die eenmalig bevroegd wordt. Een dergelijke momentopname maakt dat het design eerder statisch is. Wanneer deze procedure zich in de tijd herhaalt, of wanneer met andere woorden op verschillende tijdstippen telkens verschillende representatieve steekproeven worden bestudeerd, krijgen we een reeks opeenvolgende cross-sectionele surveys. In dit scenario wordt het meetinstrument in de mate van het mogelijke constant gehouden over de tijd, waardoor het mogelijk wordt bepaalde veranderingen en evoluties in kaart te brengen (De Vaus, 2006).

In Figuur wordt een dergelijk herhaald cross-sectioneel design voorgesteld. De metingen of cross-secties worden dan bijvoorbeeld om de twee jaar uitgevoerd. Telkens zal er een nieuwe representatieve steekproef getrokken worden uit de onderzoekspopulatie. Dit is dan ook het grote verschil met voorgaande scenario's. Waar een panel steeds dezelfde individuen bevroegt en onderzoekt, betreft het in dit scenario telkens een verschillende groep individuen. Beide designs bieden evenwel inzichten over de tijd en naargelang de onderzoeksvragen zijn er mogelijkheden en beperkingen verbonden aan elk.

Figuur 2.2 Scenario 3- opeenvolgende cross-secties

2.2.1 Voordelen

Door het verzamelen van cross-sectionele data op twee of meerdere tijdstippen, wil men longitudinale gegevens verzamelen zonder de typische nadelen van panels. Herhaalde surveys proberen een alternatief te bieden voor panels en laten inderdaad ook analyses toe van veranderingen en trends op geaggregeerd niveau. Waar panelstudies het meest geschikt zijn voor de studie van individuele verandering, zijn opeenvolgende cross-secties bijzonder aangewezen bij de studie van *sociale verandering* (De Vaus, 2007a). Omdat herhaalde cross-secties ideaal zijn voor het bestuderen van trends, worden deze vaak trendstudies genoemd.

Specifiek voor een onderzoek naar loopbanen, kunnen herhaalde cross-sectionele surveys bruikbaar zijn om geaggregeerde trends in *arbeidsmobiliteit* te beschrijven en kunnen ze in zekere mate aangewend worden voor beleidstoepassingen (Van Wissen en Meurs, 1989). Deze data kunnen een algemeen zicht geven op de arbeidsmarkt, op basis van beschrijvende statistiek en trendanalyses. Op basis van cross-sectionele gegevens is er bijgevolg een kennisopbouw in de breedte mogelijk. De data kunnen bovendien op elk tijdstip gezien worden als een afzonderlijke cross-sectie. Na elke meting is het mogelijk om analyses uit te voeren, waardoor er relatief *snel* resultaten gerapporteerd kunnen worden. Het is in dit geval dus niet noodzakelijk om data op lange termijn te hebben zoals bij panels (Menard, 1976).

Indien de groepen individuen ook gelijkaardig zijn van de ene periode tot de andere, dan kunnen *vergelijkingen* gemaakt worden tussen deze periodes (Menard, 1976). Om de vergelijkbaarheid van de metingen te garanderen, zou evenwel steeds dezelfde vragenlijst voorgelegd moeten worden (Ruspini, 2002). Men kan bij elke meting ook telkens andere groepen nemen naargelang de interesse van het beleid. Als bijvoorbeeld de vergrijzingsproblematiek hoog op de agenda staat, dan kan ervoor gekozen worden om in de volgende cross-sectionele meting de vragenlijst en doelpopulatie te richten op oudere werknemers. Dit creëert de mogelijkheid om een zekere flexibiliteit in te bouwen in de vragenlijsten over de jaren heen in functie van het beleid.

De voordelen van opeenvolgende cross-secties bestaan erin de nadelen die intrinsiek verbonden zijn aan panels, te verminderen of zelfs te elimineren. Het feit dat niet telkens dezelfde individuen opgevolgd moeten worden, brengt immers een aantal voordelen met zich mee. Dit maakt het design ten eerste een stuk eenvoudiger. Opeenvolgende cross-sectionele designs zijn, volgens Menard (1976), zelfs de meest *eenvoudige* longitudinale designs. De kostprijs kan lager liggen omdat niet telkens dezelfde mensen opgevolgd moeten worden en de organisatie- en coördinatiekosten lager liggen. Daarbij is een langdurige investering in dit design niet noodzakelijk. Herhaalde cross-secties zijn minder onderhevig aan controle-effecten, en kunnen gewoonlijk meer individuen opnemen in hun design voor een gegeven budget (Bauer, 2004).

Er is bovendien geen oversampling nodig, noch is er sprake van bias door attritie wat zoals gezegd vaak een belangrijk probleem is bij panelstudies. De technieken om contacten te behouden doorheen de tijd, de kosten voor de nodige incentives en de nood aan gedetailleerde documentatie van de data zullen ervoor zorgen dat de tijd en kosten vereist voor panels veel groter zullen zijn dan die voor cross-sectionele designs (Bauer, 2004).

2.2.2 Nadelen

De sterktes van longitudinale panels zijn in feite de zwaktes van cross-sectionele studies (Cohen en Manion, 1980). Opeenvolgende cross-sectionele surveys laten niet toe om veranderingen op individueel niveau te onderzoeken. Dit is meteen ook het grootste nadeel van het gebruik van cross-sectionele data. Het feit dat er meerdere cross-sectionele surveys na elkaar worden afgenomen bij telkens verschillende individuen laat dan wel meting van veranderingen toe, maar enkel op geaggregeerd niveau. Men kan geen uitspraken doen over individuele veranderingstrajecten (De Vaus, 2007b). Hiervoor zijn immers longitudinale data van dezelfde individuen nodig.

Verdere nadelen van dit design zijn vooral gerelateerd aan inhoudelijke beperkingen van een onderzoek met cross-sectionele data. Het laat kennisopbouw toe in de breedte, maar niet in de diepte. Er kunnen dan wel snellere resultaten gepresenteerd worden, maar dergelijk onderzoek kan enkel een statisch beeld geven van het bestudeerde proces. Vaak willen onderzoekers daarbij niet alleen de proportie van de populatie in een specifieke staat op verschillende tijdstippen onderzoeken, en trends vaststellen, maar ook de duur in deze staat bestuderen en de transities in en uit deze staten. In geval van loopbanen bijvoorbeeld willen onderzoekers de duur van werkloosheid, de transities in en van de arbeidsmarkt analyseren, en verschillende hypothesen testen omtrent de causale factoren in dit proces (De Keulenaer, 2007). Cross-sectionele studies kunnen echter niets zeggen over de richting, oorzaak of gevolg, van het verband. In dergelijk onderzoek zoekt men vaak naar lineaire verbanden en correlaties tussen variabelen. De causale relaties tussen exogene factoren en arbeidsmarktresponsen op bepaalde effecten kunnen bovendien ook pas na bepaalde tijd zichtbaar zijn. Ook andere dynamische kenmerken kunnen niet geanalyseerd worden aan de hand van cross-sectionele data (Van Wissen en Meurs, 1989). Er kunnen wel bepaalde technieken gebruikt worden om specifieke types van verandering en verklaringen te onderzoeken, maar ook deze zijn niet zonder bias. Het gebruik van retrospectieve vragen bijvoorbeeld laat toe om de causale orde van bepaalde gebeurtenissen te bepalen, maar kunnen vertekeningen vertonen omwille van herinneringsbias (De Vaus, 2006).

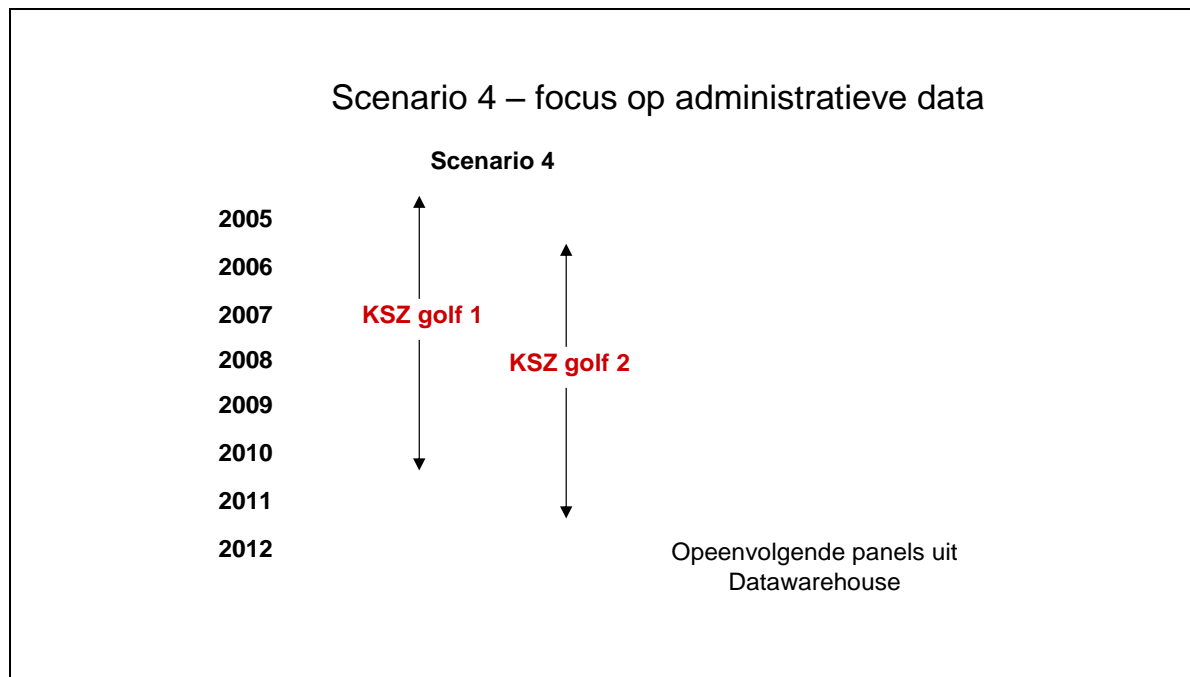
Hoewel scenario 3 een aantrekkelijk design lijkt in termen van kosten en flexibiliteit, is het minder geschikt om levensloopeffecten vast te stellen. Het is moeilijk om leeftijd- en cohorteffecten te onderscheiden. Cohorteffecten kunnen namelijk niet uitgeschakeld worden gezien leeftijd en cohorte volledig gecorreleerd zijn. Selectie-effecten kunnen eveneens niet goed geëlimineerd worden, omdat er geen informatie voorhanden is over kenmerken voorafgaand aan de transities (Kalmijn, 2002). Deze cohort- en leeftijdseffecten kunnen wel onderscheiden worden aan de hand van paneldata.

2.3 Scenario 4: Administratieve data

Het Datawarehouse Arbeidsmarkt en Sociale Bescherming is een administratieve databank waarin sociaal-economische gegevens van verschillende sociale zekerheidsinstellingen gekoppeld worden. Administratieve data bevatten informatie die verzameld werd voor administratieve doeleinden. De verschillende gegevens worden gekoppeld door de Kruispuntbank Sociale Zekerheid (KSZ). Over het individu is er informatie beschikbaar over onder meer persoonskenmerken, gezinssamenstelling, arbeidsprestaties en RVA-statuuat, kinderbijslag, pensioen of invaliditeit (Van der Hallen en Van Mechelen, 2001; Geurts, 2006).

Het Datawarehouse omvat alle individuen die in de loop van een kwartaal gekend zijn bij een van de sociale zekerheidsinstellingen. De gegevens van deze individuen werden achtereenvolgens gekoppeld aan gegevens van de gezinsleden via opzoeking in het Rijksregister. De dataset is ook bruikbaar voor wetenschappelijk onderzoek. Hiervoor dient er een gegevensaanvraag ingediend te worden bij de KSZ volgens een vaste procedure². Scenario 4 tracht gebruik te maken van deze dataset door uit dit register een steekproef te trekken waarvan de individuen opgevolgd kunnen worden doorheen de tijd. Hierdoor heeft de beschikbare informatie een panelkarakter en kunnen individuele veranderingen bestudeerd worden (zie Figuur).

Figuur 2.3 Scenario 4- focus op administratieve data



² <http://www.ksz.fgov.be/nl/index.asp>

2.3.1 Voordelen

Een belangrijke meerwaarde van het Datawarehouse is het gedetailleerde niveau waarop gegevens beschikbaar zijn. Er is *uitgebreide informatie* beschikbaar over werkenden (sector, arbeidsregime, loon), werkzoekenden (duur werkloosheid), en de niet-beroepsactieven (pensioen, invaliditeit). Ook een aantal demografische gegevens (leeftijd, nationaliteit, gezinssamenstelling) zijn gedetailleerd aanwezig. Deze data leveren met andere woorden de mogelijkheid voor loopbaanonderzoek om de dynamische processen te beschrijven. Er is een schat aan informatie aanwezig om arbeidsmarktonderzoek te verrichten. De mobiliteit van individuen kan zowel naar aard als naar omvang in kaart gebracht worden. Daarenboven is er een koppeling van gezinsvariabelen met arbeidsmarktdata waardoor de gezinscontext kan meegenomen worden in verklarende analyses (De Lathouwer, et al., 2006; Geurts, 2006).

Administratieve data worden verzameld voor een belangrijk deel van de Belgische populatie. Hierdoor is het risico op 'sampling errors' veel kleiner. Ook voor kleine subgroepen in de populatie valt het klassieke risico op te weinig observaties hierdoor weg (Ruspini, 2002). Deze data maken het aldus mogelijk om specifieke gebeurtenissen of transities te bestuderen die in gewone surveys bij onvoldoende respondenten voorkomen. In de PSBH zijn er bijvoorbeeld over de 11 jaren heen slechts 141 voltijdse loopbaanonderbrekers aanwezig (zie Frans en Mortelmans, 2008). Dit kleine aantal belemmert een uitvoerig onderzoek naar de gevolgen van een onderbreking op de verdere loopbaanontwikkeling waarbij ook rekening wordt gehouden met exogene factoren.

Het voordeel is dat de data in principe representatief zijn voor heel Vlaanderen en dat de gegevens gekoppeld kunnen worden over de tijd. Hierdoor heeft dit design een *panelkarakter* en kunnen de ontwikkelingen van individuen over de tijd gevolgd worden. De paneldata in dit design vertonen geen bias ten gevolge van attritie. De eventuele uitval die nog optreedt, is beperkt en wellicht minder selectief. In administratieve data is er sprake van een meer natuurlijke uitval (vanwege bijvoorbeeld sterfte) in plaats van non-respons of attritie. Bovendien kunnen er grote steekproeven getrokken worden, wat uitval minder schadelijk maakt.

Een panel met administratieve data heeft als bijkomend voordeel dat lange *tijdreeksen* van dezelfde individuen kunnen worden bestudeerd. Dit laat toe om bepaalde effecten in te schatten op lange termijn, waardoor de impact (zelfs na vertraging) gemeten kan worden. Hier komt bovenop dat de gegevens op korte termijn telkens opnieuw geregistreerd worden (i.e. *kwartaalgegevens*), waardoor zeer korte tijdsintervallen bestudeerd kunnen worden. Dit kan een belangrijk voordeel zijn voor loopbaanonderzoek. Transities en statuten op de arbeidsmarkt kunnen namelijk op zeer korte tijd veranderen. Op een jaar tijd kunnen individuen bijvoorbeeld meerdere transities maken.

Administratieve panels zijn ongetwijfeld de minst indringende basis om longitudinale data te verzamelen. Een groot voordeel bij het gebruiken van administratieve data is dat de data niet zelf verzameld moeten worden, waardoor in feite enkel de analyses dienen te gebeuren. De data zijn vaak elektronisch beschikbaar, wat het transport vergemakkelijkt. Ook de analyse zelf is relatief goedkoop gezien de meeste databases in gestandaardiseerde formats te verkrijgen zijn en bijna alle analyses op de computer uitgevoerd kunnen worden met SAS, SPSS e.d. (Billings, 2003). Voor dit scenario zal de kostprijs aldus minimaal zijn (Ruspini, 2002).

Administratieve data worden ook verondersteld een grote *accuratheid* te hebben. De informatie wordt eerst en vooral nagenoeg op het moment zelf geïnventariseerd. Gezien de inventarisatie niet op bevraging van individuen is gebaseerd, stelt zich evenmin het risico van een herinneringsbias. Ten tweede is het via deze weg mogelijk om over gevoelige informatie te beschikken over zaken die in survey-onderzoek vaak (item)non-respons opleveren. Gegevens over bijvoorbeeld het inko-

men kunnen hier met een redelijk grote accuraatheid verkregen worden (Moffit en Ver Ploeg, 2001).

2.3.2 Nadelen

Administratieve data komen evenwel niet zonder problemen of beperkingen. Ten eerste betreft het *ruwe data*. De gegevens zijn in eerste instantie verzameld voor bepaalde administratieve doelstellingen en niet vanuit het perspectief van een bepaalde wetenschappelijke probleemstelling. Dit kan een aantal praktische problemen met zich meebrengen. De inhoud kan namelijk afwijken van wat vanuit onderzoeksperspectief gewenst is. Omdat het interessegebied niet centraal staat in de data, kan er informatie ontbreken en afhankelijk van administratieve procedures, kunnen gegevens van bepaalde groepen systematisch ontbreken (Borghans en Kriechel, 2008). Daarbij komt dat de definitie van de variabelen waarschijnlijk zal verschillen wanneer zij verzameld worden voor administratieve doeleinden en niet met het oog op specifieke onderzoeksdoelen (m.a.w. hoe de onderzoeker ze zou definiëren) (De Keulenaer, 2007).

De gebruikswaarde van administratieve data bereikt aldus limieten op inhoudelijk vlak en ook op vlak van analyse. Vele interessante onderzoeksvragen kunnen niet onderzocht worden omdat de informatie simpelweg ontbreekt. De beschikbare administratieve bestanden bevatten minder en andere variabelen dan databestanden die gebaseerd zijn op wetenschappelijke enquêtes (Borghans en Kriechel, 2008). Subjectieve gegevens zoals percepties, intenties en motivaties kunnen bijvoorbeeld niet onderzocht worden. Gegevens betreffende de arbeidsinhoud en kwaliteit van de arbeid mankeren eveneens. Analyses zullen zich bijgevolg moeten beperken tot de variabelen in de originele survey van administratieve gegevens. Bovendien is er veelal een gebrek aan de nodige controlevariabelen. Zo bijvoorbeeld ontbreekt het onderwijsniveau, een belangrijke controlevariabele, in het datawarehouse AM&SB.

Administratieve data geven aldus zelden een compleet beeld van het studieobject, waardoor verdere analyses vaak noodzakelijk zijn. Er kunnen patronen ontdekt worden en hypothesen geformuleerd worden, maar in vele gevallen is bijkomende informatie nodig om de volgende stap te bepalen. De data kunnen daarbij ofwel onvolledig, niet aanwezig of ook fout zijn. Voorzichtigheid bij het gebruik van dergelijke data is daarom essentieel (Billings, 2003).

De grote omvang van de data kan eveneens technische complicaties met zich meebrengen. Het werkgeheugen van moderne computers kan overschreden worden zodat een efficiënte selectie en voorbereiding van de data belangrijk is (Borghans en Kriechel, 2008). De toegang tot dergelijke administratieve data kan ernstig bemoeilijkt worden ten gevolge van privacy- en vertrouwensredenen (Moffit en Ver Ploeg, 2001). Het gebruik van de data wordt vaak gehinderd door wetten betreffende databescherming, die het moeilijk kunnen maken om toegang te verkrijgen tot dergelijke data (Ruspini, 2002).

Een bijkomend nadeel is de geringe actualiteit van vele gegevens uit administratieve databanken. Registers kunnen een behoorlijke vertraging oplopen waardoor de informatie niet voldoende actueel is. Ook de toestemming van de privacycommissie omvat vaak een zekere tijdspanne waarin recente gegevens niet vrijgegeven kunnen worden. Dit ondermijnt de relevantie voor beleidsactoren (Bakker, et al., 2005). Beleidsmakers moeten namelijk steeds op de hoogte zijn van de huidige stand van zaken om bepaalde maatregelen te kunnen treffen.

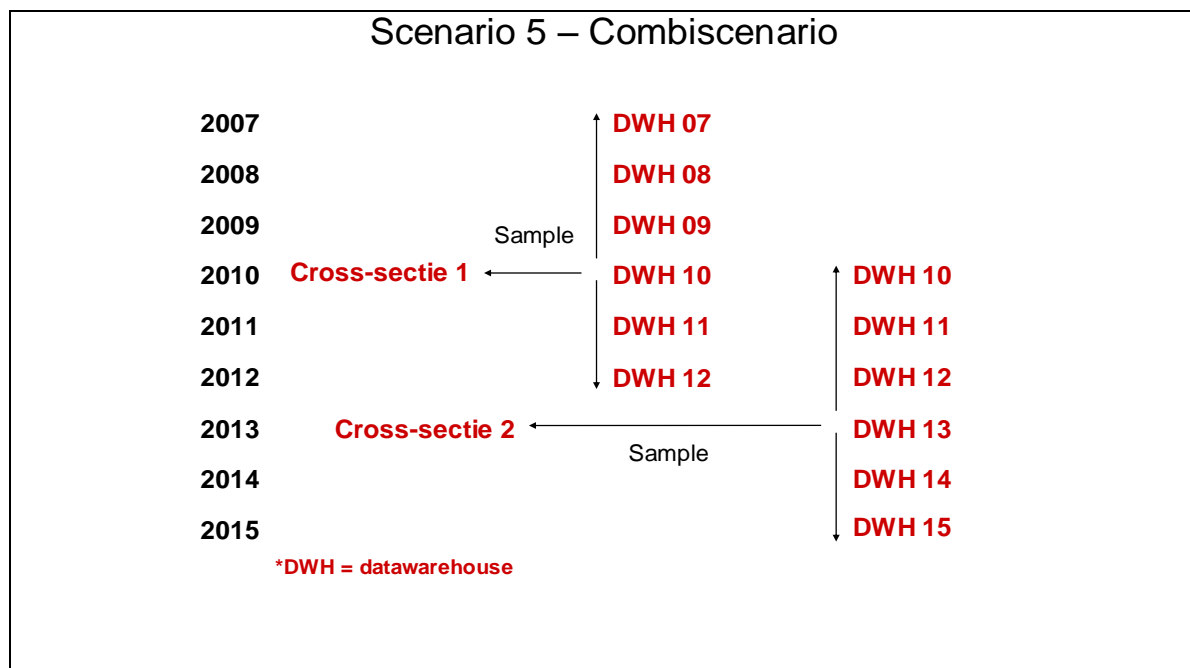
2.4 Scenario 5: Administratieve data gekoppeld aan cross-sectie

Met het oog op de dynamische aard van de onderzoeksvragen is idealiter een longitudinaal panel-design nodig. Zo kunnen veranderingen van individuen in waarden en gedragingen tijdens hun loopbaan continu worden opgevolgd. Zuivere panels (zie 0 scenario 1/2) zijn echter kostelijk in termen van geld en tijd. Een goedkopere en snellere manier zou een panel zijn via administratieve data (zie 0 scenario 4). Hoewel het longitudinale karakter behouden blijft en er een grote hoeveelheid informatie vervat zit in administratieve data, stoot dit design echter nog steeds op duidelijke beperkingen (vanwege het administratieve karakter) (Borghans en Kriechel, 2008).

De nadelen zouden echter beperkt kunnen worden door een combinatie van verschillende scenario's. Een minder duur en indringend design is dan bijvoorbeeld het linken van surveys aan een grote hoeveelheid individuele informatie aanwezig in administratieve databanken. Rekening houdend met de eigenheid en specificiteit van elke bron, kan het denken in termen van en/en in plaats van of/of een meerwaarde bieden voor verder loopbaanonderzoek. Door complementair gebruik kunnen namelijk de sterktes van beide bronnen worden gecombineerd en de zwaktes worden geëlimineerd (Van der Hallen en Vanheerswyngheles, 2004).

We bespreken twee dergelijke scenario's in deze haalbaarheidsstudie. Een eerste scenario omvat het koppelen van administratieve data aan cross-sectionele surveys, m.a.w. een combinatie van scenario 3 en scenario 4 (Figuur). Hier wordt een bepaalde groep individuen opgevolgd in de tijd in de administratieve data, en deze groep zal ook eenmalig bevraagd worden voor bijkomende informatie belangrijk voor het onderzoek. Een tweede scenario (zie verder) kijkt eveneens een groep individuen doorheen de tijd in administratieve data, maar zal deze regelmatig opnieuw bevragen. Dit is dus een koppeling van scenario 1/2 en scenario 4 (zie Figuur). Omdat beide scenario's fundamenteel verschillen op een aantal vlakken, bespreken we ze apart. In dit deel (scenario 5) worden de voor- en nadelen van de koppeling van administratieve data met cross-sectionele gegevens tegen elkaar afgewogen.

Figuur 2.4 Scenario 5. Koppeling administratieve data met cross-sectie(s)



2.4.1 Voordelen

Het gebruik van administratieve data stoot op een aantal *beperkingen* (zie 0 2.3.2 Nadelen). Het belangrijkste nadeel is het gebrek aan belangrijke informatie in administratieve data. Niettemin zijn er belangrijke voordelen verbonden aan het gebruik van deze datasets (waaronder een grote accuraatheid). Om deze voordelen niet verloren te laten gaan, kunnen administratieve gegevens gekoppeld worden aan surveygegevens. De rijke dataset wordt aangevuld met cross-secties om zo een deel van deze beperkingen te voorkomen.

Het grootste voordeel van een koppeling van administratieve data aan cross-sectionele bevestigingen is dat de cruciale variabelen die ontbreken in de administratieve dataset kunnen worden opgenomen in de survey. Het Datawarehouse bevat bijvoorbeeld geen onderwijs- of opleidingsgegevens van individuen (Geurts, 2006). Het mag duidelijk zijn dat het onderwijsniveau van groot belang is op de arbeidsmarkt, voor de loopbaanontwikkeling van individuen en dus ook voor loopbaanonderzoek. Deze informatie kan dan verkregen worden aan de hand van cross-secties. Zoals De Lathouwer (2006) vaststelt:

“Een aantal belangrijke gegevens zullen hoe dan ook niet, of zeer moeilijk, via administratieve weg verzameld kunnen worden. Het verzamelen van gegevens via enquêtes blijft daarom onontbeerlijk. Idealiter worden administratieve gegevens en enquêtegegevens aan elkaar gekoppeld.”

Een cross-sectionele bevestiging kan bovendien peilen naar andere bijkomende informatie. Zo kunnen additionele factoren onderzocht worden zoals gezondheid, preferenties, motivaties, attitudes, tevredenheid, kwaliteit van het werk, werkomstandigheden, enz. Deze factoren worden in onderzoek steeds meer naar voren geschoven als belangrijke verklarende variabelen. Zeker in het licht van de vergrijzing worden er verklaringen gezocht in deze richting om mensen langer op de arbeidsmarkt te houden (Bakker, et al., 2005). Afhankelijk van het onderwerp van bevestiging kunnen tevens retrospectieve data worden verzameld met een relatief hoge graad van betrouwbaarheid om bepaalde gedragingen die geobserveerd werden in administratieve data meer te duiden of te verklaren.

Administratieve registers kunnen de kosten helpen verlagen en de steekproeftrekking optimaliseren. Op basis van registers kunnen de steekproeven immers efficiënter getrokken worden omdat deze gestratificeerd kunnen worden naar bepaalde kenmerken aanwezig in de databanken (Bakker, et al., 2005). Hierdoor krijgen we een redelijk accuraat steekproefkader. De vragenlijst kan daarbij ook aanzienlijk verkort worden omdat er al veel informatie beschikbaar is. Hierdoor kan men zich ook veel meer richten op bepaalde aspecten ter verklaring van loopbaanbeslissingen. Ook hier zullen de gegevens vrijwel meteen beschikbaar zijn. Dit laat toe dat analyses en resultaten relatief snel gepresenteerd kunnen worden.

2.4.2 Nadelen

Een koppeling van cross-sectionele data aan administratieve data stoot nog steeds op enkele minpunten. Gezien er een eenmalige momentopname gekoppeld wordt aan administratieve data zijn er geen effectmetingen mogelijk van acties op gedrag of attitudes. Door het cross-sectionele karakter van de data is onderzoek naar causaliteitsverbanden nog steeds aan beperkingen onderhevig. Er zijn wel effectmetingen mogelijk, maar deze blijven beperkt tot effecten van variabelen uit de administratieve databank (bv. het wel of niet maken van een transitie in jaar t) op variabelen uit de bijkomende survey (bv. een attitude in jaar t+1) en van variabelen uit de bijkomende survey (bv. de loopbaantevredenheid of scholingsinspanning in jaar t+1 op het wel of niet maken van een transitie

in jaar t+2, tenminste wanneer de surveygegevens niet alleen aan retrospectieve administratieve data gekoppeld worden, maar ook verrijkt worden met administratieve data uit de jaren na de bevraging). Causaliteitsuitspraken binnen de context van de bijkomende survey zijn echter onmogelijk, omdat elke survey op nieuwe steekproeven is gebaseerd.

Verder is veel afhankelijk van de aard van de in administratieve databanken beschikbare data. Variabelen die niet beschikbaar zijn, moeten in de survey bevroegd worden. Indien die survey cross-sectioneel is, dan kan daarbij enkel een tijdsdimensie ingebracht worden door retrospectieve vraagstelling en bijgevolg 'historische' observatie. Deze zijn echter nooit zonder bias. Retrospectieve vragen zijn gelimiteerd doordat enkel herinnerbare gebeurtenissen redelijk accuraat opgehaald kunnen worden (o.a. huwelijk, intrede arbeidsmarkt, geboorte kind).

Daarnaast kunnen er ook problemen verwacht worden bij de koppeling. Om een beroep te doen op de diensten van de Kruispuntbank Sociale Zekerheid, moet een aanvraag ingediend worden die een afdoende beschrijving geeft van het geplande wetenschappelijk onderzoek en van de gegevens die vereist zijn om het onderzoek tot een goed einde te brengen. De aanvraag voor niet-gecodeerde sociale gegevens van persoonlijke aard³, dient ingediend te worden bij de Kruispuntbank, waarvan de Beheerscomité beslist of al dan niet op het verzoek wordt ingegaan. Hiervoor dient het Sectoraal comité van de sociale zekerheid zijn machtiging te verlenen. Indien het verzoek wordt goedgekeurd en een machtiging wordt verleend, worden de niet-gecodeerde sociale gegevens van persoonlijke aard van de instellingen van sociale zekerheid verzameld, gekoppeld en meegedeeld door de Kruispuntbank⁴.

De Kruispuntbank en de instellingen van sociale zekerheid verschaffen in principe enkel anonieme gegevens of gecodeerde sociale gegevens van persoonlijke aard aan onderzoeksinstellingen. Ze besteden hierbij bijzondere aandacht aan het minimaliseren van de mogelijkheid om de personen op wie de gegevens betrekking hebben te heridentificeren. De privacy bij een dergelijke koppeling is daarbij cruciaal. Het kan daarbij een moeilijk dossier worden bij KSZ omwille van de combinatie van beide methoden. Niet-gecodeerde sociale gegevens van persoonlijke aard worden aldus slechts in uitzonderlijke gevallen meegedeeld. Dit is enkel het geval als het strikt noodzakelijk is voor de realisatie van de wetenschappelijke doeleinden. Er is daarbovenop ook toestemming nodig van de respondent zelf. Het effect op de non-respons is echter moeilijk in te schatten op voorhand. Er kan weliswaar verwacht worden dat de non-respons bij het linken van administratieve data en survey gegevens groter is. Bij cross-sectionele bevraging zal er steeds een non-respons zijn. Er zullen mogelijk meer individuen afhaken als zij toestemming dienen te geven voor een koppeling van hun antwoorden aan hun persoonlijke gegevens via administratieve data.

³ Niet-gecodeerde sociale gegevens kunnen wél door de onderzoeksinstellingen worden teruggebracht tot individuele personen (bijvoorbeeld voor steekproeven voor enquêtes)

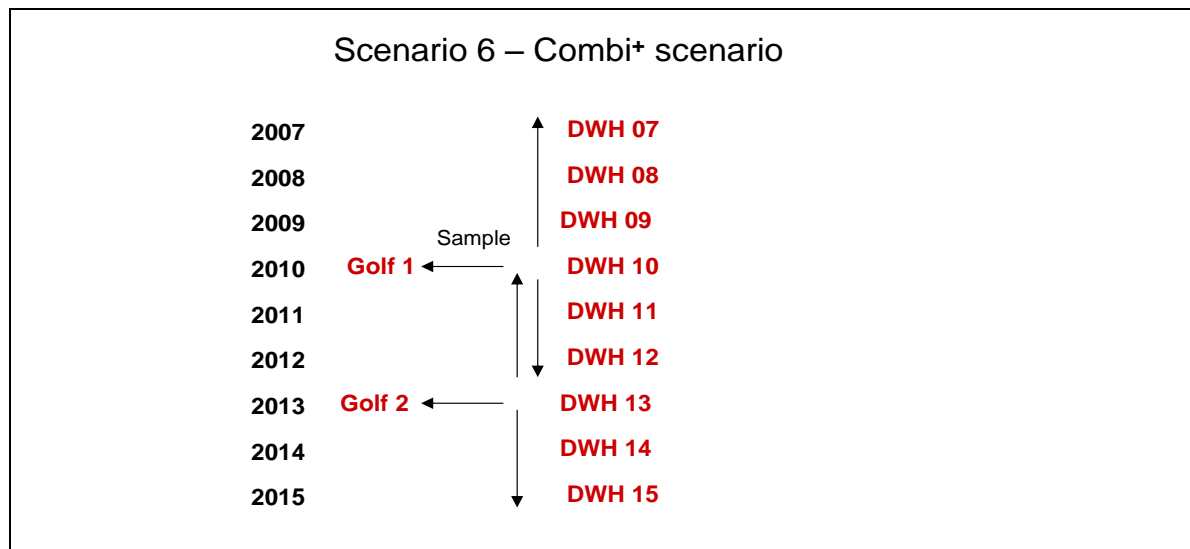
⁴ <http://www.ksz.fgov.be/>

2.5 Scenario 6: Administratieve data gekoppeld aan panelstudie

Scenario 6 gaat, zoals reeds aangegeven, een stapje verder dan scenario 5 en koppelt een panel-survey aan een administratief panel. Vanuit beleids perspectief is het namelijk belangrijk om beleidsevaluaties te kunnen maken. In dit scenario wordt uit de administratieve databank een steekproef getrokken worden van individuen die opgevolgd zullen worden in de tijd. Daarnaast zal dezelfde groep individuen ook nog eens herhaaldelijk bevestigd worden zoals in een panel het geval is (zie Figuur).

Waar voorgaand scenario een cross-sectionele survey koppelt aan administratieve gegevens, zal in dit scenario de survey bij dezelfde respondenten herhaald worden in de tijd. Dat betekent dat de eerste golf van dit scenario identiek is aan één bevraging in scenario 5. Op deze manier verkrijgen we aldus twee naast elkaar bestaande panels van telkens dezelfde individuen. Dit scenario wordt weergegeven in Figuur .

Figuur 2.5 Scenario 6. Koppeling administratieve data met panelsurvey



2.5.1 Voordelen

Gezien het om een panel gaat, zijn de voordelen van scenario 1 en 2 ook van toepassing op dit scenario. Het verschil schuilt vooral in de hoeveelheid informatie die beschikbaar wordt door de koppeling aan administratieve data. Hierdoor is er minder nood aan erg lange vragenlijsten. Bovendien zijn de gegevens vanuit administratieve data over variabelen die gewoonlijk een hoge non-respons opleveren in surveys, zoals het inkomen, redelijk accuraat. Een ander voordeel is dat de opvolging van de individuen efficiënter kan gebeuren. In administratieve data is namelijk informatie beschikbaar over de woonplaats, waardoor individuen bijvoorbeeld na een verhuis teruggevonden kunnen worden. Dit kan de hoge kost van panels in termen van tijd en geld om mensen te volgen doorheen de tijd, verlagen en de uitval verminderen.

Er zijn ook een aantal inhoudelijke motivaties voor dit scenario. De grootste meerwaarde van een koppeling van administratieve data met een longitudinaal panel is dat evoluties in de tijd kunnen worden gemeten voor elementen zoals genderrolopvattingen, motivaties, intenties, percepties, gezondheid, tevredenheid, werkomstandigheden. Deze veranderingen kunnen zoals gezegd niet vastgesteld worden aan de hand van cross-sectionele data.

Dit scenario geeft ongetwijfeld de ruimste mogelijkheden voor verklarende analyses, en dit zowel op meer objectieve loopbaanmerken als op meer subjectieve aspecten van werk- en loopbaanbeleving. Vooral het feit dat herhaalde metingen van zowel 'harde' loopbaanmerken als meer 'zachte' aspecten van loopbaanbeleving, onderscheidt dit scenario van het vorige. Enkel in dit scenario wordt het mogelijk om op een robuuste wijze te onderzoeken welke effecten bepaalde loopbaantransities (bv. onderbreking van de loopbaan, korte periode van werkloosheid, verandering van werkgever, transitie van werknemer naar zelfstandige, etc.) hebben op loopbaantevredenheid, psychisch welzijn, job(on)zekerheid, employability, e.d.m. De rijkdom aan informatie voor beleidsvoorbereidend onderzoek is hier dan ook maximaal.

2.5.2 Nadelen

Ook hier kunnen we teruggrijpen naar de nadelen eigen aan panelstudies die vermeld werden bij scenario's 1 en 2. De non-respons kan aanzienlijke problemen veroorzaken in het geval van panels. Noodzakelijk in dergelijke scenario's is daarbij de toestemming van de respondenten om surveygegevens te koppelen aan administratieve gegevens. Er kunnen aldus problemen verwacht worden op dit vlak, zoals een hogere non-respons, een weigering tot medewerking. Deze weigering zal wellicht groter zijn dan bij de andere designs vanwege de koppeling aan administratieve data. Dit kan leiden tot een bias in de data.

Daarenboven is net zoals in voorgaand scenario de toestemming van de Privacycommissie vereist om survey gegevens te koppelen aan administratieve gegevens. Typisch aan panels en ook een nadeel in dit scenario is de attritie, de uitval door de jaren heen. Een oversampling in de eerste golf is dan ook vaak vereist om representatief te blijven doorheen de tijd. Niettemin kan de paneluitval kleiner zijn dan bij een zuiver panel net door de koppeling met administratieve data. Individuen kunnen hier namelijk gemakkelijker opgevolgd worden. Bovendien blijft ook voor individuen die uitvallen door attritie informatie beschikbaar op basis van de administratieve data.

Dit scenario houdt eveneens een langetermijninvestering in, wat zeker een probleem kan zijn in een opzet afhankelijk van beleidsinstanties die regelmatig wisselen. Een onzekere financiering zal de toekomst van een dergelijke opzet in gedrang brengen waardoor een grote hoeveelheid aan informatie verloren zal gaan.

2.6 Scenario 7. Loopbaan-cohortenstudie (SONAR-scenario)

In scenario 7 wordt een longitudinale cohortstudie voorgesteld. Een cohortstudie is een type panel dat zich richt op één welomschreven subpopulatie, i.e. één of enkele cohorten, die dezelfde gebeurtenis(sen) tijdens dezelfde periode heeft meegemaakt (bv. op hetzelfde moment geboren zijn, getrouwd zijn tijdens dezelfde periode of afgestudeerd zijn in een bepaald jaar). Zulke cohortpanels worden in de literatuur ook wel fixed panel surveys genoemd. De leden van een panel kunnen immers niet veranderen over de tijd en dus zijn de regels om de steekproefeenheden te volgen eenvoudig. In elke nieuwe golf wordt gepoogd data van alle originele leden te verzamelen. Een cohortpanel wordt vaak opgezet om langetermijn veranderingsprocessen, zoals transities naar volwassenheid, te bestuderen (De Keulenaer, 2007).

Omdat dit panel niet helemaal uit het niets te moeten opstarten, zo kunnen aangesloten worden op bestaande cohortpanels. Er is in Vlaanderen één cohortpanel lopende onder de naam SONAR, Studie van de overgang van Onderwijs naar ARBeidsmarkt. SONAR tracht de overgang van school naar werk te bestuderen door verschillende jongerencohorten van 1976, 1978 en 1980 op te volgen. Deze jongeren werden allen geïnterviewd op 23-jarige leeftijd. De eerste twee cohorten werden ook benaderd op 26-jarige leeftijd en de eerste cohorte werd uiteindelijk ook op 29-jarige leef-

tijd bevroegd. Deze databank bevat een aantal mogelijkheden voor beleidsgericht onderzoek en kan ter beschikking gesteld worden van andere onderzoekers. Er wordt hierbij ook onderzocht in hoeverre de gegevens aangevuld kunnen worden via administratieve data⁵.

Scenario 7 omvat een voortzetting van het SONAR-panel in functie van het WSE-loopbaanonderzoek. Hiervoor zou de laatste cohorte van 1980 in aanmerking komen om verder te bevragen. Op het moment van de dataverzameling zullen deze respondenten de leeftijd hebben van 29.

2.6.1 Voordelen

Het continueren van een al bestaande cohortstudie kan interessant zijn omdat reeds een hoeveelheid informatie beschikbaar is over deze individuen. Binnen het beperkte budget is het mogelijk om deze cohorte nog eenmaal face-to-face te bevragen op 29-jarige leeftijd. Gegeven de schaarse middelen in Vlaanderen is het efficiënt om de cohorte, oorspronkelijk bevroegd voor een specifieke school-werk transitie, verder te zetten en op te nemen in een breder kader rond loopbaanonderzoek. Uit financiële overwegingen kan men zich daarbij initieel richten op één cohorte. Naarmate de tijd verstrijkt, kunnen er vervolgens bijkomende cohorten gevolgd worden.

Dergelijke longitudinale panelopzet heeft ook de gekende voordelen van panelstudies. Causale verbanden kunnen nagegaan worden en effectmetingen zijn mogelijk, waardoor de impact van beleidsmaatregelen opgespoord kunnen worden. Gezien op diverse tijdstippen informatie verzameld wordt bij dezelfde individuen, kunnen ook dynamische loopbaanevoluties correct in hun sociale en individuele context geplaatst worden. Dit type panel volgt de veranderingen op van een groep respondenten naarmate ze verouderen. Of individuen meer of minder traditionele genderrolopvattingen hebben dan vroeger kan bijvoorbeeld niet nagegaan worden door opeenvolgende cross-sectionele surveys omdat deze niet kunnen onderzoeken of de verschillen te wijten zijn aan cohort- of aan levensloopverschillen.

2.6.2 Nadelen

De cohorten vanuit SONAR dienen echter aangevuld te worden, want voor loopbaanonderzoek zijn gegevens op huishoudniveau belangrijk. De gezinssituatie, de partner en de kinderen kunnen immers een belangrijke invloed uitoefenen op de beslissingen van individuen op de arbeidsmarkt en in hun loopbaanontwikkeling. Bovendien moeten deze gegevens gekoppeld kunnen worden aan administratieve gegevens. Zo kunnen er lange tijdreeksen onderzocht worden en kan de loopbaanontwikkeling tamelijk accuraat in beeld gebracht worden.

Een nadeel van een cohortpanelstudie is de mogelijkheid dat de resultaten enkel betrekking hebben op een bepaalde cohorte waaruit de sample geselecteerd werd (Duncan & Kalton, 1987). Aangezien leeftijd en periode samen veranderen, is het onmogelijk om hun aparte effecten te onderscheiden. Een uitbreiding van de cohortebenadering is het gebruik van verscheidene cohorten, die gevolgd worden in gelijkaardige panelsurveys, waardoor het mogelijk wordt om leeftijd en periode-effecten apart te bestuderen (De Keulenaer, 2007).

Een beperking is aldus dat er momenteel slechts één cohorte kan worden bestudeerd. Dit heeft als bijkomend nadeel dat het onderzoek zich beperkt tot de beginfase van de carrière. In het licht van de groeiende beleidsaandacht, is het echter tevens noodzakelijk informatie te verzamelen over ou-

⁵ <http://www.steunpuntloopbanen.be/vorm1.html>

dere werknemers, werkzoekenden en beroepsinactieven. Hiervoor kan er een oudere cohorte bijgetrokken worden. Het opstarten van deze cohorte zal echter weer veel tijd en geld kosten.

Een laatste nadeel van het gebruik van reeds bestaande cohortstudies, is het feit dat er al een paneluitval heeft plaatsgevonden. Deze individuen opnieuw interviewen zal een bijkomende attritie (non-respons) veroorzaken, waardoor de bias accumuleert. We behouden daardoor een heel selecte groep in de cohorte, wat de representativiteit mogelijk in het gedrang brengt. Om deze reden zal een bijkomende trekking noodzakelijk zijn wat de kost verhoogt.

2.7 Samenvatting: overzicht van de mogelijke scenario's

Gezien de keuze voor het type dataverzameling nog open ligt, werden in deel één van deze haalbaarheidsstudie zeven mogelijke scenario's voor een toekomstige WSE-loopbaanonderzoek beknopt besproken. Hieruit blijkt duidelijk dat elk scenario zijn specificiteit heeft. Er zullen dan ook afwegingen gemaakt moeten worden in het licht van de onderzoeksvragen en het doel van het onderzoek binnen het beperkte budget. Om een zicht te krijgen op deze verschillende mogelijkheden, werden de belangrijkste voordelen en nadelen van elk scenario kort besproken. Op basis hiervan moet een keuze gemaakt te worden voor het meest aantrekkelijke scenario voor wetenschappelijk onderzoek rond loopbanen en arbeidsmarktbeleid.

Tabel vat voorgaande bespreking samen. Hier worden de pluspunten en minpunten weergegeven voor elk type design voor het loopbaanonderzoek. De plussen en minnen in deze tabel dienen echter niet te strikt geïnterpreteerd te worden. Deze tabel geeft namelijk een overzicht in de mogelijke meerwaarde van elk scenario ten opzichte van de andere scenario's. Dit wil bijvoorbeeld niet zeggen dat de inhoudelijke scope niet groot kan zijn in cross-secties. Over het algemeen hebben panels echter een meerwaarde op inhoudelijk vlak. Ook non-respons kan in elk scenario een groot representativiteitsprobleem veroorzaken, maar kan in een cross-sectie doorgaans beter aangepakt worden.

In dit vergelijkend opzicht dient eveneens niet louter een optelsom gemaakt te worden van plussen of minnen. Bepaalde punten wegen namelijk zwaarder dan andere. De non-respons kan bijvoorbeeld minimaal gehouden worden door bepaalde technieken mee op te nemen in het design. In hoeverre er een budget is voor de scenario's is daarentegen een cruciale overweging die men in rekening moet nemen bij de keuze voor een bepaald scenario. De inhoudelijke doelstellingen en de onderzoeksvragen zijn belangrijke en vaak doorslaggevende factoren voor de keuze voor het type design. Wat willen we onderzoeken? Wat is de meerwaarde voor toekomstig loopbaanonderzoek?

Op basis daarvan, kan er besloten worden dat de laatste drie scenario's inhoudelijk en beleidsmatig de interessantste perspectieven bieden voor toekomstig onderzoek. Deze pistes zullen daarom dieper uitgewerkt worden in deel 2 van deze haalbaarheidsstudie. Hier zal dieper ingegaan worden op de implicaties van de drie verschillende designs op vlak van de inhoudelijke, methodologische, financiële en valorisatie dimensie.

Tabel 2.1 Voor- en nadelen van de mogelijke scenario's besproken in deel 1

	Scenario 1/2 Zuiver panel	Scenario 3 Opeenvolgende cross-secties	Scenario 4 Administratieve data	Scenario 5 Administratieve + cross-sectie	Scenario 6 Administratieve + panel	Scenario 7 Administratieve + cohort (SONAR)
<i>inhoudelijke scope</i>	+	+/-	-	+	+	+
<i>meten van individuele veranderingen</i>	+	-	+	+	+	+
<i>gedetailleerde dataset</i>	+	+/-	+/-	+	+	+
<i>kennisopbouw in de diepte</i>	+	-	-	-	+	+
<i>vernieuwende inzichten</i>	+	+/-	-	+/-	+	+
<i>up-to-date gegevens</i>	-	+	-	+/-	-	+/-
<i>vaststellen van causaliteit</i>	+	-	+/-	+/-	+	+
<i>beleidsevaluatie</i>	+	-	+/-	+/-	+	+
<i>flexibiliteit in vragenlijst</i>	-	+	-	+	-	-
<i>gegevens op huishoudniveau</i>	+	+	-	+	+	+
<i>eenvoudig design</i>	-	+	+	+/-	-	-
<i>steekproefomvang</i>	-	+	+	+	-	-
<i>representativiteit van de steekproef</i>	-	+	+	+	-	-
<i>non-respons bias</i>	-	+	+	-	-	-
<i>attritie</i>	-	nvt	nvt	nvt	-	-
<i>scheiding leeftijd-periode effecten</i>	+	-	-	-	+	+
<i>aflevering resultaten</i>	-	+	+	+	+/-	-
<i>tijdsinvestering</i>	-	+	+	+	-	-
<i>budgetvriendelijk</i>	-	+	+	+	-	+
<i>valorisatie</i>	+	+	-	+	+	+

3. Haalbaarheidsstudie voor Scenario 5, 6 en 7

Het eerste deel van deze haalbaarheidsstudie handelde over de algemene voor- en nadelen verbonden aan de zeven scenario's voor dataverzameling omtrent loopbanen van individuen. Na een overweging van de belangrijkste factoren, blijken de laatste drie (combinatie-)scenario's het meest opportuun voor verder loopbaanonderzoek. Zowel vanuit wetenschappelijke als vanuit beleidshoek biedt een koppeling van surveys met administratieve data namelijk de meeste perspectieven. Dergelijke data laten toe om verdergaande analyses uit te voeren en beleidsaanbevelingen aan te reiken. De enquêtegegevens kunnen daarbij telkens bij een verschillende groep individuen verzameld worden (cross-secties) ofwel bij dezelfde groep individuen (panel of cohort).

Omdat de keuze nog open ligt, gaan we in dit deel dieper in op de haalbaarheid van scenario 5, 6 en 7. De kosten voor het verzamelen en analyseren van longitudinale data kunnen immers aanzienlijk oplopen, zowel monetair als in termen van tijd. Dit is zeker zo als men betrouwbare informatie wil verzamelen over het gedrag van grote populaties over een langere tijdspanne. Het is daarom ook belangrijk om een gepast design te selecteren waarbinnen de verzamelde data toelaten om de onderzoeksvragen zo accuraat mogelijk te beantwoorden binnen de financiële mogelijkheden. De haalbaarheidsstudie zal in dit deel dieper ingaan op de technische details van de drie designs om uiteindelijk een optimaal design te kiezen voor onderzoek naar loopbanen bij individuen.

In dit deel bekijken we de haalbaarheid van de designs vanuit diverse perspectieven. We gaan hiervoor dieper in op een aantal belangrijke dimensies. Een eerste dimensie betreft de **methodologische dimensie** van een design. Hier worden de methodologische keuzes beschreven die gemaakt moeten worden betreffende de steekproeftrekking, non-respons strategie, onderzoekseenheid, design en sampling en bevragsingsmethode. Eveneens zal hier onderzocht worden welke de mogelijkheden zijn om survey-gegevens te koppelen aan administratieve data, en welke de beperkingen zijn op het vlak van privacy en wettelijke richtlijnen.

In scenario zeven wordt getracht een cohortsurvey te koppelen aan administratieve data. Omdat dit een voortzetting zou zijn van de al bestaande SONAR-studie, zal in een apart deeltje extra aandacht worden besteed aan de daar aanwezige informatie (i.e. in het bijzonder wordt er ingegaan op de achtergrond, steekproef, vragenlijst, non-respons en attritie van de SONAR data). Hier zal dus dezelfde opbouw toegepast worden als in de methodologische dimensie waarbij ook het nut voor dit onderzoek zal worden besproken.

De **financiële dimensie** gaat dieper in op het kostenplaatje verbonden aan de verschillende scenario's. De kosten zijn afhankelijk van de gekozen dataverzamelmethode. Ook hier zullen afwegingen gemaakt moeten worden. Sommige methodes bieden namelijk grotere voordelen (bv. face-to-face bevraging) maar zijn veel duurder in uitvoering. Ook het beheer van de data, de inschatting van het benodigde personeel voor de cleaning en het databeheer zijn belangrijke elementen in het totale kostenplaatje. Om tot een zo realistisch mogelijk scenario te komen, worden kostprijsimplicaties bij elk design als een belangrijk criterium naar voren geschoven.

Een derde dimensie heeft te maken met de **utiliteit en valorisatie** van de designs. Hier wordt in kaart gebracht hoe de gegevens die verzameld worden, verspreid zullen worden. Wordt de data vrij beschikbaar gesteld voor wetenschappelijk onderzoek of tegen betaling? Er groeit een besef in Europa dat data verzameld door de overheid maximaal gebruikt moet kunnen worden, zoals bij de ESS (European Social Survey)-studie die gratis te downloaden is. Doorgaans worden de gegevens in Vlaanderen tamelijk afgeschermd. In deze haalbaarheidsstudie zal afgetast worden in hoeverre

het ESS-model haalbaar dan wel wenselijk is, wat zal leiden tot een voorstel om de valorisatie van de panelgegevens te optimaliseren.

Het **inhoudelijk deel** van de scenario's wordt besproken in een vierde dimensie van de haalbaarheidsstudie. Deze zal echter pas in een volgend rapport besproken worden samen met de uitwerking van de vragenlijst omwille van de onderlinge verbondenheid. Meer specifiek bekijken we dan een aantal theoretische modellen, van waaruit de onderzoeksvragen gedestilleerd worden. We bekijken met welk detail een bepaald scenario bepaalde onderzoeksvragen kan beantwoorden. Hierbij zal vervolgens ingegaan worden op de vragenlijst en de informatie die zeker aanwezig moet zijn voor het beantwoorden van bepaalde onderzoeksvragen. Uiteindelijk wordt er nagegaan in hoeverre het design toelaat om concrete beleidsmaatregelen te evalueren.

3.1 Methodologische dimensie

Wat de methodologie betreft, moeten er beslissingen genomen worden over het soort design. Er zijn verschillende aspecten in een design waarover keuzes gemaakt moeten worden: (1) het steekproefdesign: de doelpopulatie, steekproefkader en steekproefomvang; (2) de steekproefmethode: de initiële sampling en de opvolgregels ('following rules'), tijdsregelingen zoals panellengte en interval tussen de waves, en de onderzoekseenheden; en (3) de bevragingmethode en non-responsstrategie: panelonderhoud, interviewtechnieken, en dataverzameling zoals interviewmethode.

3.1.1 Steekproefdesign

3.1.1.1 Doelpopulatie

Een eerste cruciaal element in een design is een duidelijke definitie van de doelpopulatie. Met het oog op een volledige inventarisatie van loopbaantransities, beoogt deze studie idealiter een getrouwe weerspiegeling van de totale bevolking op arbeidsleeftijd, met andere woorden inclusief de werklozen en niet-beroepsactieven. Integratie van deze twee laatste groepen laat ook toe om de kenmerken van niet-beroepsactieven in kaart te brengen, en ook de antecedenten en gevolgen van (vervroegde) uittrede en herintrede op de arbeidsmarkt te inventariseren. We gaan er daarbij vanuit dat individuen interessant worden voor het onderzoek vanaf het moment dat ze potentieel actief worden op de arbeidsmarkt en tot het moment dat ze de arbeidsmarkt definitief verlaten. We nemen de minimumleeftijd van 16 gezien jongeren vanaf deze leeftijd een leercontract kunnen aanvragen en dus deeltijds kunnen werken. De transitie naar pensioen kan in een longitudinaal perspectief gevolgd worden. Hierdoor hoeven gepensioneerden niet noodzakelijk vanaf de eerste golf aanwezig te zijn.

Een bijzonder aandachtspunt is de voldoende ruime vertegenwoordiging van specifieke subgroepen met een bijzondere beleidsrelevantie. Hier kan onder meer gedacht worden aan allochtonen, werklozen en vijftigplussers. Indien voor deze specifieke deelpopulaties representatieve uitspraken met een voldoende graad van detail beoogd worden, zal in de steekproeftrekking gemikt moeten worden op een proportionele oververtegenwoordiging. Dit lijkt gezien de bijzondere aandacht voor de eindloopbaanproblematiek het meest relevant voor vijftigplussers.

Uit de onderzoekspopulatie, hier zijn dat Vlaamse individuen op arbeidsleeftijd, wordt een steekproef genomen. De omvang van en het soort steekproef zijn afhankelijk van het scenario. Deze steekproef zal in administratieve data gevolgd worden in de tijd en afhankelijk van het scenario een

of meerdere keren bevroegd worden. Gezien we conclusies en aanbevelingen willen doen over de ganse populatie in Vlaanderen, zal de representativiteit gegarandeerd moeten worden. Deze representativiteit is afhankelijk van steekproefmethode, steekproefkader, steekproefomvang en responsgraad.

3.1.1.2 Steekproefkader

Een steekproefkader bestaat uit een lijst met alle eenheden van de doelpopulatie, in dit geval de bevolking op arbeidsleeftijd. Er kunnen hiervoor verschillende registers geraadpleegd worden zoals het telefoonregister, de kieslijsten of het bevolkingsregister. De keuze voor het type register dat uiteindelijk gebruikt zal worden als basis voor de steekproef, is uiteraard afhankelijk van de doelpopulatie en de mate waarin deze volledig gedekt wordt (Decorte, et al., 2008).

Elk steekproefkader zal niettemin een zekere bias met zich meebrengen. Het steekproefkader dient zo gekozen te worden dat de dekkingfout of 'coverage error' minimaal is en een zo accuraat mogelijk weerspiegeling is van de doelpopulatie. Als er bijvoorbeeld individuen of groepen worden uitgesloten die wel behoren tot de onderzoeks- of doelpopulatie, dan zal dit zorgen voor een bias in de data. Daarnaast zullen tevens de surveymethode en budgetbeperkingen de keuze voor het kader beïnvloeden (Buck, et al., 1995; Decorte, et al., 2008).

Het Rijksregister van België is voor loopbaanonderzoek het meest geschikt om steekproeven te trekken. Het Rijksregister is een systeem van informatieverwerking, dat informatie over de natuurlijke personen bevat. Deze houdt de personen in die ingeschreven zijn in de bevolkingsregisters of in de vreemdelingenregisters gehouden in de gemeenten, de personen die ingeschreven zijn in de registers gehouden in de diplomatieke zendingen, de consulaire posten in het buitenland, en de personen die ingeschreven zijn in het wachtregister. Ieder persoon krijgt vervolgens bij zijn eerste inschrijving in het Rijksregister een identificatienummer⁶.

De volgende informatie wordt opgenomen en bewaard in het Rijksregister:

- naam en voornamen;
- geboorteplaats en -datum;
- geslacht;
- nationaliteit;
- hoofdverblijfplaats;
- (plaats en datum van het overlijden);
- beroep;
- burgerlijke staat;
- samenstelling van het gezin;

Dit register als steekproefkader gebruiken, kan voor het onderzoek een belangrijke troef zijn om de representativiteit te garanderen en te toetsen. De PSBH (Panel Studie van Belgische Huishoudens) en de EU-SILC (Survey in Income and Living Conditions) baseren zich eveneens op het

⁶ Wet van **8 augustus 1983** tot regeling van een Rijksregister van de natuurlijke personen. (B.S. van 21 april 1984). http://www.juridat.be/cgi_loi/loi_N.pl?cn=1983080836

Rijksregister van de natuurlijke personen. Het Rijksregister geeft informatie over de officieel geregistreerde situatie van individuen. Een aantal niet-geregistreerde groepen zullen daardoor niet aanwezig zijn in dit register, zoals illegalen, mensen zonder vaste verblijfplaats en daklozen. In sommige gevallen kan de officieel gedocumenteerde situatie ook afwijken van de werkelijke situatie. Een individu kan bijvoorbeeld officieel gedomicilieerd zijn bij de ouders, terwijl deze in feite samenwoont met diens partner (Geurts, 2006). Zo ook kan er enige tijd overgaan voor een verandering in de feitelijke situatie opgenomen wordt in het register (zoals geboorte, dood, verhuizingen) (Molenberghs, et al., 2003).

Het Datawarehouse Arbeidsmarkt en Sociale Economie baseert de indeling van de bevolking in gezinsposities op persoonskenmerken die geregistreerd worden in het Rijksregister (het geslacht, de leeftijd, aantal personen in het huishouden). De wetgeving op de bescherming van persoonlijke gegevens staat de toegang tot adresgegevens voor wetenschappelijke steekproeven niet in de weg. In de praktijk kunnen er echter verplichtingen zijn die een goede aanpak in de weg kunnen staan. De Privacy Commissie dient zijn toestemming te verlenen om de data te gebruiken.

3.1.2 Steekproefomvang

De steekproefomvang is belangrijk voor de nauwkeurigheid voor de populatieschattingen. De omvang van de steekproef staat echter niet vast. Bij het bepalen van de grootte van de steekproef, zijn verschillende elementen van belang. Het is afhankelijk van het niveau van precisie dat aanvaardbaar is. Bepaalde doelpopulaties hebben een grotere steekproef nodig om accurate schattingen te maken. Het onderzoeksonderwerp en het uiteindelijke doel spelen hier eveneens een belangrijke rol (Decorte, et al., 2008).

Er moet een gepast aantal observaties zijn voor elke gebeurtenis of personen in de (sub)populaties. Wanneer bijvoorbeeld het aantal individuen dat arbeidsmarkttransities maakt tijdens hun carrière beperkt is, zijn er grote steekproeven nodig. Dit geldt eveneens wanneer men gedetailleerde uitspraken wil doen over bepaalde subgroepen in de populatie. Gezien de vroege uittrede van de arbeidsmarkt een belangrijk onderwerp is voor het beleid, dienen oudere werknemers extra vertegenwoordigd te worden in de steekproef.

Er moet daarbij rekening gehouden worden met twee factoren. Ten eerste zal **attritie** het aantal observaties verminderen over de tijd. Bepaalde beleidsrelevante subgroepen kunnen daarbij een grotere uitval kennen over de tijd (bv. de jonge, meer mobiele groep). Het belangrijkste aspect bij een panel is uiteraard dat er een voldoende aantal observaties overgehouden wordt op elk meetmoment voor de populatie en mogelijke subpopulaties. Ten tweede kan de omvang van de steekproef ook beperkter zijn. Wanneer de focus in het panel ligt op een transitie die regelmatig voorkomt, en de **observaties accumuleren** over de golven heen, dan ontstaat er na enkele golven reeds een voldoende groot aantal transities die bestudeert kan worden. Een voorbeeld is de transitie naar werkloosheid. Hoewel het aantal transities uit werkloosheid bvb. klein kan zijn tussen de golven, kan dit significant groter worden na een aantal golven. Omgekeerd geldt natuurlijk ook dat de steekproefomvang voldoende groot moet zijn wanneer ook zeldzame observaties bestudeerd worden (Buck, et al., 1995; De Keulenaer, 2007).

Beide scenario's (met cross-sectie of met een panel) starten op dezelfde wijze: met een steekproef uit het Rijksregister. Voor een panel is het echter essentieel dat deze steekproef van goede kwaliteit is voor toekomstig succes van het panel (cf. infra). Om een representatieve steekproef te verkrijgen, moeten de aantallen zo groot mogelijk zijn. Het is uiteraard altijd beter een grote steekproef te hebben dan een kleine. Maar er moet hierbij rekening gehouden worden met de financiële haal-

baarheid en de tijdslimiet (Magnusson en Bergman, 1990). Met hetzelfde budget kunnen er bijvoorbeeld in het scenario met cross-sections meer individuen bevestigd worden dan in een scenario met een panelopzet.

De PSBH (Panel Studie van Belgische huishoudens) verzamelde informatie van iets minder dan 5000 huishoudens in België, wat neerkomt op ongeveer 2000 huishoudens voor Vlaanderen. Deze steekproefgrootte is adequaat voor toekomstig loopbaanonderzoek om representatief te zijn voor de Vlaamse populatie. Dit zal dan ook het doel zijn voor de omvang van beide scenario's. Wat betreft de cohortstudie van SONAR dienen de leden van de oudste cohort terug bevestigd te worden. Aangezien hier echter sprake zal zijn van attritie, zullen er bijkomende leden getrokken moeten worden om representatief te zijn. Op de SONAR-studie gaan we later nog dieper in.

3.1.3 Steekproefmethode

De keuze voor het steekproefkader heeft vanzelfsprekend diverse implicaties voor de selectie van de meest aangewezen steekproeftechniek. Er zijn verschillende manieren waarop een steekproef getrokken kan worden. De keuze voor een bepaalde methode is afhankelijk van het steekproefkader, de grootte van de steekproef, het steekproefdesign en de responsgraad (Fowler, 1988). De grote uitdagingen op het vlak van panelbehoud of 'panel maintenance' moeten bij de selectie ook in rekening worden gebracht.

Over het algemeen zijn er twee grote types steekproeven. De basis steekproefmethode is de *enkelvoudige toevalssteekproeftrekking (simple random sampling)* in de doelpopulatie. Men kan individuen of huishoudens selecteren aan de hand van een toevalsselectie. Dit is een selectie van individuen van een voorgedefinieerde populatie waar alle steekproeven van een gegeven grootte dezelfde kans hebben om geselecteerd te worden. De basiseenheid van de steekproeftrekking hoeft daarvoor niet het individu te zijn, maar kan ook het huishouden zijn.

Dit komt in de praktijk echter zelden voor en is niet altijd mogelijk of wenselijk. Omdat deze mensen gecontacteerd en face-to-face bevestigd moeten worden, is het minder kostelijk in termen van geld en tijd en gemakkelijker voor veldwerk om andere methoden of combinaties van methoden te gebruiken zoals de *gestratificeerde steekproeftrekking (stratified sampling)* (Buck, et al., 1995). In dit type steekproef wordt er eerst een relevante stratificatievariabele genomen (zoals gemeente of arrondissement) op basis waarvan een correcte representatie van de populatie gewaarborgd wordt. Dan zal het samplingkader in groepen verdeeld worden naargelang de stratificatievariabele waarna een systematische sampling uitgevoerd wordt om een gepast aantal individuen in alle strata te bekomen (Da Vaus, 1986).

Voor grote populaties, zoals voor de Vlaamse bevolking, wordt de populatie soms verdeeld in subpopulaties zoals gemeenten, waarbinnen random samples worden genomen. Een belangrijke reden hiervoor is de bescherming tegen een vertekende selectie (bias) (Goldstein, 1979). Bij de random selectie kunnen de kenmerken van de steekproef namelijk ietwat verschillen van de populatie waaruit ze getrokken wordt (Fowler, 1988). Gestratificeerde steekproeven worden dus bij voorkeur toegepast om meer accurate steekproeven te produceren. Dit brengt evenwel een grotere kost mee van een meer gecompliceerde procedure.

3.1.4 Initiële steekproef en opvolgregels

De steekproef in de eerste golf is essentieel voor toekomstig succes van een panelstudie. Dit heeft een aantal implicaties voor panelstudies. Conventioneel is de steekproefdesign gelijkaardig aan die

in cross-sectionele surveys, geclusterd om administratieve kosten van interviews te verminderen en gestratificeerd om de precisie van schattingen te verhogen in de clusters. Door clustering is er een kleinere variantie van de schattingen, omdat de observaties niet geheel onafhankelijk zijn, wat de effectieve steekproefgrootte vermindert (Buck, et al., 1995).

Het is dus van groot belang dat het paneldesign van dergelijke kwaliteit is die representatief is voor de eerste golf en daarenboven representatief blijft in de daaropvolgende golven. Om dit te garanderen is er een hoge kwaliteit nodig in de eerste steekproef die aan de hand van enkele regels geactualiseerd kan worden (De Keulenaer, 2007).

A. Oversamplen

Het aantal observaties voor subgroepen kan in panels te beperkt zijn of doorheen de tijd te klein worden (De Keulenaer, 2007). Bepaalde specifieke subgroepen kunnen echter beleidsmatig en voor loopbaanonderzoek erg relevant zijn, bv. bruggepensioneerden, loopbaanonderbrekers, tijdelijke werknemers, werklozen, enz. Het zijn dan ook vaak deze groepen die in algemene panels in te kleine aantallen aanwezig zijn waardoor effectevaluaties van beleidsmaatregelen of transitieanalyses niet meer mogelijk zijn. Panelstudies kunnen zich hiervoor toespitsen op specifieke groepen van werkenden en werkzoekenden zoals in het SHARE-‘retirement panel’ dat zich enkel richt tot werknemers vanaf 45 jaar en het Nederlandse CERRA-panel of sectorspecifieke panels.

Een *oversampling* van de subgroepen in een panel kan eveneens een oplossing zijn voor het potentieel probleem van kleine steekproefgroottes (De Keulenaer, 2007). Het overrepresenteren van bepaalde subgroepen in een steekproef zal de kost niettemin sterk verhogen.

Bij het oversamplen moeten er twee factoren overwogen worden: welke kenmerken er geschikt zijn als basis voor de oversampling en hoe dit gedaan moet worden. Oversamplen op basis van een aantal kenmerken die variëren over de tijd en die beleidsrelevant zijn, veroorzaken verschillende problemen. Oversamplen op basis van werkloosheidsstatus bijvoorbeeld kan gepast zijn wanneer het beleid geïnteresseerd is in transitie uit deze status. Maar als het belang ligt bij transitie naar deze staat, is dergelijke oversampling niet gepast. Dan zouden er factoren moeten gebruikt worden die de kans op deze status verhogen (Buck, et al., 1995).

B. Opvolgregels

Bovenop de beslissingen met betrekking tot de steekproef in de eerste golf, dienen er keuzes gemaakt te worden over welke steekprofeenheden zeker behouden moeten worden in elke golf. Populaties veranderen immers doorheen de tijd. Bij het scenario met cross-sectionen zijn er geen opvolgregels nodig omdat er telkens een andere groep individuen wordt bevraagd. Voor de overige scenario's zijn er wel een aantal *opvolgregels* die men kan hanteren voor de opvolging van de leden en de mogelijke update van de steekproef zodat op elke meting in een panel de sample cross-sectioneel representatief blijft. Deze opvolgregels zijn richtlijnen om leden toe te voegen aan de sample en om leden te verwijderen (De Keulenaer, 2007).

Voor een cohortsurvey zijn de opvolgregels eenvoudig, aangezien de definitie van het lidmaatschap voor deze groep vaststaat. Op elk meetmoment moet men trachten interviews bij alle originele cohortleden te verkrijgen. Er zal daarbij eveneens telkens getracht worden informatie te verzamelen over de individuen in het huishouden van de originele cohortleden, maar deze dienen niet opgevolgd te worden omdat zij geen deel uitmaken van het lidmaatschap van de cohort. Afhankelijk van het onderzoeksdoel kunnen er daarbij wel individuen toegevoegd worden doorheen de jaren, zoals bvb. immigranten van dezelfde geboortecohort als de originele leden (Buck, et al., 1995).

Een panel vereist een complexer design, want het is belangrijk om de originele leden (huishoudens of individuen) te behouden doorheen de tijd. Hiervoor moet er up-to-date informatie beschikbaar zijn over de woonst van de respondenten. Het gebruik van administratieve data zal deze taak vergemakkelijken, gezien deze gegevens van de leden er aanwezig zijn. Ook hier houdt men best contact met de respondenten door bijvoorbeeld postkaartjes, verjaardagskaartjes, kort overzicht van de resultaten, enz. om panelleden geïnteresseerd te houden, om ze te behouden in het panel (De Keulenaer, 2007).

C. Verwachte responsgraad

In het algemeen zien we een daling in de responsgraad over de tijd heen. Dit kan de omvang van de steekproef beïnvloeden, maar ook de representativiteit aantasten. Non-respons kan problematisch zijn en is een belangrijke bron van fouten in surveys (bias in de data). De responsgraad kan wel berekend worden, maar er is vaak weinig geweten over het effect van de non-respons op de data omdat het moeilijk is om de niet-meewerkende individuen te onderzoeken. In hoeverre men bereid is mee te werken, is doorgaans afhankelijk van de leeftijd, het al dan niet gehuwd zijn en het arbeidsstatuut. Weigeringen op vragenlijsten via mail vertonen vaak een bias op onderwijsniveau en interesse. Non-respons in telefoonsurveys heeft een bias naar leeftijd en onderwijs. Deze effecten moeten zo minimaal mogelijk gehouden worden en verdienen prioriteit in het totale design van de surveys (Fowler, 1988).

Non-respons kan een probleem zijn in cross-sectioneel onderzoek, maar is potentieel een groter probleem in panelstudies. Hier moeten immers steeds dezelfde mensen herhaaldelijk gecontacteerd worden. In cross-sectionele surveys maakt het niet uit wie gecontacteerd wordt op elk meetmoment zolang de steekproef maar representatief is voor de doelpopulatie. Non-respons kan in een panel daarentegen op elk meetmoment plaatsvinden. Het cumulatief effect kan dan leiden tot een bias. Zelfs een gematigde uitval elk jaar zal accumuleren tot een grote non-respons over de tijd met incomplete data. Een hoge responsgraad van 90% bijvoorbeeld kan leiden tot tamelijk incorrecte schattingen van verandering (Goldstein, 1979).

In hoeverre dit problematisch is, hangt af van de omvang van de steekproef alsook van de kenmerken van de uitval. Als deze uitval random was, zou er geen bias ontstaan in de data (hoewel het design zou veranderen). De meeste studies wijzen echter uit dat de individuen die niet (meer) meewerken in een panel, veelal verschillen van de meewerkende individuen in bepaalde karakteristieken die meestal van belang zijn voor het onderzoek. Het grootste probleem bij non-respons en attritie is dus het feit dat de uitval meestal disproportioneel verdeeld is. Attritie verwijst naar het verlies van respondenten uit een panel, een non-respons in de tijd. Het zijn daarbovenop vaak bepaalde groepen van de populatie, die interessant zijn voor onderzoek, die de meeste veranderingen vertonen over de tijd, die meer mobiel zijn, beter geschoold,... die meer waarschijnlijk uitvallen doordat zij onbereikbaar zijn, of niet beschikbaar of weigeren mee te werken (Goldstein, 1979).

Deze selectieve uitval leidt tot een bias in de data, en kan er bijvoorbeeld toe leiden dat de reële arbeidsmarktsituatie verkeerd ingeschat wordt. Een gevolg van attritie kan tevens zijn dat de steekproef niet meer representatief is voor de doelpopulatie en/of subpopulaties, waardoor er geen valide uitspraken gedaan kunnen worden over deze groepen. Daarom moeten we een aantal strategieën inbouwen om de non-respons en uitval zoveel mogelijk op te vangen.

Op het gebied van non-respons en paneluitval is voorkomen beter dan genezen. Dit kan men bereiken door het optimaliseren van de contactpogingen, het geven van incentives, het opleiden van de interviewers, of het combineren van verschillende dataverzamelmethode (de Leeuw en Hox, 1998). De beste strategie bij uitval is ook deze bij aanvang zoveel mogelijk proberen te mini-

maliseren. Dit is mogelijk door contact te blijven houden met de respondenten doorheen het jaar (Magnusson en Bergman, 1990). Het is namelijk zeer belangrijk in longitudinale studies om elk meetmoment zoveel mogelijk individuen van de originele sample trachten te behouden (Bijleveld en Van der Kamp, 1998). In het geval dat attritie reeds is opgetreden, kunnen er individuen bijgetrokken en toegevoegd worden aan het panel. In de PSBH bijvoorbeeld werd een dergelijke bijtrekking uitgevoerd van ongeveer 850 Vlaamse huishoudens bij het vaste panel in de zevende golf (Jacobs, et al., 1991).

Als er informatie beschikbaar is over deze non-respondenten dan kunnen er ook gepaste correcties doorgevoerd worden in de schattingen. Bijvoorbeeld als de gemiddelde leeftijd van de niet-respondenten hoger ligt dan bij de respondenten, kan deze laatste groep gecorrigeerd worden door middel van wegingscoëfficiënten. De methode om met attritie om te gaan, bestaat er dus uit te ontdekken hoe niet-respondenten van respondenten verschillen, hoe beide groepen bijvoorbeeld verschillen in een aantal achtergrondkenmerken zoals leeftijd, geslacht, onderwijsniveau,... (Goldstein, 1979).

Een panel kan een mogelijkheid bieden om de effecten van de uitval in te schatten en er aldus voor te corrigeren wanneer de uitval optreedt in latere golven van een panel. Er is dan immers informatie beschikbaar over deze individuen in eerdere golven. Deze gegevens kunnen dan gebruikt worden om de mogelijke bias te schatten die gecreëerd werd door de uitval. Daarbovenop kan er ook informatie ontbreken in de data voor bepaalde individuen op bepaalde momenten in de tijd (item-nonrespons). Deze partiële uitval kan eveneens problemen veroorzaken vooral voor de analysemethoden die complete informatie vereisen waar dus rekening mee gehouden dient te worden in de analyses (Magnusson en Bergman, 1990).

3.1.5 Panelduur, interval tussen en aantal golven

Voor de panelduur, de intervals tussen en het aantal golven, zijn er een aantal criteria die in rekening moeten worden gebracht. Wat willen we meten? Hoeveel verandering willen we meten? En welke zijn de beperkingen voor de organisatie en kosten geassocieerd met de designs. De lengte van het panel, het interval en het aantal golven beïnvloeden uiteraard ook de kwaliteit van de verzamelde data en de hoeveelheid fouten in de data (De Keulenaer, 2007).

A. Lengte van het panel

De lengte van het panel wordt mede bepaald door het doel van het onderzoek, door wat men wil onderzoeken. Als men bijvoorbeeld de inkomensdynamiek wil meten, dan kan een panel van vijf jaar voldoende zijn. Een studie naar de loopbaanontwikkeling en de werk-leven balans in een levensloopperspectief zal echter een langere observatieperiode vereisen om de ontwikkelingen en veranderingen in kaart te brengen. Langere panels bieden uiteraard altijd meer informatie.

Hoe langer een panel is, hoe meer werkloosheidsperiodes geobserveerd kunnen worden tijdens de looptijd en bijgevolg hoe meer accurate schattingen gemaakt kunnen worden voor overlevingsfuncties. Panels met een lange looptijd hebben evenwel nadelen. Zo zal het moeilijk zijn om een representatief cross-sectionele panel te behouden doorheen de tijd (De Keulenaer, 2007).

B. Interval tussen golven

Voor het interval tussen de interviews moeten drie overwegingen gemaakt worden. Ten eerste de hoeveelheid tijd nodig om een significante verandering of gebeurtenis waar te nemen. Het opspo-

ren van veranderingen in tewerkstelling of financiële patronen kan relatief snel plaatsvinden. Een individu kan tevens de arbeidsmarkt verschillende keren in- en uittreden op een jaar tijd. Hierdoor vereist een onderzoek naar veranderingen in status en financiële patronen relatief korte tijdsintervallen. Maar de impact van een loopbaanonderbreking kan pas na een aantal jaren te zien zijn, waardoor langere intervallen meer geschikt zijn (Kasprzyk, et al., 1989).

Ten tweede bepalen ook methodologische argumenten de tijd tussen de golven. Naarmate het interval groter is, vergroot ook de kans op fouten omwille van geheugeneffecten, is er een grotere kans dat men weigert om mee te werken, en verhoogt dus de kans op attritie. De optimale frequentie voor dataverzameling is afhankelijk van de mogelijkheid van individuen om informatie op te halen, daar panels vaak retrospectieve vragen stellen voor informatie over de verstreken periode sinds de laatste survey. Dit is namelijk nodig om continue informatie te verzamelen over de veranderingsprocessen.

Tenslotte is het budget van het onderzoek een belangrijke overweging bij het bepalen van het tijdsinterval. Als de onderzoeker een fenomeen over een vaste tijd wil bestuderen, dan zullen meerdere interviews uiteraard meer geld vereisen. Hoe langer het interval tussen de golven, hoe lager de kost. Hoe langer de intervallen, hoe langer de referentieperiode waarvan de respondent gebeurtenissen moeten ophalen (Kasprzyk, et al., 1989). Panels als de PSBH verzamelen ieder jaar informatie van de respondenten. Het arbeidsaanbodpanel van OSA daarentegen verzamelt elke twee jaar gegevens van de respondenten.

3.1.6 Onderzoekseenheid

Het OSA-arbeidsaanbodpanel en de *Labour Force Survey* zijn typische voorbeelden van surveys die individuen als onderzoekseenheden hebben. Andere studies, zoals de *PSBH*, laten een aggregatie op huishoudniveau toe omdat hier het huishouden als onderzoekseenheid gekozen werd. Verrijking van de Europese huishoudpanels met 'metavariabelen' laat zelfs aggregatie op het niveau van landen en regio's toe. We bekijken in deze paragraaf de voor- en nadelen van de verschillende opties voor dynamisch arbeidsmarktonderzoek.

De eenheid van analyse in longitudinale data is doorgaans het individu en niet het huishouden. Het is immers niet evident om een longitudinaal huishouden op een accurate manier te definiëren zodat het zou toelaten om de eenheden te volgen over de tijd. Nieuwe gezinnen en huishoudens worden regelmatig gevormd en bestaande gezinnen zijn onderhevig aan veranderende leden. Het individu daarentegen is stabiel in een longitudinale context. Dit betekent niet dat longitudinale surveys niets kunnen zeggen over huishoudens, integendeel. Maar de noodzakelijke informatie wordt afgeleid van individuen die gerelateerd zijn aan hun gezin of huishoudcontext (die verandert over de tijd) (Buck, et al., 1995). Een steekproef van individuen via een gestratificeerde sampling op basis van gemeenten uit het Rijksregister heeft als bijkomend voordeel de eenvoud van analyse voor externe gebruikers.

Maar een onderzoek van een steekproef op basis van individuen heeft ook nadelen. Er is aan de ene kant de hogere kost die eraan verbonden is. En anderzijds negeert men de invloed van intra-huishoudprocessen op loopbaanbeslissingen en transitie van individuen omdat deze niet rechtstreeks onderzocht werden. Een aantal specifieke huishoudtypes zou in een steekproef van individuen eveneens te klein worden om representatieve conclusies te trekken. Steekproeven op basis van individuen zouden met name een ondervertegenwoordiging creëren van eenpersoonshuishoudens, huishoudens van bejaarden en eenoudergezinnen. Dit zou ook leiden tot een oververtegen-

woordiging van huishoudens met drie of meer kinderen en samengestelde huishoudens (Jacobs, et al., 1991).

Onderzoek naar individuen én huishoudens kan dan ook enkel effectief door een steekproef van huishoudens te nemen. Een zelfwegende steekproef voor individuen geeft immers geen zelfwegende steekproef voor huishoudens, terwijl dit omgekeerd wel zo is. Het huishouden als basiseenheid sluit geenszins een analyse uit op individueel niveau. Als voorbeeld kunnen we hiervoor grijpen naar de PSBH-dataset. Inhoudelijk is het eveneens aantrekkelijker om huishoudens als basiseenheid van de steekproef te nemen. We zijn namelijk geïnteresseerd in de veranderingen op de arbeidsmarkt als gevolg van overwegingen op huishoudniveau. Voor toekomstig loopbaanonderzoek is het daarbij ook interessant te zien hoe de carrières van alleenstaanden en eenoudergezinnen evolueren naast die van tweeverdienergezinnen bijvoorbeeld.

Een steekproef van huishoudens heeft als bijkomende voordelen dat ze goedkoper zijn en relevanter omwille van clustering van loopbaanbeslissingen in het huishouden. Je moet immers minder huishoudens bezoeken voor dezelfde steekproefomvang. De representativiteit op huishouden en op individueel niveau en grotere knowhow internationaal omwille van brede verspreiding van huishoudpanels spreken eveneens in het voordeel van huishoudpanels.

Als nadelen zijn er natuurlijk de non-repons van huishoudens, de complexere weegfactoren en complexere datastructuur. Een typisch huishoudpanel vereist daarbij een complexer design dan een cohortsurvey omdat ze zowel representatief wil zijn op individueel als op huishoudniveau. Hiervoor zijn er mechanismen om een goede representativiteit te verkrijgen van de populatie van personen en huishoudens waarin zij verblijven over de tijd, en vooral voor nieuwkomers in deze populatie. Dergelijke complexiteit verhoogt de vereiste tijd en kosten voor het opzetten van zo een panel.

De opvolgeregels, die de representativiteit van cross-sectionele gegevens vertegenwoordigen, gaan als volgt. In de eerste golf worden alle individuen en huishoudens als originele sample leden (OSL) gedefinieerd. In de daaropvolgende golven wordt getracht interviews af te nemen van alle leden of huishoudens met ofwel een OSL ofwel een individu geboren aan een OSL ongeacht ze lid waren van de originele sample. In praktijk is er echter een verschil in de behandeling van nieuwe leden die niet meer samenleven met OSL. Deze mensen worden doorgaans niet meer betrokken (zie bv. BHPS) (Buck, et al., 1995).

In huishoudpanels kan de doelgroep van huishoudens in de eerste golf verzameld worden van een sampling kader gebaseerd op huishoudens. Hier zullen de personen (en hun huishoudens) geselecteerd worden en daarna gevolgd worden in de tijd ongeacht ze in hetzelfde huishouden blijven of niet. Dit design bereikt meer cases en laat een veralgemening toe naar een bredere populatie dan een cohort design (Buck, et al., 1995).

3.1.7 Bevragingmethode en non-responsstrategie

A. Bevragingmethode

De bevragingmethoden bepalen in sterke mate de kosten van het scenario. Er is een ruime keuze aan mogelijkheden, gaande van telefonisch, per post of via het web, tot face-to-face. Er is ook steeds meer ervaring met multi-method benaderingen, waarbij verschillende van deze technieken gecombineerd worden (met veelbelovende effecten op de responsgraad). De keuze van de bevragingmethode heeft repercussies voor de inhoud van de bevraging, de validiteit en betrouwbaar-

heid van de antwoorden, de responsgraad en drop-out in geval van een panelopzet, maar ook de kostprijs. Deze criteria worden mee in rekening gebracht bij de uitwerking van een 'optimaal design'. De keuze voor de bevragingmethode is sterk afhankelijk van de scope van het onderzoek, hoeveel informatie men wil verzamelen en in welke mate men op elk onderwerp wil ingaan.

Buck (1995) geeft een aantal bijkomende overwegingen mee die van invloed kunnen zijn op de selectie van de meest aangewezen bevragingmethode. Bij een panel zou in het hart van de vragenlijst een vaste vragenlijst moeten bestaan die elke golf opnieuw herhaald wordt. De selectiecriteria voor de samenstelling van dit 'stabiele' luik zijn de beleidsrelevantie, het onderzoeksbelang, en de kans op verandering. De vragen moeten zo identiek mogelijk zijn over de tijd. Het is wenselijk om bij het begin van het panel zoveel mogelijk informatie te verzamelen over het proces dat het individu heeft geleid tot zijn huidige situatie (en dit met de bijhorende datums). Dit kan een levensloopreconstructie zijn over verschillende onderwerpen, of een opsomming van datums van begin van verschillende statussen (zoals arbeidsmarktstatus, burgerlijke status). Dit kan belangrijk zijn om problemen van links censoring in analyse van transitie te verminderen.

Om de continuïteit te garanderen, wordt best zoveel mogelijk informatie verzameld over veranderingen tussen twee golven (in plaats van een meting van de status op het moment van de bevraging). Dit is noodzakelijk voor een juiste ordening van gebeurtenissen en veranderingen en om complete informatie te hebben over de vele veranderingen die kunnen plaatsvinden tussen de golven. Om meetfouten te kunnen inschatten, is het ook wenselijk om de referentieperiode voor de continue informatie te laten overlappen met die van de voorgaande golf. Om de representativiteit van het panel te controleren, is bovendien herneming van belangrijke vragen van andere nationale surveys belangrijk (met het oog op vergelijking). Deze factoren verlengen de interviewtijd en dus ook de kosten. De gemiddelde tijd per individu voor het verzamelen van huishoudinformatie, een *drop off* vragenlijst af te werken en opvolginformatie te bekomen, bedraagt ongeveer een uur in BHPS. Dit zou een tamelijk aanvaardbare limiet zijn voor een hoge respons op langere termijn.

Omdat er al een heel deel van de informatie vervat zit in administratieve data, zal de lengte van de vragenlijst en het interview mogelijk korter zijn ten opzichte van andere panels. Er is namelijk continue informatie over de loopbaanevoluitie van individuen. Tevens zijn er gegevens beschikbaar over de huishoudens en de interhuishoudelijke relaties. In de vragenlijst is het dan ook de bedoeling te peilen naar belangrijke additionele informatie over individuen. Cross-sectionele surveys zullen zowel retro- als prospectieve vragen inhouden om voldoende informatie te verkrijgen over de loopbaanpatronen van individuen. In een panelopzet kan men daarbij eveneens peilen naar de intenties en toekomstige gedragingen (om vervolgens in een volgende golf te meten of deze zich uiteindelijk voldaan hebben).

B. Bevragingmethode en non-respons

De meest klassieke dataverzamelmethode is de face-to-face bevraging. Deze methode veronderstelt persoonlijk (face-to-face) contact tussen de interviewer en de respondent. De interviewer stelt de vragen aan de respondent en noteert vervolgens de antwoorden. De interviewer kan verdere duiding geven bij bepaalde vraagstellingen en kan de respondent stimuleren om te antwoorden. Deze methode geeft over het algemeen de grootste responsgraad. Een nadeel van dergelijke methode is echter de hoge kostprijs (interviewers, verplaatskosten, langere tijdspanne). Interviewers dienen daarbij ook getraind te worden om de vragenlijst met minimale bias af te nemen. De aanwezigheid van de interviewer alleen al kan bv. leiden tot meer sociaal wenselijke antwoorden (Decorte, et al., 2008). Deze methode blijkt nog steeds de meest effectieve om tamelijk accurate data te verzamelen voor een representatief staal van de bevolking.

Een telefoonenquête laat in tegenstelling tot de face-to-face enquête toe om informatie op een relatief korte tijdsperiode te leveren aan een relatief goedkoop tarief (hoewel ook deze kosten al snel kunnen oplopen en vaak onderschat worden). Inhoudelijk heeft deze methode echter te kampen met grote nadelen. De vragen moeten namelijk zo eenvoudig en kort mogelijk zijn. Er kunnen slechts een beperkt aantal vragen gesteld worden. Ook is hier een tamelijk grote non-respons aan verbonden. Bovendien vallen individuen zonder een vaste telefoon buiten het steekproefkader (Decorte, et al., 2008). Op deze methode gaan we niet dieper op in gezien deze inadequaat is om voldoende data te verzamelen voor vernieuwend loopbaanonderzoek.

De schriftelijke vragenlijsten worden verstuurd en terug verzameld via de post. Er is hierbij dus geen interviewer aanwezig. De respondent moet dus de tijd nemen om de vragenlijst in te vullen en moet kunnen lezen en schrijven. Deze methode brengt echter grotere non-respons met zich mee wat de validiteit van het onderzoek in gevaar kan brengen. Ook hebben de onderzoekers weinig controle of de juiste persoon bereikt werd en wie de vragenlijst effectief ingevuld heeft (Decorte, et al., 2008). Ook hier dienen de vragen eenvoudig en duidelijk te zijn voor de respondent wegens een gebrek aan duiding door een interviewer. De afwezigheid van een interviewer creëert wel een gevoel van privacy wat positief kan zijn voor de accuraatheid van de gegevens.

Deze drie methoden (face-to-face, telefoon- en postenquête) zijn de klassieke methoden voor dataverzameling in sociaal-wetenschappelijk survey-onderzoek. Meer en meer worden hierbij computers ingeschakeld als een manier om data te verzamelen. Dit kan verschillende vormen aannemen zoals een deel van een persoonlijk interview (CAPI of computer assisted personal interviewing), van een telefonisch interview (CATI of computer assisted telephone interviewing) of 'pen en papier' interview (CASI of computer assisted self interviewing). De groeiende populariteit van deze methoden komt voort uit bepaalde technologische voordelen zoals vereenvoudiging van de routing. Het zou eveneens voor een snelle en gestandaardiseerde dataverzameling zorgen wat kostenbesparend is. Respondenten blijken gevoelige vragen ook gemakkelijker te beantwoorden door middel van computer. Respondenten en interviewers dienen niettemin over een aantal computervaardigheden te beschikken. Het effect van deze methoden op de kwaliteit van de data is echter nog niet uitvoerig bestudeerd (de Leeuw, et al., 1995).

De laatste jaren wordt het Internet ook meer en meer als medium gebruikt voor dataverzameling. Hierbij wordt de vragenlijst doorgestuurd naar de respondent die zijn of haar antwoorden terug doorstuurt naar de onderzoeker door middel van een computernetwerk. De opkomst van surveys via het Internet is te wijten aan de vele voordelen die hieraan verbonden zijn. De hoge snelheid en de lage kosten zijn de belangrijkste troeven van deze methode van dataverzameling. Niettemin zijn er ook nadelen aan verbonden die de kwaliteit van de surveydata in het gedrang kunnen brengen. Het feit dat niet iedereen Internet of een computer heeft (error m.b.t. de bedekkingsgraad en sampling error), over de nodige software beschikt (metingserror) of te weinig gestimuleerd is om de enquête in te vullen (nonresponsgraad), stellen de grootste problemen (Heerwegh, 2005).

3.1.8 Een korte inleiding tot SONAR

Omdat scenario 7 expliciet uitgaat van het uitbouwen van een loopbaancohortstudie, op basis van de reeds bestaande studie SONAR, willen we de lezer een korte introductie geven. Op die manier kan de haalbaarheid van het continueren en verruimen van de cohortstudie beter gekaderd worden.

3.18.1 Achtergrond

SONAR, de 'Studie van de overgang van ONderwijs naar ARbeidsmarkt', is een longitudinaal onderzoeksprogramma dat een geregelde bevraging van leeftijdscohorten inhoudt. Het vormt een antwoord op de vraag van de afdeling Beleidscoördinatie van het Departement Onderwijs van de Vlaamse Gemeenschap naar de manier waarop jongeren de overgang maken tussen school en werk. SONAR is een interuniversitaire samenwerking: zowel de Universiteit Gent, Universiteit Antwerpen, de Vrije Universiteit Brussel en het Hoger Instituut voor de Arbeid werkten mee (Laurijssen, 2005).

Centraal staat het in kaart brengen van:

- de verschillende opleidingen die jongeren volgen en hun attitudes omtrent de gemaakte studiekeuze,
- de mate van werkervaring die men opdoet tijdens de schoolloopbaan,
- het zoeken naar werk, jobverwachtingen en de kenmerken van de jobs die ze uitoefenen.

Verder komen onder andere ook jobtevredenheid en de verschillende attitudes van jongeren t.a.v. huidig en toekomstig werk aan bod. Het doel van het onderzoek is om op termijn in de uitbouw van een gegevensbank te voorzien die toelaat inzicht te verwerven in de dynamiek en processen die de onderwijs- en arbeidsloopbanen van jongeren kenmerken (Belet, et al., 2004).

Drie geboortecohorten van 1976, 1978 en 1980 werden bevroegd op 23-jarige leeftijd. De twee oudste cohorten werden opnieuw bevroegd op 26-jarige leeftijd. Enkel het oudste cohort werd een derde maal aan de vragenlijst onderworpen op 29-jarige leeftijd. Voor 2009 wordt een tweede meetmoment voorzien van de 1980 cohort dat dan 29 jaar zal worden. In onderstaande tabel worden de cohorten, de grootte van de initiële steekproef en de timing van de bevestigingen samengevat (Laurijssen, 2005).

Tabel 3.1 De verschillende cohorten van SONAR: steekproefomvang en moment van bevraging

<i>Naam</i>	<i>Geboortejaar</i>	<i>N</i>	<i>Bevroegd op 23 jaar</i>	<i>Bevroegd op 26 jaar</i>	<i>Bevroegd op 29 jaar</i>
C76	1976	3015	C76(23) (1999)	C76(26) (2002)	C76(29) (2005)
C78	1978	3002	C78(23) (2001)	C78(26) (2004)	
C80	1980	2993	C80(23) (2003)		C80(29) (2009?)

Bron: Laurijssen, 2005 ⁷

De SONAR data en steekproef vormen een interessant vertrekpunt voor een WSE-loopbaanpanel. In de eerste plaats maakt de longitudinale aanpak causale analyse mogelijk. Daarnaast zorgt de bijkomende informatie over de schoolloopbaan voor een dieper inzicht in richtinggevendende factoren voor arbeidsmarktparticipatie.

⁷ In wat volgt worden de afkortingen, zoals voorgesteld in dit kader, gebruikt om de cohorten en waves te benoemen.

3.1.8.2 De steekproef

Voor elk cohort werd getracht een representatieve steekproef te trekken van 3000 respondenten, verspreid over het Vlaamse landschap. In opeenvolgende golven werd informatie verzameld over cohorten uit 1976, 1978 en 1980. Steeds begon men met een steekproef van 23 jarigen die gevolgd werden en om de drie jaar aan dezelfde vragen onderworpen werden⁸ (Belet, et al., 2004).

Om representativiteit over Vlaanderen te verkrijgen en geografisch gebaseerde bias te voorkomen, werd een gestratificeerde toevalssteekproef gebruikt. In een eerste stap werd het aantal jongeren bepaald dat per gemeente bevestigd zou worden. Hiertoe werden 100 steekproefpunten verdeeld over 82 gemeenten, evenredig met het aantal 20 tot 24 jarigen dat in elk van deze gemeente ingeschreven stond. Het vooropgestelde doel was om voor elk steekproefpunt 30 jongeren te interviewen (totaal ca. 3000). Met het oog op het opvangen van non-respons werden voor elke initieel gekozen respondent drie reserverespondenten gezocht (gebaseerd op gelijk geslacht en gelijkaardige woonplaats). Dit resulteerde in een totaal van 12.000 adressen. Deze werden geclusterd in groepen van vier, bestaande uit telkens een startrespondent en drie reserverespondenten (Belet, et al., 2004).

De representativiteit van de eerste golf van een panelstudie is bijzonder belangrijk om representativiteit in volgende golven mogelijk te maken. Ondanks ingebouwde maatregelen tegen responsselectiviteit (wijze van bevraging, opvang van non-respons door vervanging,...) bleken bepaalde groepen oververtegenwoordigd (cf. infra). Deze over- en ondervertegenwoordiging, evenals selectieve uitval in opvolgende cohorten werden opgevangen door weging (Laurijssen, 2005).

Interviews werden face-to-face afgenomen. Alvorens ze het veld werden opgestuurd, kregen de enquêteurs een korte training. Met uitzondering van de eerste golf van de c76 cohorte werden alle interviews afgenomen met hulp van CAPI (De Rick, et al., 2007). De vragen die aan bod kwamen, werden ingedeeld in zeven luiken. Aanpassingen voor de opeenvolgende golven en cohorten varieerden van data (jaren en maanden) en periodes tot het onderscheiden van huwelijk en samenwonen⁹. Na omzetting werden de data van verschillende golven samengebracht tot één databestand per cohort (Belet, et al., 2004; De Rick, et al., 2007)

3.1.8.3 Vragenlijsten

In wat volgt worden de verschillende luiken van de vragenlijst belicht aan de hand van de versie die gebruikt werd voor c80(23). Een eerste reeks vragen peilt uitvoerig naar de onderwijsachtergrond. Zowel studierichting, afstudeergraden en –scores als meer subjectieve vragen naar tevredenheid

⁸ Zoals eerder werd vermeld is enkel voor het c76 cohort een meetmoment verkregen op 23-jarige, 26-jarige en 29-jarige leeftijd. Het c78 cohort werd enkel op 26 jaar en op 23 jaar bevestigd. Het c80 cohort werd voorlopig enkel bevestigd op 23-jarige leeftijd.

⁹ In c76(23) werd geen onderscheid gemaakt tussen samenwonen met een partner of een huwelijk. Bovendien was een periode van samenwonen of huwelijk niet geregistreerd als een periode met begin- en einddatum, maar als twee aparte registraties die het begin en het einde van zo'n periode omschreven: gaan samenwonen of huwen, en (echt)scheiding. Op basis van de betrokken informatie werden periodes van samenwonen en van huwelijk samengesteld De Rick, K., Duquet, N., Glorieux, I., Laurijssen, I., Maes, L., Saliën, K. en Van Dorsselaer, Y. (2007) Hoe maken Vlaamse jongeren de overgang van school naar werk? Cohorte 1976 (Derde Golf) Technisch Rapport. In SONAR (Ed.) (pp. 1-18): Steunpunt Beleidsrelevant Onderzoek 'Loopbanen van leerlingen en studenten in het onderwijs en de overgang van onderwijs naar arbeidsmarkt'.

en studiekeuze komen aan bod. Dit onderdeel is belangrijk in het onderzoek naar causale verbanden tussen opleiding en werk (Belet, et al., 2004).

In het tweede luik wordt gekeken naar zowel betaalde als onbetaalde werkervaring tijdens de studies. Hierbij gaat het zowel om werk als jobstudent en vakantiejobs als om stages. Het volgende onderdeel focust op zoekgedrag van de werkzoekende (Belet, et al., 2004). In een vierde deel komt het werken zelf uitgebreid aan bod. De respondent wordt o.m. gevraagd het begin en einde van de werkperiode te melden, zijn/haar eerste job te evalueren, te vertellen waaruit deze bestond, welke positie men innam in de organisatie en eventuele opleidingen binnen of buiten de firma die gevolgd werden in het kader van het werk op te noemen. Daarnaast worden ook de niet-werkenden bevroegd. In een volgende module wordt naar de huidige situatie van de respondent gepeild. Net zoals in deel vijf wordt gevraagd naar begin- en einddatum van de huidige job, studie of periode van werk zoeken. Daarnaast wordt ook gevraagd naar het aantal kinderen, broers en zussen en de visie van de respondent op zijn eigen situatie wanneer hij de leeftijd van 30 heeft bereikt (Belet, et al., 2004).

In een zevende en laatste deel staan de achtergrondgegevens centraal. Naast nationaliteit, taal, burgerlijke staat en geslacht wordt o.m. gevraagd naar levensonderhoud, activiteit in het verenigingsleven en achtergrondkenmerken van de ouders en de partner. Tenslotte wordt verzocht de interviewer te quoteren. In een bijkomend onderdeel beschrijft de interviewer de eigen ervaring met de respondent (Belet, et al., 2004).

3.1.8.4 Non-respons en attritie

A. Non-repons

Elk panelonderzoek wordt geplaagd door non-respons (niet deelnemen aan de bevestigingen) en attritie (uitval van respondenten na een of meerdere golven van dataverzameling). Ook het SONAR onderzoek bleef hiervan niet gespaard. Hoewel door het nauwkeurig samplen de verdeling over woonplaats en geslacht nauwelijks vertekeningen vertonen, blijkt dit wel het geval te zijn voor een aantal socio-economische verschillen. Uit een vergelijking van de percentages in de steekproeven c76(23) en c80(23) en de totale populatie bleek dat hooggeschoolden en autochtonen licht oververtegenwoordigd waren in de steekproef. Verder bleek dat het wonen in openstaande bebouwing een hogere kans op deelname betekende. Dit laatste, in combinatie met de relatief grote groep hooggeschoolden en autochtonen in de steekproef, doet een oververtegenwoordiging van sterkere sociale klassen vermoeden (Laurijssen, 2005).

De c80(23) en c76(23) cohorten verschilden eveneens wat betreft de samenstelling van werkenden en studenten. Dit hoeft echter niet noodzakelijk het gevolg te zijn van selectieve respons maar kan bijvoorbeeld ook afhangen van veranderende arbeidsmarktparticipatie (Laurijssen, 2005).

Vertekening van de initiële steekproef stelt de representativiteit van het hele cohort in vraag. Om dit te neutraliseren werden de waarnemingen gewogen. Eerst werd een wegingscoëfficiënt opgesteld, rekening houdend met geslacht, regionale spreiding en opleidingscategorie. Dit gebeurde door de proportie van de groep in de populatie¹⁰ te delen door de proportie van deze in de steekproef. Vervolgens werden de waarnemingen volgens de aldus verkregen coëfficiënten gewogen. Uit ver-

¹⁰ Berekend op basis van de Volkstelling van 2001 Laurijssen, I. (2005) Technisch verslag: Respons en wegingscoëfficiënten SONAR-cohorten. In TOR, O. (Ed.) (pp. 1-32). Brussel: Vakgroep Sociologie; VUB.

gelijking bleek dat de steekproef na de weging dichter aanleunde bij de populatie (Laurijssen, 2005).

B. Attritie

Attritie of selectieve uitval over verschillende golven heen vormt een ander probleem dat zich vooral stelt in het eerste cohort c76 dat bevraagd werd over drie meetmomenten.

Tabel 3.2 De verschillende cohorten van SONAR en de non-respons (attritie) door de jaren heen

Geboortejaar/ leeftijd	23-jarigen	26-jarigen	29-jarigen
1976	C76(23) N=3 015	C76(26) N=2 060	C76(29) N=1 657
1978	C78(23) N=3 002	C78(26) N=2 102	
1980	C80(23) N=2 993		

Bron: Laurijssen, 2005

Uit Tabel blijkt dat van de eerste bevraging (c76(23)) naar de tweede (c76(26)), bijna één derde van de respondenten is uitgevallen. De uitval naar de derde golf was weliswaar minder dramatisch (ongeveer 20%). Toch brengt deze uitval de representativiteit van de studie in gevaar, in het bijzonder indien het zou gaan om selectieve uitval (Laurijssen, 2005).

Uit een vergelijking van beide golven bleek inderdaad een toegenomen vertekening van de steekproef. Vooral op het vlak van onderwijsniveau bleek de bias ten opzichte van de vorige golf te zijn toegenomen. Een nieuwe weging werd uitgevoerd om de steekproef opnieuw aan te passen (Laurijssen, 2005).

3.1.8.5 Relatie van SONAR tot het nieuwe loopbaanpanel

In deel twee werd al een scenario uitgetekend waarin de SONAR-steekproef voorgesteld wordt als basis voor de dataverzameling in het WSE-loopbaanpanel. Niet alleen biedt de cohortpaneldata van SONAR een interessant vertrekpunt voor de analyse van causale relaties met betrekking tot de arbeidsmarkt¹¹. Ze laat ons ook toe om verbanden binnen een breder sociaal kader te bestuderen.

Gezien de steekproef al getrokken is en de data beschikbaar, wordt de investering in het SONAR project een tweede maal benut. De bruikbaarheid van de data voor onderzoek werd al eerder aangetoond, o.m. in de studie van Duquet et al. (2005) naar zittenblijvers, het watervaleffect en de ongekwalificeerde uitstroom uit het onderwijs. Zoals eerder vermeld zou in het WSE-panel de jongste cohort (c80) nog één maal geïnterviewd kunnen worden op 29-jarige leeftijd. De nieuwe data zouden dan met de oude gecombineerd kunnen worden.

Echter, er zijn ook nadelen verbonden aan het gebruik van de SONAR cohorten. Hoewel respondenten worden bevraagd naar de huishoudsituatie, is de verkregen data hieromtrent op zijn best beperkt. Daardoor is het niet zeker of de volledige invloed van partner en kinderen op de loopbaan

¹¹ De laatste golf van PSBH data werd bevraagd in 2003.

in de data is opgenomen. Eerder suggereerden we dan ook al de mogelijkheid deze data aan te vullen met administratieve gegevens.

De keuze voor één cohorte waarvan daarenboven enkel data beschikbaar is uit het begin van de carrière is een bijkomend nadeel. Hoewel het begin van de carrière zeer belangrijk is voor de verdere loopbaan zegt ze niets over de arbeidsparticipatie op hogere leeftijd, een belangrijke beleidsfocus vandaag. Daarnaast geldt dat de effecten die we zien niet noodzakelijk leeftijdseffecten zijn maar ook periode-effecten kunnen zijn die verschillende cohorten op een andere manier beïnvloeden. In zekere mate kan dit probleem echter opgevangen worden door een vergelijking te maken met de overige cohorten.

Ten slotte brengt paneluitval de representativiteit van de steekproef in gevaar. Niet alleen de absolute grootte van de steekproef maar ook haar samenstelling beïnvloedt de representativiteit. Eerder werd vermeld dat er selectieve uitval was tussen de eerste en de tweede golf van c76. Deze werd opgevangen door nieuwe wegingscoëfficiënten te berekenen. Echter, opnieuw wegen kan attritie maar tot een bepaald punt opvangen. Indien de uitval te groot wordt, blijft de steekproef problematisch. Een lichtpuntje is dat de uitval voor de derde golf van het c76 cohort aanzienlijk lager lag dan voor de tweede en voor het c78 cohort lager dan voor c76. Indien deze trend zich doortrekt naar het c80 cohort hoeft de uitval niet problematisch te zijn.

3.2 Financiële dimensie

Dataverzameling in het algemeen en veldwerk voor panelonderzoek in het bijzonder kent een hoge kost. Om tot een zo realistisch mogelijk scenario te komen, worden kostprijsimplicaties bij alle designkeuzes als een belangrijk criterium naar voren geschoven. We trachten in deze paragraaf een algemeen overzicht te geven van de kostprijs van elk van de scenario's. Deze oefening is kwetsbaar omdat prijzen steeds wijzigen en de context in sommige gevallen een sterke invloed heeft op de uiteindelijke kostprijs. Toch doen we de oefening zodat de lezer de verschillende scenario's vanuit één perspectief met elkaar zou kunnen vergelijken.

Tabel 3.3 Financiële raming – Algemeen Scenario met aanvulling vanuit de Kruispuntbank, (in Euro)

	2 000 huishoudens	
	4 500 surveys	
Veldwerk	100% response	60% response
	4 500 surveys	2 700 surveys
ONDERAANNEMING Extern bureau		
Vaste kosten		
Projectmanagement	1 000	
Opstartkost extern bureau	4 000	
Opvolging veldwerk	5 000	
Dataprocesing	1 000	
Variabele kosten		
	40 interviewers	
Training (2 dagen) + kosten: 50 euro	2 000	
Debriefing (1 dag) + kosten: 25 euro	1 000	
Kost per afgenomen interview: 80 Euro	360 000	216 000
Incentives: 10 Euro	45 000	27 000
TOTAAL onderaanneming	€ 419 000	€ 257 000
waarvan vaste kosten	11 000	11 000
waarvan variabele kosten	408 000	246 000
Steekproeftrekking		
Rijksregister ¹² (schatting op basis van tarieven 2008)	4 000	
Koppeling Rijksregister aan administratieve data	4 000	
Personeel		
2 FTE onderzoekers ¹³ / 2 jaar	255 118	
1 FTE administratieve kracht / 2 jaar	104 006	
Panel/ CohortScenario: Onderhoud panel		
Opvolging panelleden (kerst- en verjaardagskaartjes, ...)	(2 500)	(1 500)
Werkingskosten		
Kantoorbenodigdheden, telecommunicatie, software, onderhoud aangekochte uitrusting, ...	102 047	
Uitrusting		
Informaticamateriaal	6 750	
Valorisatie		
Publicaties, presentaties, conferenties, ...	2 500	
Overhead		
17 % op personeels- en werkingskosten	81 332	
-> voor panel en cohortscenario	(81 757)	(81 587)
BTW (21 %)	205 538	171 518
-> voor panel en cohortscenario	(206 152)	(171 922)
Geschatte totale kost		
Algemeen totaal: cross-sectie scenario (5)	€ 1 184 291	€ 988 271
Waarvan intern (equipe)	€ 677 301	€ 677 301
Waarvan extern (enquêtebureau)	€ 506 990	€ 310 970
Algemeen totaal: panel- cohortscenario (6/7)	(€ 1 187 830)	(€ 990 600)
Waarvan intern (equipe)	(€ 680 840)	(€ 679 424)
Waarvan extern (enquêtebureau)	(€ 506 990)	(€ 310 970)

12 Schatting op basis van tarieven 2008.

13 Met gemiddelde anciënniteit van 3 jaar.

Tabel 3.4 Financiële raming – Algemeen scenario: met aanvulling vanuit de Kruispuntbank (budget exclusief alle kosten afgedekt in de steunpuntfinanciering), (in Euro)

2 000 huishoudens		
Veldwerk	100 % response	
ONDERAANNEMING Extern bureau		
Vaste kosten		
Projectmanagement		1 000
Opstartkost extern bureau		4 000
Opvolging veldwerk		5 000
Dataprocesing		1 000
Variabele kosten		
		40 interviewers
Training (2 dagen) + kosten: 50 euro		2 000
Debriefing (1 dag) + kosten: 25 euro		1 000
Kost per afgenomen interview: 80 Euro	360 000	
Incentives: 10 Euro	45 000	
TOTAAL onderaanneming	€ 419 000	
waarvan vaste kosten	11 000	
waarvan variabele kosten	408 000	
Steekproeftrekking		
Rijksregister ¹⁴ (schatting op basis van tarieven 2008)		4 000
Koppeling Rijksregister aan administratieve data		4 000
Panel/ CohortScenario: Onderhoud panel		
Opvolging panelleden (kerst- en verjaardagskaartjes, ...)	(2 500)	
Personeel		
2 FTE onderzoekers ¹⁵ / 2 jaar		Gefinancierd vanuit WSE
1 FTE administratieve kracht / 2 jaar		Gefinancierd vanuit WSE
Werkingskosten		
Kantoorbenodigdheden, telecommunicatie, software, onder-		Gefinancierd vanuit WSE
Uitrusting		
Informaticamateriaal		Gefinancierd vanuit WSE
Valorisatie		
Publicaties, presentaties, conferenties, ...		Gefinancierd vanuit WSE
Overhead		
5 % totaal personeels- en werkingskosten		400 (op steekproef RR)
BTW		Geen BTW verschuldigd binnen WSE
-> panel/cohortscenario	(1 785)	
Geschatte totale kost		
Algemeen totaal	€ 427 400	€ 265 400
Waarvan extra op WSE-budget	€ 8 400	
Waarvan extern (enquêtebureau)	€ 506 990	
Algemeen totaal: panel- cohortscenario (6/7)	€ 431 285	€ 268 285
Waarvan extra op WSE-budget	€ 14 865	
Waarvan extern (enquêtebureau)	€ 506 990	

14 Schatting op basis van tarieven 2008.

15 Met gemiddelde anciënniteit van 3 jaar.

Tabel 3.5 Financiële raming – Algemeen Scenario met aanvulling vanuit de Kruispuntbank: webenquêteering, (in Euro)

	2 000 huishoudens 4 500 surveys	
	100 % response 4 500 surveys	60 % response 2 700 surveys
Steekproeftrekking		
Rijksregister[1] (schatting op basis van tarieven 2008)		4 000
Koppeling Rijksregister aan administratieve data		4 000
Personeel		
2 FTE onderzoekers ^[2] / 2 jaar		255 118
1 FTE administratieve kracht / 2 jaar		104 006
Werking		
Kantoorbenodigdheden, telecommunicatie, software, onderhoud aangekochte uitrusting, ...		50 000
programmakosten		50 000
Uitrusting		
Informaticamateriaal		6 750
Valorisatie		
Publicaties, presentaties, conferenties, ...		2 500
Panel/ CohortScenario: Onderhoud panel		
<i>Opvolging panelleden (kerst- en verjaardagskaartjes, ...)</i>	2 500	1 500
Overhead		
17 % totaal personeels- en werkingskosten	80 984	80 984
BTW	117 045	117 045
17 % totaal personeels- en werkingskosten: panel/cohort	81 409	81 239
BTW	117 134	117 099
Geschatte totale kost		
Algemeen totaal	€ 674 403	€ 674 403
Waarvan intern (equipe)	€ 674 403	€ 674 403
Waarvan extern (enquêtebureau)	€ 0	€ 0
Algemeen totaal: panel/ cohort	€ 523 411	€ 522 205
Waarvan intern (equipe)	€ 523 411	€ 522 205
Waarvan extern (enquêtebureau)	€ 0	€ 0

Tabel 3.6 Financiële raming – Algemeen scenario: met aanvulling vanuit de Kruispuntbank: webenquête (budget exclusief alle kosten afgedekt in de steunpuntfinanciering), (in Euro)

	2 000 huishoudens 4 500 surveys	
	100 % response 4 500 surveys	60 % response 2 700 surveys
Steekproeftrekking		
Rijksregister[1] (schatting op basis van tarieven 2008)		4 000
Koppeling Rijksregister aan administratieve data		4 000
Personeel		
2 FTE onderzoekers ^[2] / 2 jaar		
1 FTE administratieve kracht / 2 jaar		
Werking		
Kantoorbenodigdheden, telecommunicatie, software, onderhoud aangekochte uitrusting, ... programmamakosten		50 000
Uitrusting		
Informaticamateriaal		
Valorisatie		
Publicaties, presentaties, conferenties, ...		
Panel/ CohortScenario: Onderhoud panel		
Opvolging panelleden (kerst- en verjaardagskaartjes, ...)	2 500	1 500
Overhead		
17 % totaal personeels- en werkingskosten		9 860
BTW		3 751
17 % totaal personeels- en werkingskosten: panel/cohort	10 285	10 115
BTW	14 865	14 619
Geschatte totale kost		
Algemeen totaal		€ 71 611
Waarvan intern (equipe)		€ 71 611
Waarvan extern (enquêtebureau)		€ 0
Algemeen totaal	€ 85 650	€ 84 234
Waarvan intern (equipe)	€ 85 650	€ 84 234
Waarvan extern (enquêtebureau)	€ 0	€ 0

3.3 Utiliteits- en valorisatiedimensie

De valorisatie van een survey is een belangrijke volgende stap in het surveyproces na het uitvoeren van de survey en het analyseren van de resultaten. De drie scenario's hebben dezelfde impact op de valorisatie.

3.3.1 Databasemanagement

Een eerste stap in de valorisatie van een survey is het veldwerk en de dataverzameling publiek maken. Hier wordt het onderzoeksdesign in kaart gebracht, informatie voorzien over het doel en de onderzoeksvragen van de survey, de lengte van de survey, de steekproefdesign, de vragenlijst en de (non-) responsgraad. Dit zal aan de hand van onderzoeksrapporten publiek worden gemaakt.

3.3.2 Vrijgeven van de dataset

Een volgende stap in de valorisatie van de survey, is de beslissing of de dataset toegankelijk gemaakt moet worden voor derden. Moeten de data vrij beschikbaar gesteld worden voor wetenschappelijk onderzoek (bvb. ESS via het Internet) of tegen betaling? Vaak worden vier (drog)redenen gegeven om data niet toegankelijk te maken voor een breder publiek (Decorte, et al., 2008).

a. Privacy

De bescherming van de privacy van de respondenten kan een argument zijn om de data niet vrij te geven. Onderzoekers trachten echter geen uitspraken te doen op individueel niveau en de Belgische privacywetgeving beschermt respondenten door de privacy clausules in hun toegangsvoorwaarden. De toegankelijkheid tot administratieve data zoals die van de Kruispuntbank van Sociale Zekerheid toont aan dat privacy in België an sich geen reden is om data niet publiek te maken. Dat betekent evenwel niet dat de onderzoekers hier niet de nodige voorzichtigheid aan de dag moeten leggen om de data werkelijk te anonimiseren.

b. Copyright

De bescherming van de eigendomsrechten van de onderzoeker betreft een ander mogelijk argument. Ook in dit geval bieden de Belgische en Europese wetgeving bescherming voor de wetenschappelijke onderzoekers en wordt van derden steeds gevraagd te refereren aan de documentatie van de data opdat de connectie met de originele dataverzamelaars steeds duidelijk zou zijn.

c. Kwaliteitscontrole

Er kan een zekere mate van terughoudendheid zijn vanuit de hoek van de onderzoekers om de originele dataset vrij te geven omwille van de mogelijkheid om hun analyses en bevindingen te controleren. Dit mag echter geen probleem zijn wanneer het veldwerk goed gedocumenteerd is.

d. 'Scoop'

Vaak worden de originele datasets pas een tijd na de dataverzameling pas vrijgelaten om onderzoekers de tijd te geven om eerst de analyses en resultaten af te ronden voor wetenschappelijke publicaties. Hier kunnen goede redenen zijn maar tegelijk moet een evenwicht gezocht worden tussen de analysecapaciteiten en plannen van de equipe en de actualiteitswaarde van de data voor een grote wetenschappelijke gemeenschap.

3.3.3 Toegankelijkheid en verspreiding

Er zijn verschillende mogelijkheden om de dataset te verspreiden. In een zogenaamde 'open access paradigma' kan de toegang en verspreiding van wetenschappelijke informatie via het Internet sterk bevorderd worden. Het open access paradigma staat voor het gratis en onmiddellijk toegang verlenen tot wetenschappelijke publicaties over het internet. De ESS en EVS zijn bekende voorbeelden waarvan de datasets vrij toegankelijk zijn via het internet. Daarnaast kan ook het onderzoeksteam zelf de data verspreiden. Voorbeelden hiervan kan men vinden in Nederland 'DANS', in Duitsland 'ZA' en in Groot-Brittannië 'UKDA'.

De onderzoeksresultaten die bekomen worden op basis van de data zelf zullen bekend worden gemaakt via klassieke onderzoeksrapporten, artikels die ingediend worden bij wetenschappelijke tijdschriften, en onderzoekspapers die ingediend worden bij nationale of internationale congressen. De wetenschappelijke artikels zullen handelen over de onderzoeksvragen die uit het inhoudelijk luik van de haalbaarheidsstudie naar voren geschoven worden. Verder zullen er inspanningen gedaan worden om de belangrijkste resultaten te valoriseren aan de hand van artikelen in tijdschriften met een bredere maatschappelijke toegang, zoals Over.Werk (WSE-tijdschrift), HRMagazine, e.d.

4. Conclusie: SWOT- analyse

Bij wijze van besluit van deze haalbaarheidsstudie integreren we voorgaande informatie in een SWOT-analyse. Deze analyse dient om conclusies en voorstellen inzichtelijker te maken. Ze beschrijft de sterktes (strengths) en zwaktes (weaknesses), opportuniteiten (opportunities) en bedreigingen (threats) voor administratieve data gekoppeld aan ofwel cross-sectionele surveys, ofwel panelsurveys ofwel cohortsurveys. Op basis van deze SWOT-analyses kan verder gepland worden teneinde scenario 5, 6 of 7 te realiseren.

4.1 Administratieve data gekoppeld aan cross-sectionele surveys (Scenario 5)

Strengths	Weaknesses
<p>Inhoudelijk laten deze data toe om zeer veel informatie te verzamelen doordat de rijke dataset op basis van administratieve data aangevuld wordt met cross-sectionele gegevens. Dit zal de kennis in de breedte vergoten daar we de arbeidsmarkt vrij gedetailleerd kunnen beschrijven.</p> <p>Omdat we via administratieve data aan een longitudinaal panel komen, kunnen de individuele personen gevolgd worden doorheen de tijd. Dit laat een reconstructie toe van de individuele loopbaanontwikkelingen waardoor de verschillende loopbaanpatronen in kaart gebracht kunnen worden.</p> <p>Via administratieve data kunnen er tevens gegevens op huishoudniveau verzameld worden. Dit laat toe om representatief te zijn op individueel én op huishoudniveau.</p> <p>Er is flexibiliteit in de vragenlijst die zich kan aanpassen naargelang de focus van het beleid.</p> <p>Dit type dataverzameling betreft daarbij een eenvoudig design daar de individuen slechts eenmaal bevraagd moeten worden.</p> <p>Er kan een grote steekproef getrokken worden.</p> <p>Gegeven het feit dat er maar één cross-sectionele bevraging plaatsvindt, is er voldoende tijd om de gegevens te verzamelen en te analyseren.</p> <p>Dit scenario is minder kostelijk in vergelijking met de andere twee. Men moet namelijk niet dezelfde individuen opvolgen en opnieuw bevragen.</p>	<p>Inhoudelijk: er is geen informatie in de diepte. Arbeidsmarktbeslissingen en gedragingen kunnen niet verklaard worden omdat de individuen slechts één keer geïnterviewd worden. Deze data zijn vooral geschikt voor beschrijvende analyses van de arbeidsmarkt en vergelijking van verschillende groepen.</p> <p>Op basis van deze data kunnen er geen effecten of causaliteitsverbanden onderzocht worden. Hierdoor kan de impact van beleidsmaatregelen ook niet ten volle onderzocht worden. Beleidsevaluaties blijven in dit scenario aldus minimaal. Retrospectieve vragen kunnen hier in zekere mate de ordening van gebeurtenissen opvragen, maar dit is vaak niet zonder (herinnerings)bias.</p> <p>Er is geen onderscheid mogelijk in leeftijd of periode-effecten. Er kan niet nagegaan worden of een bepaald gedrag te wijten is aan de leeftijd of levensfase dan wel aan de periode of dat bepaald jaar.</p> <p>Er dient een gedetailleerde aanvraag ingediend te worden voor de toegang tot de administratieve data.</p>

Opportunities	Threats
<p>Deze data kunnen zorgen voor vernieuwend loopbaanonderzoek. Er kunnen meer bepaald nieuwe en verdere inzichten verzameld worden in loopbanen en loopbaanonderzoek, die een goede up-to-date beschrijving kunnen verschaffen van de huidige arbeidsmarktsituatie.</p> <p>Door gebruik te maken van administratieve data kunnen individuele personen efficiënt opgevolgd worden met tamelijk accurate gegevens, waar ook gevoelige data doorgaans met een betere accuraatheid bekomen kunnen worden (wat in surveys vaak tot non-respons leidt).</p> <p>Omdat individuen longitudinaal opgespoord worden via administratieve data en een eenmalige survey wordt afgenomen, kan men relatief snel overgaan tot analyses en kunnen de resultaten bijgevolg snel gepresenteerd worden.</p>	<p>De Privacycommissie kan de aanvraag afkeuren waardoor er geen administratieve data gekoppeld mag worden aan surveys.</p> <p>Een grote non-respons kan ervoor zorgen dat er slechts een selectieve groep individuen in de dataset aanwezig is, waardoor de resultaten niet meer representatief zijn voor de ganse doelpopulatie.</p>

4.2 Administratieve data gekoppeld aan een panelsurvey (Scenario 6)

Strengths	Weaknesses
<p>Inhoudelijk: de hoeveelheid informatie accumuleert over de jaren heen. Bovendien zorgt de aansluiting met administratieve data ervoor dat er diepgaande analyses en verklaringen gezocht kunnen worden en accurate informatie verschaft kan worden over loon en loopbaanpatronen. Dit laat toe om de vooropgestelde onderzoeksvragen met een redelijke accuraatheid te beantwoorden.</p> <p>Er is een enorme hoeveelheid informatie in de dataset, waar een aantal thema's en arbeidsmarktgedragingen in de diepte onderzocht kunnen worden</p> <p>Omdat dezelfde individuen herhaaldelijk opnieuw worden bevraagd, kunnen effecten en causaliteit gemeten worden.</p> <p>Dit laat tevens toe om de impact van beleidsregelingen na te gaan en dus het beleid te evalueren.</p> <p>Het is mogelijk om leeftijd en periode-effecten te scheiden van elkaar.</p> <p>Een steekproeftrekking laat tevens toe om analyses op huishoudniveau te doen en rekening te houden met invloeden van de partner. Resultaten kunnen aldus veralgemeenbaar zijn op individueel niveau én op huishoudelijk niveau.</p>	<p>Het opstarten van een panel is vrij kostelijk. Door het supplementaire gebruik van administratieve data, zijn de kosten niettemin minder dan in zuivere panels, daar de opvolging en opsporing van individuen gemakkelijker is en de vragenlijst beperkter.</p> <p>Een panel heeft doorgaans een beperkte steekproefomvang (om budgettaire redenen). Door de jaren heen kan attritie aldus de representativiteit van de steekproef in het gedrang brengen.</p> <p>Bepaalde subgroepen die belangrijk kunnen zijn voor loopbaanonderzoek (vb. de oudere werknemers), kunnen ook te beperkt zijn in aantal om representatieve analyses te doen. Dit kan gecompenseerd worden door een oververtegenwoordiging van bepaalde subgroepen in de steekproeftrekking, hoewel dit ook een extra kost met zich mee zal brengen.</p> <p>Een panel heeft vaak omvangrijke vragenlijsten. Naast een huishoudvragenlijst, zijn er de individuele vragenlijsten die vaak ook retrospectieve als prospectieve vragen bevatten. Dit verhoogt de lengte per interview alsook de kost.</p> <p>Panelsurveys hebben een complexer design (vb. opvolgeregels zijn complexer, zoals wie wordt door de jaren heen opgevolgd en hoe worden nieuwe huishoudens/individuen opgenomen)</p> <p>Er dient een gedetailleerde gegevensaanvraag ingediend te worden bij de Kruispuntbank van Sociale Zekerheid.</p> <p>Er zijn daarbij ook meerdere meetmomenten nodig vooraleer overgegaan kan worden op de analyses van de verzamelde data. Omdat de dataverzameling veel tijd in beslag neemt, zullen resultaten pas op lange termijn gepresenteerd kunnen worden. De investering zal aldus pas ten volle renderen na een aantal jaren.</p>

Opportunities	Threats
<p>Dit scenario laat vernieuwend loopbaanonderzoek toe. Er kunnen verdere en diepgaande analyses gemaakt worden waardoor er meer inzichten verkregen kunnen worden in arbeidsmarktbeslissingen en –gedragingen en waarbij deze ook verklaard kunnen worden. Beleidsevaluaties en beleidsaanbevelingen kunnen hieruit voortvloeien.</p> <p>Administratieve data laten toe mensen efficiënt en effectief op te volgen zowel in de administratieve data als in de parallel lopende panelsurveys, waardoor de tijd en kosten verlaagd kunnen worden.</p> <p>Hierdoor kan tevens data over gevoelige informatie met een tamelijke accuraatheid verzameld worden.</p> <p>Deze rijke dataset kan nuttig zijn in een brede waaier van onderzoeksdomeinen en heeft een grote validiteit.</p> <p>Deze dataset zou dan het best aansluiten op de onderzoeksvragen waardoor de theoretische modellen verder geëxploreerd kunnen worden.</p>	<p>het opstarten van een panel is duur en er zal extra financiering vereist zijn om dit scenario te kunnen bekostigen.</p> <p>De Privacy Commissie kan de aanvraag voor de combinatie van administratieve data met panelsurveys afkeuren.</p> <p>Er kan een grote non-respons optreden, waardoor de representativiteit in de eerste golf in gedrang komt, alsook de volgende jaren niet zeker zijn. Daarbij kan er ook attritie optreden (non-respons in de tijd) waardoor representativiteit voor de doelgroep in gevaar komt alsook mogelijk bepaalde groepen die belangrijk zijn voor loopbaanonderzoek (zoals mobiele mensen) in grote aantallen uitvallen.</p> <p>Beleidsveranderingen kunnen ervoor zorgen dat er geen verdere financiering is waardoor de toekomst van het panel in gedrang komt.</p>

4.3 Administratieve data gekoppeld aan de SONAR cohort-studie (Scenario 7)

Strengths	Weaknesses
<p>De arbeidsmarktsituatie van de individuen kan in de diepte onderzocht worden. Diepgaande verklaringen zijn mogelijk.</p> <p>Er is reeds een grote hoeveelheid informatie aanwezig in de SONAR-dataset over de transitie van school naar werk en andere domeinen, waar we gebruik van kunnen maken en kunnen aanvullen met administratieve data en bijkomende survey op latere leeftijd.</p> <p>Effect- en causaliteit van bepaalde gedragingen kunnen onderzocht worden.</p> <p>Het is mogelijk om beleidseffecten te onderzoeken op het feitelijke arbeidsmarktgedrag waardoor beleidsevaluaties gemaakt kunnen worden.</p> <p>Dit type dataverzameling kent een relatief eenvoudig design (eenvoudiger dan een panelopzet) (bv. omdat de leden van het panel vastliggen zijn de opvolgeregels eenvoudiger).</p> <p>Omdat deze cohorte reeds bevraagd werd, is er na de survey voldoende tijd om de gegevens te verzamelen en te analyseren.</p> <p>Dit scenario kent een lagere kost dan een panel.</p>	<p>Omdat we voortbouwen op een al bestaande cohorte, is attritie al opgetreden, waardoor er een kleine sample overblijft. Dit maakt dat er een bijtrekking nodig zal zijn om terug een representatieve steekproef te krijgen.</p> <p>Ook hier kunnen de deelpopulaties te klein zijn in analyse om alleen te bestuderen ook al zijn ze interessant voor loopbaanonderzoek en beleid.</p> <p>Omdat er enkel gegevens verzameld worden van een bepaalde cohorte, is er een risico dat de resultaten enkel betrekking hebben op deze bepaalde cohorte en niet veralgemeenbaar zijn tot de ganse doelpopulatie (dus representativiteit voor Vlaamse werknemers kan in vraag gesteld worden).</p> <p>Dit scenario kan gezien worden als een verlenging van de SONAR cohorte, terwijl de cohorte in feite voortgezet dient te worden met een eigen invulling en een eigen vragenlijst.</p> <p>Er is geen onderscheid mogelijk in leeftijd en periode-effecten, tenzij er meerdere cohortes gevolgd en bevraagd worden.</p> <p>Een administratieve gedetailleerde gegevensaanvraag is vereist om toelating te krijgen tot administratieve data.</p>

Opportunities	Threats
<p>Het scenario zal vernieuwende inzichten geven voor loopbaanonderzoek. Er kunnen nieuwe en up-to-date analyses van de arbeidsmarkt gedaan worden waardoor een beter begrip van de dynamische processen hieromtrent zal ontstaan.</p> <p>Via de administratieve data kunnen we de groep individuen snel en efficiënt opsporen en volgen en kunnen we informatie verkrijgen die gevoelig ligt en wat in surveys vaak een grote non-respons oplevert.</p> <p>Er kunnen relatief snel resultaten opgeleverd worden</p> <p>Er kunnen gegevens op huishoudniveau verzameld worden en de invloed van de partner kan dan bestudeerd worden.</p>	<p>De privacycommissie kan de aanvraag afkeuren om surveygegevens te koppelen aan administratieve data.</p> <p>Een grote non-respons kan de representativiteit van de resultaten in gedrang brengen waardoor we geen uitspraken kunnen doen over de doelpopulatie. Er kan daarbij attritie optreden of al opgetreden zijn, waar een groot aantal van de bevroegde leden in het SONAR-onderzoek uitvallen of al uitgevallen zijn.</p> <p>Er kunnen moeilijkheden of budgettaire beperkingen om huishoudgegevens te verzamelen van de cohort-leden.</p> <p>Huishoudgegevens zijn mogelijk door een bevraging van de huishoudleden van de cohorte, maar mogelijk is deze niet representatief op huishoudniveau.</p>

5. Bibliografie

- Bakker, B., Arts, K. en van Toor, L. (2005) 'Arbeidsmarktstatistieken in Nederland. De opzet van het sociaal statistisch bestand', *OVER.WERK Tijdschrift van het Steunpunt WAV*, 1(jaargang 15): 50-54.
- Barnett, R. C. (1999) 'A New Work-Life Model for the Twenty-First Century', *The Annals of the American Academy of Political and Social Science*, 562: 143-158.
- Bauer, K. W. (2004) 'Conducting Longitudinal Studies', pp. 75-90, in, *New Directions for Institutional Research*. Wiley Periodicals.
- Belet, H., De Rick, K., Duquet, N., Glorieux, I., Laurijssen, I., Maes, L. en Mattheus, N. (2004) Hoe maken Vlaamse jongeren de overgang van school naar werk? Cohorte 1980 (Eerste golf) Technisch Rapport. In SONAR (Ed.) (pp. 1-181): Steunpunt Beleidsrelevant Onderzoek 'Loopbanen van leerlingen en studenten in het onderwijs en de overgang van onderwijs naar arbeidsmarkt'.
- Bijleveld, C. en Van der Kamp, L. (1998) *Longitudinal Data Analysis: Designs, Models and Methods*. London: Sage.
- Billings, J. (2003) 'Using Administrative Data to Monitor Access, Identify Disparities, and Assess Performance of the Safety Net', *Tools for Monitoring the Health Care Safety Net*.
- Borghans, L. en Kriechel, B. (2008) Verkenning van de mogelijkheden van het gebruik van administratieve data voor het Project Onderwijs-Arbeidsmarkt (pp. 15). The Netherlands: Researchcentrum voor Onderwijs en Arbeidsmarkt.
- Buck, N., Ermisch, J. en Jenkins, S. (1995) Choosing a longitudinal survey design: the issues (pp. 56). Essex, England: ESRC Research Centre on Micro-Social Change, University of Essex.
- Cohen, L. en Manion, L. (1980) *Research Methods in Education*. London: Croon Helm.
- Da Vaus, D. A. (1986) *Surveys in Social Research*. London: George Allen & Unwin.
- De Keulenaer, F. (2007) *Non-Response in the Panel Study of Belgian Households (1992-2002): An Output and Process Evaluation*. Niet gepubliceerde eindverhandeling, Universiteit Antwerpen: Departement Sociologie.
- De Lathouwer, L., Bogaerts, K. en Dillen, H. (2006) *Transities op de arbeidsmarkt. Een oefening met administratieve panelgegevens.*: Academia Press.
- de Leeuw, E. en Hox, J. (1998) 'Nonresponse in surveys: Een overzicht', *Kwantitatieve Methoden*, 19: 31-53.
- de Leeuw, E., Hox, J. en Snijkers, G. (1995) 'The Effect of Computer-Assisted Interviewing on Data Quality: A Review', *Journal of the Market Research Society*, 37(4): 39-61.
- De Rick, K., Duquet, N., Glorieux, I., Laurijssen, I., Maes, L., Saliën, K. en Van Dorsselaer, Y. (2007) Hoe maken Vlaamse jongeren de overgang van school naar werk? Cohorte 1976 (Derde Golf) Technisch Rapport. In SONAR (Ed.) (pp. 1-18): Steunpunt Beleidsrelevant Onderzoek 'Loopbanen van leerlingen en studenten in het onderwijs en de overgang van onderwijs naar arbeidsmarkt'.

- De Vaus, D. (2006) *Research Design: volume 1*. London: Sage Publications.
- De Vaus, D. (2007a) *Social Surveys 2: volume 3, Survey Applications*. Los Angeles: Sage Publications.
- De Vaus, D. (2007b) *Social Surveys 2: volume 4, Survey Quality*. Los Angeles: Sage Publications.
- De Winne, S., Plevoets, E. en Sels, L. (2003) Methoden en technieken voor de uitbouw van organisatiepanels: Een overzicht van het onderzoeksontwerp en -opzet van *PASO Flanders* (pp. 101). Leuven: K.U.Leuven.
- Decorte, T., Mortelmans, D., Tieleberghien, J. en De Moor, S. (2008) Haalbaarheidsstudie naar een repetitieve prevalentieonderzoek onder de algemene bevolking. Gent.
- Devisscher, S. en Van Pelt, A. (2005) Impactanalyse van het systeem van loopbaanonderbreking/tijdskrediet in België. Brussel: IDEA consult.
- Duquet, N., Glorieux, I., Laurijssen, I. en Van Dorsselaer, Y. (2005) Problematische schoolloopbanen; Zittenblijven, waterval en ongekwalificeerde uitstroom in het secundair onderwijs. In arbeidsmarkt', S. B. O. L. v. l. e. s. i. h. o. e. d. o. v. o. n. (Ed.) (pp. 1-54): Vakgroep Sociologie, Onderzoeksgroep TOR.
- Fowler, F. J. (1988) *Survey Research Methods*. Beverly Hills: Sage Publications.
- Frans, D. en Mortelmans, D. (2008) Tijdsregelingen in de loopbaan. Terugblik, stilstand en toekomstperspectieven. (pp. 88). Brussel: WSE-rapport.
- Geurts, K. (2006) 'De arbeidsmarktpositie van alleenstaande ouders: Nieuwe bevindingen uit het Datawarehouse Arbeidsmarkt en Sociale Bescherming', *Steunpunt WAV*: 46.
- Goldstein, H. (1979) *The Design and Analysis of Longitudinal Studies: Their Role in the Measurement of Change*. London: Academic Press.
- Hakim, C. (2000) *Work-Lifestyle Choices in the 21st Century: preference theory*. New York: Oxford university press.
- Hayward, M. D. en Grady, W. R. (1990) 'Work and Retirement Among a Cohort of Older Men in the United States, 1966-1983', *Demography*, 27(3): 337-356.
- Heerwegh, D. (2005) *Web surveys. Explaining and reducing unit nonresponse, item nonresponse and partial nonresponse*. Niet gepubliceerde eindverhandeling, Katholieke Universiteit Leuven: Faculteit sociale wetenschappen.
- Jacobs, T., Loots, I., Marynissen, R. en Scheipers, T. (1991) Steunpunt gezinsdemografisch panel. Panelstudie van Belgische huishoudens: onderzoeksplan. Antwerpen: UA.
- Kalmijn, M. (2002) 'Sociologische analyses van levensloopeffecten: een overzicht van economische, sociale en culturele gevolgen', *Bevolking en Gezin*, 31(3): 1-44.
- Kasprzyk, D., Duncan, G. J., Kalton, G. en Singh, M. P. (1989) *Panel Surveys*. New York: John Wiley & Sons.

Kuppens, A., Mortelmans, D. en Casman, M.-T. (2004) 'Analyse longitudinal des parcours professionnels avec analyse de séquence', pp. 325-360, in: Doutrelepon, R. (Ed.), *Onze Ans de vie en Belgique. Analyses socio-économiques à partir du Panel Démographie Familiale*. Gent: Academia Press.

Laurijssen, I. (2005) Technisch verslag: Respons en wegingscoëfficiënten SONAR-cohorten. In TOR, O. (Ed.) (pp. 1-32). Brussel: Vakgroep Sociologie; VUB.

Magnusson, D. en Bergman, L. R. (1990) *Data Quality in Longitudinal Research*. Cambridge: Cambridge University.

Menard, S. (1976) *Longitudinal Research*. London: Sage.

Moffit, R. en Ver Ploeg, M. (2001) *Evaluating Welfare Reform, in an Era of Transition. Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs*. Washington, DC: National Academy.

Molenberghs, G., Bruckers, L. en Tibaldi, F. (2003) Revision HIS 2001 (pp. 48). Diepenbeek, Belgium: Centrum voor Statistiek, Limburgs Universitair Centrum.

Ruspini, E. (2002) *Introduction to Longitudinal Research*. London: Routledge.

Soens, N., Buyens, D., De Vos, A., Heylen, L., Kuppens, A., Mortelmans, D. en Van Puyvelde, I. (2005) *Belgische loopbanen in kaart traditioneel of transitioneel?*. Gent: Academia Press.

Van Aerschot, M. (2004) *De combinatie van levenssferen doorheen de levensloop: Literatuurstudie*. Antwerpen: Universiteit Antwerpen.

Van der Hallen, P. en Van Mechelen, N. (2001) 'Datawarehouse: een nieuwe bron voor arbeidsmarktonderzoek?', *OVER.WERK Tijdschrift van het Steunpunt WAV*, 1.2(jaargang 11): 78-80.

Van der Hallen, P. en Vanheerswynghele, A. (2004) Noden en behoeften aan gegevens en nieuwe vormen van gegevens inzameling rond arbeid. Inventariserende studie. (pp. 31). Leuven: Steunpunt WAV.

Van Wissen, L. en Meurs, H. (1989) 'The Dutch mobility panel: Experiences and evaluation', *Transportation*, 16: 99-119.

XXX Samenstelling van de SONAR data: Vakgroep Sociologie, Onderzoeksgroep TOR.